



**Universitat Jaume I**

**Escuela Superior de Tecnología y Ciencias Experimentales  
Departamento de Lenguajes y Sistemas Informáticos**

**Integración  
de la Bioinformática  
en la Investigación Genómica  
Cardiovascular: Aplicaciones  
en el Framingham Heart Study**

**Tesis doctoral presentada por:  
Òscar Coltell i Simon**

**Dirigida por los profesores:  
Francisco Toledo Lobo  
José M<sup>a</sup> Ordovás Muñoz  
Ricardo Chalmeta Rosaleñ**

**Castellón, 10 de Junio de 2004**



**Universitat Jaume I**

**Escuela Superior de Tecnología y Ciencias Experimentales  
Departamento de Lenguajes y Sistemas Informáticos**

# **Integración de la Bioinformática en la Investigación Genómica Cardiovascular: Aplicaciones en el Framingham Heart Study**

Tesis doctoral presentada por:  
**Òscar Coltell i Simon**

Dirigida por los profesores:  
**Francisco Toledo Lobo  
José M<sup>a</sup> Ordovás Muñoz  
Ricardo Chalmeta Rosaleñ**

**Castellón, Junio de 2004**



**Universitat Jaume I**

**Escuela Superior de Tecnología y Ciencias Experimentales**

**Departamento de Lenguajes y Sistemas Informáticos**

**Integración de la Bioinformática  
en la Investigación Genómica  
Cardiovascular: Aplicaciones  
en el Framingham Heart Study**

Tesis doctoral presentada por:

**Òscar Coltell i Simon**

Ingeniero en Informática, Licenciado en Ciencias Físicas y Auditor CISA

Profesor Titular de Escuela Universitaria

Para la obtención del Grado de Doctor Ingeniero en Informática

Castellón, 2004

Dirigida por los profesores:

**Francisco Toledo Lobo**

**José M<sup>a</sup> Ordovás Muñoz**

**Ricardo Chalmeta Rosaleñ**



## **Dedicatoria**

*A Dolors, perquè sense el seu amor, suport i paciència no haguera ni escomençat aquesta aventura. I la veritable aventura s'escomença a partir d'ací.*



## Dedicatoria

Al meus pares, Isabel Simón†(6-05-2004) i  
Manuel Coltell†(31-12-1993).

Els m'han ajudat i sempre m'han acompanyat en aquest viatge acadèmic, però no podran veure el fruit final. Allà on estiguen, els dos junts, sé que espiritualment m'alluminaran el camí de la vida.





## Agradecimientos

Me hubiera gustado añadir un anexo a este trabajo para incluir en él la lista de todas las personas que, directa o indirectamente, han contribuido de alguna forma para que haya podido terminar este trabajo. Sin embargo, como no es conveniente para evitar el exceso de extensión de esta memoria, me limitaré a destacar los reconocimientos que he considerado más significativos, pero a todos los tengo en mi corazón.

Esta tesis tiene una doble dedicatoria, a mi esposa Dolors, y a mis padres. Todavía creo que no tengo clara conciencia de cuanta paciencia han tenido conmigo, antes mis padres y ahora, Dolors.

Sin Dolors no estaría aquí. Es decir, ni hubiera solicitado entrar en la universidad, ni hubiera llegado a escribir esta tesis. Entre el primer hecho y el segundo median un montón de años; de trabajo; de horas sin dormir; de nervios, de estancias de investigación y viajes; de alegrías por haber conseguido la aceptación de un artículo o la aprobación de un proyecto, de tristeza por lo contrario; y sobre todo, de apoyo mutuo y mucho amor. Bien es cierto que siempre me ha gustado la investigación pero, cuando acabé la carrera y empecé a buscar trabajo, esta vocación quedo profundamente dormida por las circunstancias vitales y laborales. Dolors ha conseguido, como si fuera la versión en negativo del cuento de la Bella Durmiente, despertar como el príncipe, después de bastantes años, a la “bella” vocación de la investigación. De todas formas, volver a la universidad y volver a imbuirse de su espíritu, cuando han transcurrido varios años desde la graduación, se hace muy duro. Y ella me ha ayudado a superar todos esos obstáculos.

Mis padres me han dado todo cuanto han podido, en la justa medida de las cosas, y siempre han respetado mis decisiones en cuanto a mi trayectoria académica, desde el paso de la escuela primaria al instituto de bachillerato, hasta que quise ir a estudiar Ciencias Físicas, en lugar de Medicina para continuar la profesión de mi padre. Esto se llega a entender cuando se es adulto y, a veces demasiado tarde, cuando ya no hay vuelta atrás. Y ahora resulta que les ofrezco la culminación de un trabajo que se ha realizado curiosamente en el ámbito de la Medicina Cardiovascular. Parece que los más profundos deseos (no expresados ni comunicados) de los padres se manifiestan de forma muy peculiar en los hijos. Ahora que ya no se encuentran entre nosotros en persona, estoy convencido de que nos inspiran en espíritu.

Una mención especial para el Dr. Francisco Toledo, quien tuvo a bien depositar su confianza en mi como supervisor de mi segunda trayectoria doctoral en el departamento. Y por la paciencia que ha tenido conmigo en mis reiterados “asaltos” a su despacho, ya que sus cargos de Vicerrector, anteriormente, y de Rector, en la actualidad, conllevan muchas obligaciones y compromisos, pero, muy a su pesar, poca disponibilidad para asuntos académicos y de investigación. Un pequeño apartado para sus sucesivas secretarías, Rosa y Ana y los administrativos del Gabinete de Rectorado, quienes han ejercido de “correvedile” entre nosotros para salvar la indisponibilidad del Dr. Toledo. Supongo que lo habrán hecho con sumo gusto.

Tampoco se puede olvidar que esta tesis casi no hubiera sido posible sin la oportunidad que se me brindó al poder colaborar, por una parte, con el Dr. José M<sup>a</sup> Ordovás y la gente del Nutrigenomics Laboratory del Human Nutrition Research Center de la Tufts University en Boston; y por la otra parte, con la gente de la Unidad de Investigación en Epidemiología Genética y Molecular (EPIGEM), cuya responsable es la Dra. Dolores Corella. Entre los primeros, quiero mencionar a E Shyong Tai (de Singapur), Chao-Quiang Lai, Larry Parnell, Lu Qi (China), Xien Adiconis, Carolina Vargas (Méjico), Josiemer Mattei, Roberto Elosua (España), Toshiko Tanaka (Japón), Pollyana (Perú-España), Maia (China) y la gente de administración. Entre los segundos, no se puede olvidar a Carmen Sáiz, José Ignacio González, Olga Portolés, Marisa Guillén, Francesc Francés, Paula Carrasco, Silvia Prado y el personal de administración y técnico Manuel Pino y Francisco Jiménez.

Por otra parte, esta tesis tampoco se hubiera podido llevar a cabo, sin contar con la confianza y apoyo de la gente de la Red Temática INBIOMED, cuyo coordinador es el Dr. Fernando Martín, siendo el responsable del nodo IRIS en la UJI el Dr. Ricardo Chalmeta. Un especial agradecimiento por su ayuda en los problemas técnicos del día a día para los investigadores colaboradores Miguel Arregui, Alex Doménech, Carlos Pérez y Mónica Molés. También para mis compañeras Cristina Campos y Reyes Grangel.

No quiero olvidar tampoco al personal del Negociado de Estudiantes y Tercer Ciclo de esta universidad, por su paciencia y competencia en la resolución de los tediosos procedimientos que conlleva la presentación, evaluación y defensa de una tesis doctoral, desde que se presenta el proyecto. Sobre todo, gracias a Belén y a Esperanza. Hasta por mención de terceros ya identifican inmediatamente mi caso sin más detalles. Es decir, que se saben de pe a pa mi vida académica de doctorando.

## AGRADECIMIENTOS

---

Casi terminando, creo que es de bien nacidos considerar los esfuerzos, a veces extraordinarios, a veces como favores, que han realizado la gente del Departamento de Lenguajes y Sistemas Informáticos de esta universidad, al cual pertenezco, tanto profesores como administrativos. Dadas las circunstancias personales dolorosas que he sufrido en este primer semestre del año, quiero dar mis más sinceras gracias especialmente al director del departamento, Dr. Pablo Aibar y a su secretario, Dr. Pedro García, así como al responsable de la comisión de doctorado, Dr. Rafael Berlanga. Su buen hacer y buen gobierno ha creado un clima excelente para que se puedan llevar a cabo, con ilusión y tranquilidad, este tipo de actividades.

Finalmente, no quiero dejar de mencionar a mi abuelo Fidel, el único ascendiente vivo que ahora me queda, al que ofrezco esta tesis para que pueda lucir de nieto doctor durante muchos años.



# ÍNDICE

|   |            |
|---|------------|
| <b>TERMINOLOGÍA.....</b>  | <b>17</b>  |
| <b>PRÓLOGO .....</b>  | <b>21</b>  |
| ANTECEDENTES E HISTORIA .....   | 21         |
| ORGANIZACIÓN DE LA EXPOSICIÓN DE CONTENIDOS Y FUNDAMENTOS .....             | 24         |
| ORGANIZACIÓN DE ESTA MEMORIA .....  | 25         |
| CONVENIOS UTILIZADOS EN ESTA TESIS .....                                    | 26         |
| <b>1. INTRODUCCIÓN.....</b>   | <b>27</b>  |
| 1.1 LA BIOLOGÍA SISTÉMICA.....  | 27         |
| 1.2 LA INFORMÁTICA BIOMÉDICA.....   | 28         |
| 1.3 LA BIOINFORMÁTICA .....   | 29         |
| 1.4 LAS ENFERMEDADES CARDIOVASCULARES .....                                 | 33         |
| 1.5 LA EPIDEMIOLOGÍA Y LAS ENFERMEDADES CARDIOVASCULARES .....              | 35         |
| 1.6 LA SIMBIOSIS ENTRE BIOINFORMÁTICA Y EPIDEMIOLOGÍA CARDIOVASCULAR .....  | 54         |
| 1.7 PROBLEMAS Y PROYECTOS PLANTEADOS EN LOS PROCESOS DE INVESTIGACIÓN ..... | 72         |
| 1.8 BIBLIOGRAFÍA .....  | 95         |
| <b>2. HIPÓTESIS Y OBJETIVOS.....</b>  | <b>115</b> |
| 2.1 HIPÓTESIS .....   | 115        |
| 2.2 OBJETIVOS .....   | 115        |
| 2.3 REFERENCIAS .....   | 118        |
| <b>3. MATERIAL Y MÉTODOS.....</b>   | <b>120</b> |
| 3.1 INTRODUCCIÓN .....  | 120        |
| 3.2 RECURSOS Y METODOLOGÍA DE EPIDEMIOLOGÍA GENÓMICA CARDIOVASCULAR ...     | 121        |

|   |            |
|---|------------|
| 3.3 MÉTODOS ESTADÍSTICOS.....                                       | 133        |
| 3.4 ENFOQUE DE INGENIERÍA INFORMÁTICA Y RECURSOS METODOLÓGICOS..... | 147        |
| 3.5 BIBLIOGRAFÍA.....   | 170        |
| <b>4. RESULTADOS .....</b>  | <b>177</b> |
| 4.1 INTRODUCCIÓN .....  | 177        |
| 4.2 RESULTADOS DE LOS PROYECTOS BIOINFORMÁTICOS.....                | 178        |
| 4.3 ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS .....                    | 208        |
| 4.4 AUDITORÍA BIOINFORMÁTICA.....                                   | 272        |
| 4.5 BIBLIOGRAFÍA.....   | 290        |
| <b>5. DISCUSIÓN.....</b>  | <b>295</b> |
| 5.1 GENERALIDADES.....  | 295        |
| 5.2 DISCUSIONES DE CADA UNO DE LOS GENES .....                      | 295        |
| 5.3 DISCUSIONES DE LA PARTE BIOINFORMÁTICA.....                     | 318        |
| 5.4 BIBLIOGRAFÍA.....   | 329        |
| <b>6. CONCLUSIONES.....</b>   | <b>343</b> |
| 6.1 APORTACIONES .....  | 343        |
| 6.2 CONCLUSIONES GENERALES.....                                     | 345        |
| <b>7. FUTUROS TRABAJOS .....</b>                                    | <b>349</b> |
| 7.1 LIMITACIONES DE LOS RESULTADOS.....                             | 349        |
| 7.2 LÍNEAS DE INVESTIGACIÓN ABIERTAS .....                          | 351        |
| 7.3 TRABAJO FUTURO.....   | 353        |
| <b>8. BIBLIOGRAFÍA GENERAL .....</b>                                | <b>355</b> |
| 8.1 INTRODUCCIÓN .....  | 355        |
| 8.2 ARTÍCULOS PUBLICADOS.....                                       | 355        |
| 8.3 BIBLIOGRAFÍA SOBRE BIOINFORMÁTICA Y DISCIPLINAS ASOCIADAS.....  | 357        |
| 8.4 BIBLIOGRAFÍA SOBRE TÉCNICAS DE DETERMINACIÓN .....              | 358        |
| 8.5 BIBLIOGRAFÍA SOBRE TÉCNICAS DE DESARROLLO DE SOFTWARE.....      | 359        |
| 8.6 BIBLIOGRAFÍA DE CETP .....                                      | 360        |
| 8.7 BIBLIOGRAFÍA DE APOE.....                                       | 363        |

---

|  |            |
|--|------------|
| 8.8 BIBLIOGRAFÍA DE APOA1 .....  | 372        |
| 8.9 BIBLIOGRAFIA DE LIPC.....  | 375        |
| 8.10 BIBLIOGRAFÍA DE SR-BI .....   | 378        |
| 8.11 BIBLIOGRAFÍA DE PLIN.....   | 382        |
| 8.12 BIBLIOGRAFÍA DE RESULTADOS DE LOS PROYECTOS BIOINFORMÁTICOS .....                               | 385        |
| 8.13 BIBLIOGRAFÍA DE AUDITORÍA Y CONTROL EN BIOINFORMÁTICA .....                                     | 390        |
| <b>9. GLOSARIO DE TÉRMINOS .....</b>   | <b>393</b> |
| <b>10. INDICE DE FIGURAS.....</b>  | <b>405</b> |
| <b>11. INDICE DE TABLAS.....</b>   | <b>411</b> |
| <b>12. ANEXOS .....</b>  | <b>415</b> |
| 12.1 CÁLCULO CLÁSICO DEL RIESGO CARDIOVASCULAR: LA ECUACIÓN CLÁSICA DE<br>FRAMINGHAM .....           | 415        |
| 12.2 EL LENGUAJE UML .....   | 420        |
| 12.3 EL PROCESO UNIFICADO DE RATIONAL.....   | 439        |
| 12.4 DESCRIPCIONES COMPLETAS DE LOS GENES .....  | 448        |
| 12.3 TABLA DE CONVERSIÓN DE UNIDADES.....  | 469        |
| 12.4 PROCESO DE IDENTIFICACIÓN DE GENES Y POLIMORFISMOS ASOCIADOS<br>A LAS ECV. TABLA COMPLETA ..... | 470        |
| 12.5 ALGORITMO PRINCIPAL DE LA HERRAMIENTA SQUEEZER2B .....  | 477        |
| 12.6 ALGORITMO PRINCIPAL DE LA HERRAMIENTA SQUEEZER3B .....  | 479        |
| 12.7 ESTRUCTURA DE DATOS PRINCIPAL USADA POR LAS HERRAMIENTAS<br>SQUEEZER2B Y SQUEEZER3B .....       | 482        |





# TERMINOLOGÍA

Las abreviaturas utilizadas en esta tesis son las siguientes:

|         |  |
|---------|--|
| β:      | Beta.  |
| μ:      | Micro.   |
| A:      | Adenina.   |
| AA:     | Aminoácido.  |
| ADN:    | Ácido desoxirribonucleico.   |
| APOA-1: | Gen de la apolipoproteína de tipo A-I.   |
| ApoA-I: | Apolipoproteína de tipo A fracción I.  |
| apoB:   | Apolipoproteína de tipo B.   |
| APOB:   | Gen de la apolipoproteína de tipo B.   |
| ApoE:   | Apolipoproteína de tipo E.   |
| APOE:   | Gen de la apolipoproteína de tipo E.   |
| ARN:    | Ácido ribonucleico.  |
| ASCII:  | American Standard Code for Information Interchange.  |
| ASI:    | Auditoría de Sistemas de Información.  |
| C:      | Citosina.  |
| CETP:   | Gen de la proteína de transferencia del éster de la colesteroína (en inglés, <i>cholesteryl ester transfer protein</i> ) |
| DT:     | Desviación típica (en inglés, SD: <i>Standard Deviation</i> ).   |
| EBCDIC: | Extended Binary Coded Decimal Interchange Code.  |
| ECV:    | Enfermedades cardiovasculares.   |
| EDTA:   | Ácido etilendiaminotetraacético.   |

|                                   |   |
|-----------------------------------|---|
| G:                                | Guanina.  |
| HDL:                              | Lipoproteínas de alta densidad.   |
| HDL-C:                            | Colesterol HDL.   |
| HPLC:                             | Cromatografía líquida de alta presión.  |
| HSL :                             | Lipasa sensitiva a hormonas.  |
| IC:                               | Intervalo de confianza (en inglés, CI: <i>Confidence Interval</i> ).  |
| IMC:                              | Índice de masa corporal (en inglés, BMI: <i>Body mass index</i> ).  |
| LCAT:                             | Gen de la lecitina-colesterol aciltransferasa.  |
| LDL:                              | Lipoproteínas de baja densidad.   |
| LDL-C:                            | Colesterol LDL.   |
| LH:                               | Apolipoproteína lipasa hepática.  |
| LIMS:                             | Laboratory Information Management Systems.  |
| LIPC:                             | Gen de la apolipoproteína lipasa hepática.  |
| M:                                | Molar.  |
| mmol:                             | Milimol.  |
| MR:                               | Masa molecular relativa.  |
| MUFA:                             | Ácidos grasos monoinsaturados.  |
| OR:                               | Odds Ratio.   |
| pb:                               | par de bases (en inglés, bp: <i>base pair</i> ).  |
| PCR:                              | Reacción en cadena de la polimerasa.  |
| PKA:                              | Proteína kinasa A.  |
| PLIN:                             | Gen de la apolipoproteína perilipina.   |
| PUFA:                             | Ácidos grasos poliinsaturados.  |
| R <sup>2</sup> (r <sup>2</sup> ): | Coefficiente de determinación (modelos de regresión lineal).  |
| RFLP:                             | Polimorfismos de longitud en los fragmentos de restricción.   |
| SATFA:                            | Ácidos grasos saturados.  |
| SATFAT:                           | Ácidos grasos saturados.  |
| SFA:                              | Ácidos grasos saturados.  |
| SNP:                              | Polimorfismos de único nucleótido.  |
| SR-BI:                            | Gen del receptor del recolector de clase B y tipo I (en inglés, <i>scavenger receptor class B type I</i> ). |

---

|        |                                     |
|--------|-------------------------------------|
| T:     | Timina.                             |
| TAG:   | Triglicéridos o triaglicéridos.     |
| TaqIB: | [Glosario] Enzima de restricción.   |
| TG:    | Triglicéridos o triaglicéridos.     |
| VLDL:  | Lipoproteínas de muy baja densidad. |



# PRÓLOGO

La ciencia, a pesar de sus progresos increíbles, no puede ni podrá nunca explicarlo todo. Cada vez ganará nuevas zonas a lo que parece inexplicable. Pero las rayas fronterizas del saber, por muy lejos que se eleven, tendrán siempre delante un infinito mundo de misterio.

Gregorio Marañón

La mayoría de las ideas fundamentales de la ciencia son esencialmente sencillas y, por regla general, pueden ser expresadas en un lenguaje comprensible para todos.

Albert Einstein

Esta tesis tiene sus inicios en el verano de 1997, cuando el doctorando, becado por el programa de becas de Bancaja, realizó una estancia como investigador invitado, entre julio y septiembre, en el *Lipid Metabolism Laboratory* del *JM-USDA Human Nutrition Research Center on Aging at Tufts*. Este es un centro de investigación en todos los campos relacionados con el envejecimiento y la nutrición cofinanciado por el gobierno federal de los Estados Unidos, a través del Departamento de Agricultura (USDA), y la Tufts University de Boston, Massachusetts (EE. UU.) El investigador que actuó como receptor y tutor del doctorando en su estancia fue el Dr. José M<sup>a</sup> Ordovás.

## Antecedentes e historia

En 1997 estaba ya en marcha el Proyecto del Genoma Humano, con las dos líneas de trabajo en declarada competencia, la pública, y la privada. Todavía estaba lejos la foto de la “reconciliación” mediada por el ex-presidente de los EE. UU. Bill

Clinton. En este gran marco científico, y recogiendo los frutos que se iban desgranando, la unidad de investigación del Dr. Ordovás, una autoridad mundial en lípidos y posteriormente en Nutrigenómica, desarrollaba sus experimentos en el ámbito del *Framingham Study* (iniciado como *Framingham Heart Study* y posteriormente seguido por el *Framingham Offspring Study*). Estos experimentos se realizaban aplicando rigurosamente los procedimientos y protocolos de la Biología Molecular y Genética para la extracción de ADN de muestras de sangre de los sujetos del estudio, y el tratamiento posterior del ADN extraído hasta llegar a detectar la expresión genética correspondiente.

El tratamiento de los datos adquiridos de los sujetos y los generados en los experimentos se apoyaba simplemente en la utilización de hojas de cálculo Microsoft Excel (y en su momento, de la suite Wordperfect). Y el análisis de dichos datos se realizaba con los paquetes estadísticos SPSS y SAS. Por tanto, vista la forma de trabajar con la información, la rigurosidad excelsa de los protocolos de biología molecular no tenía el suficiente reflejo en cuanto a la gestión de los datos biomédicos. Esta es una de las primeras observaciones que pudo realizar el doctorando.

En esa época, se hablaba de un conjunto de herramientas informáticas que se habían desarrollado, fundamentalmente por los biólogos y científicos de perfil similar que trabajaban en el Proyecto del Genoma Humano, para manejar y analizar la ingente cantidad de datos que generaba diariamente dicho proyecto. Estas herramientas, desarrolladas casi exclusivamente para plataformas UNIX i linux, se distribuían libremente a través de páginas Web. Es de destacar una de las más famosas, la de Jurgen Ott de la Rockefeller University. Había algunas excepciones en cuanto a herramientas hechas para las plataformas de APPEL. Sin embargo, no había ninguna conexión entre dichas herramientas y un planteamiento más formal sobre la forma de abordar los problemas en genética y en el descubrimiento del genoma.

Por similitud, se empezaba a hablar de la Biología Computacional (*Computational Biology* o *Computer Biology*), como aquella disciplina que, intentando mejorar la capacidad de la Biología tradicional, había incorporado métodos, técnicas y herramientas de la Ciencia de la Computación (*Computer Science*). Otros hablaban de la Genética Computacional (*Computational Genetics* o *Computer Genetics*) dando a entender que realmente se estaba extendiendo la capacidad científica de la Biología Genética (y Molecular), y no de la Biología entera. Sin embargo, ha prosperado curiosamente un término que no encaja con el uso anglosajón (por el adjetivo “*Computer*” o “*Computational*” calificando a un sustantivo derivado de la disciplina a nombrar): “*Bioinformatics*”. El doctorando supone malévolamente que es un término “ecléctico” inventado para contentar a

todos y no favorecer a nadie. Así, es entre 1997 y 1998 cuando va calando el nombre “Bioinformatics” (muy posteriormente se empezaría a utilizar el equivalente en español de “Bioinformática”).

En la fecha en que se está escribiendo esta memoria, junio de 2004, se tiene la impresión (como el tiempo todo lo cura y borra memorias neurológicas) de que la Bioinformática es ya una disciplina consolidada cuyas raíces se remontan quizá al siglo XIX. Nada más lejos de la realidad. Si bien el ritmo de consolidación de la disciplina ha sido fortísimo, debido al incremento exponencial de esfuerzos e investigadores aplicados en ella, no se puede hablar todavía de una disciplina (¿ciencia?, ¿ingeniería?, ¿tecnología?) madura. Todavía falta bastante para ello, y esta tesis tiene, como una de sus misiones, ponerlo en discusión.

Entonces, a partir de 1997 y durante los años en que se han desarrollado los trabajos de investigación de esta tesis, el marco científico y técnico ha estado caracterizado por la constante evolución de la Bioinformática, complementado por una evolución casi paralela de otra nueva disciplina, la Epidemiología Genómica, surgida de la Epidemiología Genética. Se puede afirmar que esta tesis se gestó con un encontronazo poco amistoso entre la Ciencia de la Computación y la Epidemiología Genética, y mediante los auspicios de la Biología Computacional, ha terminado en un bien avenido matrimonio entre la Epidemiología Genómica y la Bioinformática.

Hay que hacer constar, por otro lado, que la colaboración con el HNRC at Tufts, a través del Dr. Ordovás, se ha mantenido constante además de las sucesivas estancias de investigación realizadas en dicho centro en 1998, 1999, 2000, 2001, 2002 y 2003. Y queda pendiente, y a punto de iniciarse en la fecha de redacción de esta memoria (junio de 2004) una en el año 2004. Fruto de esta constante colaboración han sido los artículos que constituyen la fuente de esta tesis. Sin contar con gran cantidad de artículos y comunicaciones a congresos internacionales y nacionales.

Por otra parte, los resultados y conclusiones vertidos en esta tesis provienen también de la participación del doctorando en varios proyectos de investigación. Algunos de ellos han sido financiados por la Generalitat Valenciana y conducidos por el Grupo de Investigación en Epidemiología Genética y Molecular de la Dra. Dolores Corella (Departamento de Medicina Preventiva y Salud Pública de la Universitat de València). Otros por la Comisión de Ciencia y Tecnología (CICYT) y del extinto Ministerio de Ciencia y Tecnología, liderados por el Dr. Ricardo Chalmeta (Departamento de Lenguajes y Sistemas Informáticos de la Universitat Jaume I). Y fundamentalmente, esta tesis incluye los resultados obtenidos hasta la fecha de la Red Temática de Investigación Cooperativa en Informática Biomédica,



denominada “INBIOMED. “, cuyo coordinador es el Dr. Fernando Martín (Área de Bioinformática y Salud Pública del Instituto de Salud Carlos III de Madrid). Además, los conocimientos de Inteligencia Artificial que han inspirado algunos de los trabajos de esta tesis y, en su origen, la aplicación de enfoques basados en la lógica, así como otros aspectos metodológicos relacionados con la verificación de software, provienen de la estrecha colaboración con el Dr. Francisco Toledo.

## Organización de la exposición de contenidos y fundamentos

La normativa sobre la estructura de las memorias de tesis doctoral resulta bastante flexible y permite, como es el caso de las universidades estadounidenses y de bastantes universidades públicas españolas, confeccionar una memoria a base de reunir varios artículos, la mayoría previamente publicados en revistas de reconocido prestigio. Este fue el planteamiento inicial en la presente memoria, puesto que, hasta la fecha, se dispone de cinco artículos previamente publicados y otro más recientemente aceptado para publicación, todos ellos en revistas indexadas en el *Journal Citation Reports Index* de Thomson ISI (<http://www.isinet.com/>). En cambio, después de haber hecho una profunda reflexión sobre el hecho de que en esta memoria se presentan contenidos pertenecientes a diversas disciplinas que no suelen reunirse habitualmente, se ha optado por aplicar una estructura y enfoques clásicos en donde se incluyen los planteamientos y resultados de cada artículo en forma distribuida. Además, otro aspecto considerado para esta decisión es que la estructura y estilo de dichos artículos, determinado por las revistas a las que se han remitido para su publicación, aparece poco habitual según los cánones empleados en las publicaciones de Ciencia de la Computación e Ingeniería Informática. Así, resulta mejor extraer lo esencial de cada artículo e integrarlo en un esquema tradicional compuesto por las secciones de introducción, material y métodos, resultados, discusión y conclusiones.

Cabe destacar por otra parte, que los artículos que sirven de base a esta tesis presentan un grado de calidad alto en cuanto a importancia e interés científico, contenido y estructura formal. Esto viene apoyado en que las revistas que los han publicado están en la primera mitad de su categoría (alguna está en primer lugar) por su índice de impacto en la clasificación *ISI Journal Citation Reports* que se publica anualmente (<http://go5.isiknowledge.com/portal.cgi>).

En este contexto, el proceso de generación de dichos artículos viene determinado por el diseño y forma de trabajar de los grandes proyectos en donde se han obtenido los resultados, como, por ejemplo, el *Framingham Offspring Study*. Dichos proyectos tienen tal envergadura en recursos y tiempo que deben distribuir

su desarrollo entre varios centros y grupos de trabajo, así como a lo largo de varias generaciones de investigadores. Su potencial de producción científica es enorme, pero, por el contrario, el control de resultados específicos por una sola persona es casi inimaginable. Además, dado que el tratamiento de las “materias primas”, personas, muestras biológicas, etc., y de la información generada es tan complejo, dichos proyectos necesitan el concurso de expertos y personal de muy diversas disciplinas que van, desde la Biología Molecular y Genética, pasando por la Epidemiología Genómica y la Medicina Clínica, hasta la Informática Médica, Estadística, y la Bioinformática. Todos ellos trabajando en equipos multidisciplinares según el enfoque más efectivo fomentado por el Proyecto del Genoma Humano. Por lo tanto, muchas veces existen verdaderas dificultades para deslindar en trabajos publicados surgidos de estos macroproyectos la aportación de cada uno de los autores. Conscientes, tanto el doctorando como sus directores, de dicha dificultad, en esta memoria se ha tratado de reflejar y destacar con la propiedad debida las aportaciones del candidato al título de doctor.

## Organización de esta memoria

Esta memoria se estructura en 11 capítulos y un capítulo final que engloba todos los anexos. Los capítulos son los siguientes:

- Capítulo 1. Introducción.
- Capítulo 2. Hipótesis y objetivos.
- Capítulo 3. Material y métodos.
- Capítulo 4. Resultados.
- Capítulo 5. Discusión.
- Capítulo 6. Conclusiones.
- Capítulo 7. Proyecciones de la tesis.
- Capítulo 8. Bibliografía.
- Capítulo 9. Glosario de términos.
- Capítulo 10. Índice de figuras.
- Capítulo 11. Índice de Tablas.
- Capítulo 12. Anexos.

## Convenios utilizados en esta tesis

Se ha procurado seguir en todo momento las normas de publicación promulgadas por Publicacions de la Universitat Jaume I, que están en consonancia con las normas internacionales. No obstante es necesario hacer algunas indicaciones para evitar confusiones en el manejo del texto:

1. Se escriben en cursiva los términos en lengua extranjera insertos en el texto.
2. Se escriben en cursiva los términos bajo las siguientes condiciones: cuando se enuncian conceptos en una lista en el mismo párrafo, cada vez que se abre un nuevo párrafo para describir el concepto correspondiente, el término que denomina el concepto aparece por primera vez en cursiva.
3. También se utiliza las letras cursivas para destacar conceptos cuando aparecen por primera vez en el texto.
4. Se escriben también en cursiva determinados párrafos que se quieran destacar.
5. Los nombres técnicos u otros términos que interesa destacar, se escriben con letras en negrita.
6. Los ejemplos de instrucciones de lenguajes de programación o similares se escriben con letra courier de tamaño 9. Tanto en los ejemplos aislados como si hay instrucciones insertadas en el texto.

Ningún descubrimiento se haría si ya nos contentásemos con lo que sabemos.

Lucio Anneo Séneca

## 1.1 La Biología Sistémica

El siglo XXI ha significado un cambio de rumbo en la forma de ver las diversas disciplinas que forman una entidad conceptual denominada “Ciencias de la Vida”. En el caso de las disciplinas más básicas, como Biología Molecular, Biología Celular, Bioquímica, Genética etc., se ha abierto un profundo debate basado en el axioma general que afirma que “el todo es la suma de las partes”. Por una parte, el enfoque reduccionista, que es el que se ha venido aplicando desde principios del siglo XX, establece una compartimentación casi rigurosa entre las distintas disciplinas, pero ha sido muy productivo en los pasados 60 años (Jackson et al., 2003). Por otra parte, el enfoque integrador, el del presente siglo, promueve la colaboración de las distintas disciplinas, trabajando todas bajo unos objetivos comunes pero cada una de ellas integrando sus aportaciones particulares.

Este es el planteamiento que Jackson et al.(2003) aplican en la investigación de los sistemas biológicos fundamentales. Ellos afirman que ya se ha llegado a comprender que los sistemas biológicos son mayores que la suma de sus partes. Es decir, que un enfoque integrador en el estudio de los sistemas biológicos puede dar más y mejores frutos que los estudios centrados en las partes de dichos sistemas. Este es, por tanto, el espíritu que anima la creación de una nueva disciplina denominada Biología Sistémica, como una aproximación integrada de principios, teorías, metodologías, técnicas e instrumentos, con el objetivo de analizar grandes cantidades de datos biológicos de muy diversa naturaleza.

Y como efecto colateral, también se está replanteando la forma de ver la Bioinformática, no solamente como una disciplina técnica que se dedica al estudio y tratamiento de las secuencias (ADN, ARN, AA, etc.), sino como una disciplina

científica y técnica que puede compartir el mismo espacio de investigación que las disciplinas mencionadas anteriormente. Es más, la integración de la Bioinformática en cualquier proceso de investigación en Ciencias de la Vida es una necesidad vital y, en algunos ámbitos específicos, un hecho. Quizá deba replantearse la relación entre la Biología Sistémica y la Bioinformática como una jerarquía de generalización-especialización. Esto es, la Biología Sistémica es generalización de la Bioinformática y, en sentido inverso, la Bioinformática es una especialización de la Biología Sistémica. Esta visión filosófico-científica de alto nivel es la que define la forma de abordar los trabajos de investigación de esta tesis.

## 1.2 La Informática Biomédica

La comunidad de investigadores, médicos clínicos y otros profesionales que trabajan en el ámbito de la Biomedicina están planteando un problema común: existe una gran necesidad de procesar la gran cantidad de información adquirida, producida y almacenada en los distintos niveles de información sobre salud (población, enfermedad, paciente, órgano o tejido y molecular o genético) (Martín-Sánchez et al., 2004). Hay un conjunto de subdisciplinas de las Tecnologías de la Información que tradicionalmente han podido gestionar la información de salud en su ámbito de actuación como las siguientes:

- La Informática en Salud Pública ha sido capaz de enfrentarse y resolver problemas a nivel poblacional.
- La Informática Médica ha generado diversas soluciones, la mayoría de ellas inconexas, para manejar con eficacia la información de paciente y de enfermedad.
- La Informática Gráfica y el Tratamiento Digital de Señales, aplicadas en el procesamiento de imágenes médicas, han desarrollado herramientas de gestión y análisis de imágenes en la realización de técnicas no invasivas para el estudio de órganos y tejidos (Martín-Sánchez et al., 2004).
- Fuera del enfoque tradicional, la Bioinformática es la proveedora de herramientas para la gestión y análisis de los datos moleculares, genéticos y proteómicos disponibles.

Sin embargo, las necesidades actuales de la investigación de enfermedades complejas, multigénicas y de alta influencia de factores ambientales no pueden ser cubiertas por las disciplinas anteriores. En cambio se necesitan acciones integradas entre distintas especialidades, tanto médicas como tecnológicas, que ayuden al estudio de las interacciones entre todos los distintos niveles de información

(Martín-Sánchez et al., 2004). Este tipo de acciones ha provocado el surgimiento de una nueva disciplina apoyada por otras con mayor antigüedad: la Informática Biomédica.

Por tanto, la Informática Biomédica es una disciplina que pretende establecer los fundamentos teóricos, las metodologías y las técnicas, y diseñar los procedimientos e instrumentos, que permitan integrar la información gestionada en los distintos niveles de información sobre salud (población, enfermedad, paciente, órgano o tejido y molecular o genético) en un sistema conceptualmente homogéneo. Se cree que la Informática Biomédica desempeñará un papel preponderante en la integración de todos estos nuevos datos y enfoques a la hora de realizar las aplicaciones sanitarias (Martín-Sánchez et al., 2003).

En España, se ha tomado conciencia de la necesidad de fomentar la investigación pública en Informática Biomédica. Fruto de ello es la Red de Informática Biomédica INBIOMED (INBIOMED, 2004), que pretende diseñar nuevos enfoques y nuevas tecnologías ante la necesidad creciente de sistemas capaces de almacenar, relacionar y gestionar todos los datos y la información en biomedicina, así como de desarrollar herramientas de análisis que permitan generar, predecir, modelar y visualizar conocimiento biomédico a partir de los datos experimentales (Martín-Sánchez et al., 2004).

### 1.3 La Bioinformática

La Bioinformática ha surgido como consecuencia de una convergencia multidisciplinar bajo el amparo del Proyecto del Genoma Humano (Collins, 2003). Esta disciplina está orientada fundamentalmente a la investigación aplicada relacionada principalmente con las siguientes áreas: biología molecular, biología molecular y genética (genómica, proteómica, etc.), ciencias biomédicas, ciencias de la computación, matemáticas, física y estadística (Figura 1.1). También es un enfoque interdisciplinario, puesto que al menos un científico experto en biología y otro científico experto en ciencias de la computación, deben colaborar estrechamente para alcanzar un objetivo común, para lo cual necesitan un lenguaje común y, además, aprender uno del otro los respectivos conocimientos fundamentales.

Una definición compendiada de la Bioinformática establece que se trata de una disciplina científica y tecnológica en la que interaccionan en armonía los planteamientos investigadores de la Biología Genética y Molecular, con los enfoques metodológicos y tecnológicos de la Ciencia de la Computación y la Ingeniería Informática, para la obtención y gestión del conocimiento biológico

genómico y proteómico (Coltell et al., 2002). Gracias a la Bioinformática, la secuenciación del genoma ha podido finalizar antes de lo previsto inicialmente, y en la denominada era Postgenoma, la Bioinformática está siendo fundamental en la translación de la información genética a sus aplicaciones biomédicas y sociales. Sin embargo, una de las principales limitaciones en la eclosión de esta disciplina en España ha sido la escasez de profesionales formados en la misma, ya que tradicionalmente las principales disciplinas científicas implicadas han ocupado compartimentos estancos con escasa comunicación entre ellos. Es necesario pues intensificar los esfuerzos en el diseño de los nuevos Planes de Estudio, y en la formación de postgrado para asegurar una colaboración fluida entre los investigadores de las principales disciplinas implicadas.

En esta sección se va a establecer la relación exacta entre la Epidemiología Genómica y la Bioinformática y se va a describir las soluciones que aporta la segunda agrupadas por categorías.

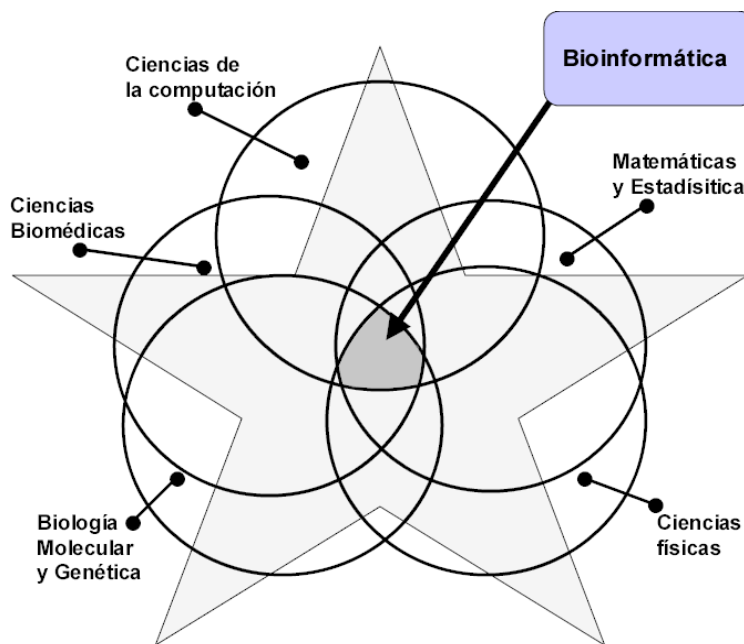


Figura 1.1. La Bioinformática como convergencia multidisciplinaria

### 1.3.1 Relación con la Informática Médica y la Informática Biomédica

Existe también confusión en la delimitación conceptual de la Bioinformática y, aunque algunos autores no son partidarios de establecer una diferenciación entre Informática Biomédica y Bioinformática, sí que está más clara la diferencia entre Informática Médica y Bioinformática. En una primera aproximación, se podría afirmar que la Bioinformática no es Informática Médica, dados los planteamientos clásicos de la última.

En la Informática Médica, existe una clara separación entre el especialista informático frente al usuario, que es el médico clínico, o el administrativo, o el gestor, u otros profesionales de la sanidad, incluso los pacientes. El informático tiene como responsabilidades el diseño, la construcción, implantación gestión y mantenimiento de las soluciones tecnológicas en los sistemas informáticos sanitarios, y también resuelve las incidencias técnicas de cualquier tipo que surgen como consecuencia de la utilización de las soluciones y los sistemas.

En cambio, el usuario se limita a la aplicación de las soluciones tecnológicas en su actividad clínica o administrativa, siendo algunas veces los propios generadores de incidencias. Esta disciplina es relativamente genérica y extensa y está enfocada más hacia los aspectos de desarrollo, innovación y producción (Van Bommel et al., 1997) (Sackman, 1997). En cambio, en la Bioinformática se da la convergencia multidisciplinar, donde al menos un científico biólogo y otro científico informático colaboran estrechamente para alcanzar un objetivo común, para lo cual necesitan un lenguaje común y, además, aprender uno del otro los respectivos conocimientos fundamentales. Esta disciplina tiene un horizonte más estrecho orientado fundamentalmente a la investigación aplicada (a veces pura) relacionada principalmente con las áreas que se muestran en la Figura 1.1.

Sin embargo, la Informática Médica también ha ido incorporando en los últimos años una importante faceta investigadora (tratamiento de imágenes, diseño de medicamentos, terapias asistidas, etc.) Además, ha adoptado con rapidez las tecnologías relacionadas con la genómica y la proteómica con el objetivo de potenciar la investigación clínica y epidemiológica aplicadas (Jones, 1998) (Roberts, 2000). Esto ha dado lugar a la aparición de una nueva disciplina, denominada Informática Biomédica, que se puede ver como la intersección entre la Informática Médica y la Bioinformática. De esta forma, la Informática Biomédica trata aspectos que no están cubiertos por las otras disciplinas como diagnóstico genético, la integración de información clínica, genética y medioambiental y el modelado, todo ello orientado a encontrar la relación entre problemas clínicos y moleculares en el marco de la salud humana. Precisamente, para el desarrollo de la Informática Biomédica en España, se ha creado recientemente la red de



investigación INBIOMED (Plataforma para la integración de datos clínicos, genéticos, epidemiológicos, ambientales e imágenes orientada a la investigación sobre patologías), financiada por el Instituto de Salud Carlos III. INBIOMED reúne a los principales grupos de investigación españoles en esta disciplina, y en la que colabora también nuestro grupo. Concretamente, la Epidemiología Genómica de las Enfermedades Cardiovasculares (ECV), se ha propuesto como modelo experimental de requisitos de usuario para conseguir la integración de los datos genéticos y ambientales en la investigación sobre patologías.

### 1.3.2 Relación con la Epidemiología Genómica

La Epidemiología Genómica es una disciplina encuadrada en el ámbito de las ciencias biomédicas (Figura 1.2), que está relacionada a su vez con la Bioinformática (Figura 1.1). Así, se puede deducir que la intersección entre Bioinformática y Epidemiología Genómica es un subconjunto del área que resulta de la intersección entre la Bioinformática y las ciencias biomédicas.

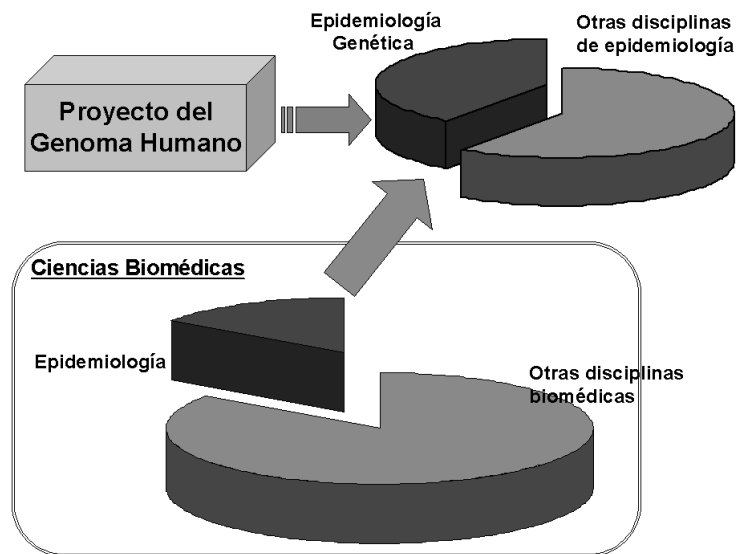


Figura 1.2. La Epidemiología Genética se adscribe en el contexto de las ciencias médicas y es una parte de la epidemiología general, pero con una contribución del Proyecto del Genoma Humano

Esto significa que no todas las soluciones bioinformáticas son imprescindibles en la Epidemiología Genómica. Por ejemplo, tendría una aplicación más tangencial la representación de estructuras proteicas en dos o tres dimensiones (2D ó 3D) para estudiar la bioquímica molecular de las proteínas, etc.

Puesto que la Informática Médica ha ido incorporando en los últimos años una importante faceta investigadora, mediante la aplicación de las tecnologías relacionadas con la genómica y la proteómica, con el objetivo de potenciar la investigación clínica y epidemiológica aplicadas (Jones, 1998) (Roberts, 2000), se ha confundido a veces con la Bioinformática. Sin embargo, en la subsección anterior se ha visto que esta diferenciación disciplinar está bastante clara. Además, la Bioinformática es fundamental en la Epidemiología Genómica para la investigación de las causas y del tratamiento de las enfermedades al permitir la integración del conocimiento genético en la etiología de las mismas.

## 1.4 Las Enfermedades Cardiovasculares

Existe un grupo de enfermedades cuyo origen es genético. O dicho en términos médicos, su etiología es de origen genético. En la Tabla 1.1 se puede ver la clasificación de las mismas con algún ejemplo asociado. Por otra parte, otra característica de las enfermedades de etiología genética es la penetrancia. La penetrancia se entiende, en términos precisos, como la proporción de individuos con un genotipo específico que lo manifiestan al nivel fenotípico. Siendo el genotipo la estructura genética específica de un organismo y el fenotipo los rasgos observables resultantes de la interacción entre el genotipo y los factores ambientales.

**Tabla 1.1. Etiología de las enfermedades desde un punto de vista genético**

| Categoría por origen                        | Ejemplos   |
|---|--|
| <i>Monogénica</i>                           | Errores congénitos del metabolismo: fenilcetonuria y galactosemia.       |
| <i>Cromosómica</i>                          | Síndrome de Down y síndrome del maullido del gato.                       |
| <i>Multifactorial de Alta Heredabilidad</i> | Defectos de nacimiento aislados: defectos del tubo neural.               |
| <i>Multifactorial de Baja Heredabilidad</i> | Enfermedades coronarias y enfermedades pulmonares obstructivas crónicas. |

Sin embargo, se puede entender la penetrancia desde un punto de vista más intuitivo según lo muestra el modelo simulado de la Figura 1.3. En dicha Figura se representan algunos de los casos de enfermedades genéticas en relación con unas

hipotéticas unidades de riesgo de padecer la enfermedad (URE  $\equiv$  RUD: *Risk Unit of Disease*) en función del tiempo en años. Suponemos previamente que cada una de las curvas representa la media de individuos afectados por cada una de las enfermedades. En primer lugar, tenemos la enfermedad de tipo A, que no tiene ninguna interacción ambiental y se manifiesta en el instante cero, manteniéndose en el mismo nivel de RUD durante toda la vida del individuo. En segundo lugar, las enfermedades del tipo B empiezan a manifestarse a cierta edad, y se incrementa su expresión a medida que transcurren los años. En tercer lugar, las enfermedades del tipo C empiezan a manifestarse a una edad determinada, debido a la interacción con algún factor ambiental, y se incrementa su efecto a medida que transcurren los años. Finalmente, en las enfermedades de tipo D, ocurre lo mismo inicialmente pero, en otro momento distinto y posterior, la interacción con otro factor ambiental distinto puede provocar un efecto positivo y disminuir el nivel de RUD.

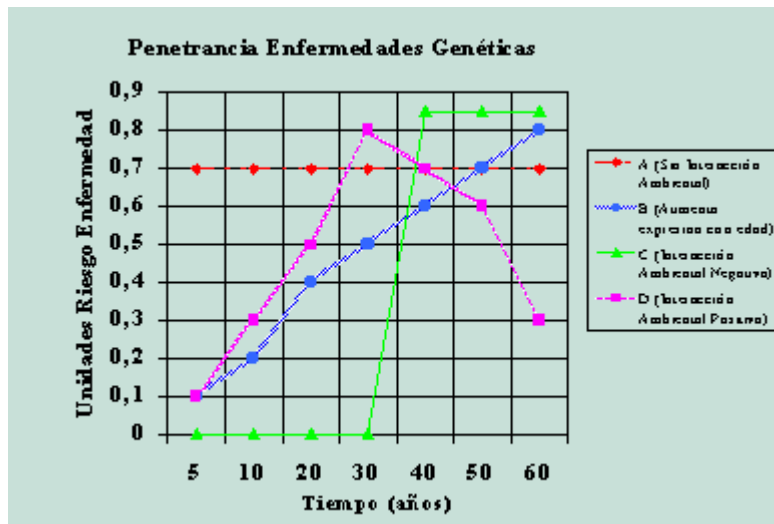


Figura 1.3. La penetrancia en las enfermedades genéticas

Por otra parte, existen enfermedades que pueden combinar varios de estos comportamientos. No se han representado en la Figura 1.3 por motivos de claridad.

Las enfermedades cardiovasculares, son enfermedades cromosómicas multifactoriales de baja heredabilidad. Es decir, por una parte, existen varios genes participando con distintos grados de influencia en la aparición de cardiopatías y aterosclerosis (Ordovás et al., 1995) (Figura 1.4).

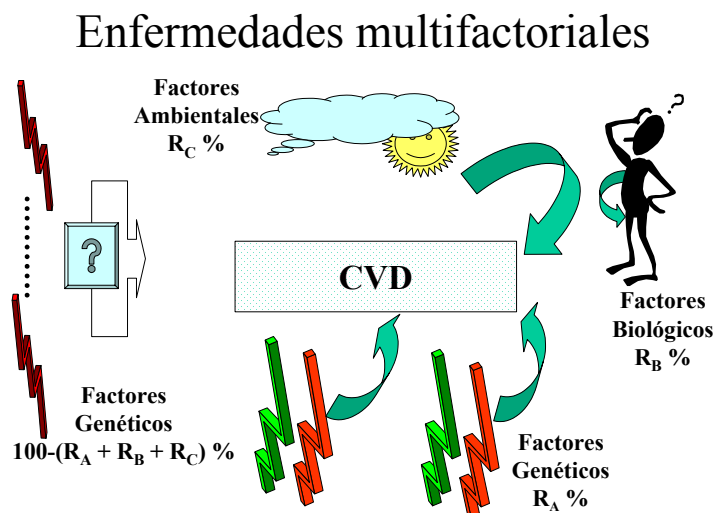


Figura 1.4. Las interacciones de varios factores en las enfermedades cardiovasculares

Y por otra parte, este tipo de enfermedades no se manifiestan forzosamente en todos los descendientes familiares, sino que se ha podido comprobar la existencia de factores ambientales que determinan la prevalencia y la penetrancia de dichas enfermedades (Khoury et al., 1993), tales como la ingesta de grasas saturadas, de colesterol y de alcohol en la dieta diaria, más el riesgo del tabaco, y los hábitos de conducta (Ordovás et al., 1995). Hay también otros factores dependientes de la morfología del propio individuo, como el índice de masa corporal (IMC) y el índice cintura/cadera, principalmente.

## 1.5 La Epidemiología y las enfermedades cardiovasculares

En 1988 se creó la *Human Genome Organisation* (HUGO), para coordinar mundialmente todos los esfuerzos individuales en investigación genómica, contando con 23 países fundadores. Así se marcó uno de los hitos importantes respecto del verdadero inicio de la “Era de la Revolución Genómica” (McKusic, 1989), cuya primera etapa culmina con la publicación del borrador del genoma humano en 2001 (Science’s News Staff, 2001), que se ha completado formalmente en 2003 (Collins et al., 2003 –publicado simultáneamente en *Nature* y *Science* pero como dos artículos distintos). Los resultados parciales que se iban obteniendo, posibilitaron el surgimiento de una nueva disciplina con la misión de comprender la influencia de los factores genéticos en la ocurrencia de enfermedades en las poblaciones humanas: era la Epidemiología Genética (Khoury, 1997). Los primeros

estudios llevados a cabo en el marco de la Epidemiología Genética estaban basados en la observación de la agregación familiar para ciertas enfermedades. Así, se estudiaban varias decenas de familias con varios de sus miembros afectados por una cierta enfermedad hereditaria (por ejemplo, cáncer de mama familiar) (Khoury et al., 1993), y mediante modelos matemáticos complejos, se intentaba estudiar la heredabilidad de la enfermedad, avanzando posteriormente hacia la localización concreta del posible gen implicado en lugar cromosómico mediante los populares estudios de ligamiento o *linkage* con los cálculos de la *LOD score*. Estos estudios familiares se complementaban también con estudios en gemelos, homocigóticos y dicigóticos; en niños adoptados o en pares de hermanos afectados (Khoury et al., 1993).

En esta sección se procederá en primer lugar a caracterizar la Epidemiología Genómica. Después se describen brevemente las enfermedades cardiovasculares, que son el centro del estudio epidemiológico en este trabajo. A continuación se procede a una caracterización general del riesgo cardiovascular y de los factores que lo determinan. Finalmente, se explica el procedimiento clásico para calcular el riesgo cardiovascular.

### **1.5.1 La Epidemiología Genómica**

La Epidemiología Genética potencia su desarrollo con los resultados del Proyecto del Genoma Humano (HGP: *Human Genome Project*), debido a la posibilidad de estudiar directamente las variaciones en el ADN de cada uno de los individuos. Esta detección molecular se ve reflejada en la denominación de la disciplina incorporando también el calificativo de “Molecular”. Aunque durante toda la década de los 90 (Khoury, 1997), se ha venido utilizando la denominación de Epidemiología Genética y Molecular, la necesidad de detectar e integrar las denominadas interacciones gen-ambiente en la investigación etiológica de las enfermedades dio lugar a que, a finales de los años 90, se prefiriera la denominación de Epidemiología Genómica, que recoge mejor los fines de la misma. Por lo tanto, una definición ya estable de la Epidemiología Genómica la presenta como la disciplina que estudia la contribución relativa y posible interacción de los factores genéticos y ambientales en la etiología, distribución y prevención de la enfermedad en los humanos. Uno de los axiomas principales de esta disciplina es que las enfermedades humanas están causadas por la interacción entre factores genéticos (variaciones en el genoma) y ambientales (estilos de vida). Y por tanto, dicha interacción debe tenerse en cuenta explícitamente en el diseño y análisis de los estudios epidemiológicos.

### 1.5.2 Las ECV como ejemplo de interacción entre genes y ambiente

Las enfermedades cardiovasculares (abreviadamente ECV) constituyen actualmente la primera causa de mortalidad a escala mundial (Figura 1.5). En España, estas enfermedades superan al cáncer como primera causa de defunción. Esto representa un importante problema de salud pública que precisa intensificar esfuerzos, no sólo en su tratamiento, sino fundamentalmente en la investigación de las causas (factores) de riesgo que originan las ECV para poder actuar sobre ellas de manera preventiva.

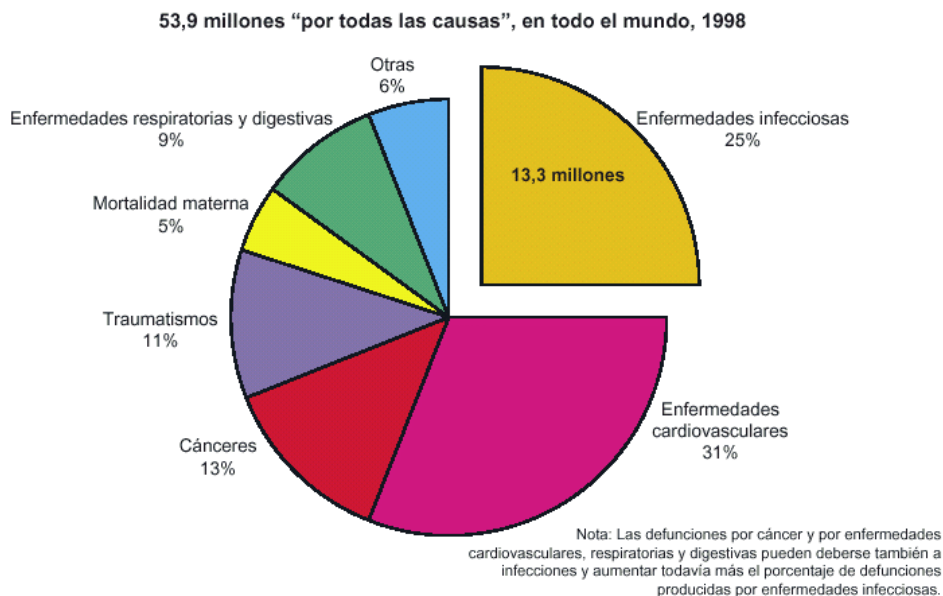


Figura 1.5. Estadísticas de mortalidad de las EVC en todo el mundo según datos de las OMS en 1998

Las ECV son enfermedades complejas y multifactoriales. Con excepción de algunas cardiopatías congénitas y otros síndromes poco prevalentes con un gran componente genético, la mayoría de ECV son el resultado de la interacción entre varios factores genéticos y ambientales que determina el riesgo de padecer un incidente cardiovascular. Estos factores o causas se denominan formalmente factores de riesgo expresando así su componente probabilístico más bien que determinista. De manera clásica los factores de riesgo, también denominados determinantes de las enfermedades, se han clasificado en cuatro grandes grupos: medio ambiente, estilo de vida, biología humana y asistencia sanitaria. Las ECV

también responden a este esquema general (Figura 1.6). El medio ambiente comprende los factores ambientales poco modificables por el individuo, entre ellos se encuentran las radiaciones ionizantes, los contaminantes atmosféricos, las bacterias, los virus, la temperatura, la humedad, etc. El estilo de vida comprende los comportamientos que adopta el individuo en su manera de vivir (consumo de tabaco, alcohol, ejercicio, consumo de alimentos, drogas, etc.). La biología humana hace referencia a la dotación genética de los individuos, mientras que la asistencia sanitaria comprende todos los cuidados de salud que proporciona el sistema sanitario. Simplificando este esquema clásico, la Epidemiología Genómica, considera sólo dos factores de riesgo: los genéticos (genoma) y los ambientales (factores no genómicos). Dentro de los factores ambientales se condensan los tres anteriores: medioambiente, estilo de vida y asistencia sanitaria.

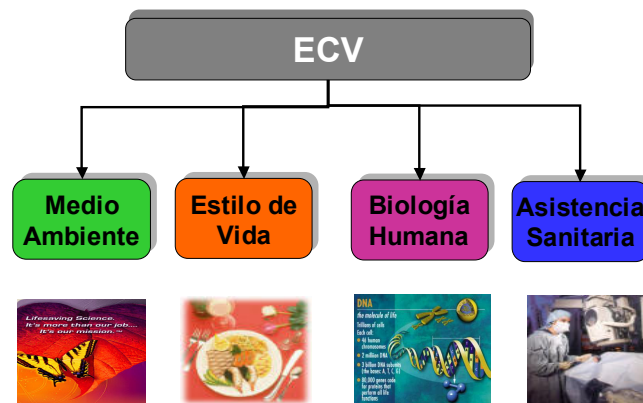


Figura 1.6. Los determinantes de la salud en ECV (adaptado de la propuesta general de Laframboise (1973))

En la Figura 1.7, se resumen el enfoque actual de la Epidemiología Genómica para la clasificación de los factores de riesgo: los factores de riesgo genéticos y los ambientales. Concretamente, para las ECV, los principales factores genéticos pueden estudiarse mediante la detección y caracterización de las principales variaciones en genes candidatos (actualmente se estima en miles de genes los implicados en las ECV), mientras que los principales factores ambientales en estas enfermedades serían la dieta, tabaco, alcohol, actividad física, estrés, etc.

La Bioinformática va a permitir la integración del estudio de ambos tipos de factores de riesgo con los recursos tecnológicos creados al efecto. Por tanto, para poder realizar la adquisición y el análisis de los datos genéticos y ambientales, es necesario seguir intensificando los esfuerzos ya que en la actualidad existen herramientas que tratan de forma parcial dicha integración.



Figura 1.7. El esquema del que parte la Epidemiología Genómica en el estudio de las ECV

Por ello están plenamente justificados los esfuerzos de la red de investigación temática INBIOMED, en la que parcialmente se han desarrollado los trabajos de investigación de esta tesis, que tiene por objetivo la obtención de soluciones integradas para el tratamiento de información relacionada con las enfermedades a distintos niveles, desde el molecular, hasta los estudios poblacionales.

### 1.5.3 Genes implicados como factores de riesgo en las ECV

Se han descrito miles de genes posiblemente implicados en el riesgo de ECV actuando a distintos niveles: metabolismo lipídico, tensión arterial, estrés oxidativo, intolerancia a la glucosa, obesidad, inflamación, hiperhomocisteinemia, etc. En esta tesis se han elegido, como principales genes candidatos, los más relevantes en el metabolismo lipídico (CETP, APOE, APOA1, LIPC y SR-BI), incluyendo también el análisis de un gen menos estudiado y posiblemente implicado en la etiología de la obesidad y resistencia a la insulina, el gen PLIN.

A continuación se presenta una breve descripción de cada uno de estos genes así como de los polimorfismos posteriormente analizados.



### 1.5.3.1 El gen CETP

La proteína de transferencia del éster de la colessterina (en inglés *cholesteryl ester transfer protein*, abreviado como CETP) facilita el intercambio de los triglicéridos y los ésteres de la colessterina entre partículas de lipoproteínas. En los humanos, el ARN mensajero (en inglés mRNA) de la CETP codifica un polipéptido de 53.000 MR que es n-glicosilado en los cuatro extremos, permitiendo alcanzar la forma madura de la CETP de 74.000 MR (Drayna et al., 1987). La CETP se expresa primariamente en el hígado, el bazo y el tejido adiposo, y se han detectado niveles bajos en el intestino bajo, la glándula suprarrenal, el corazón, los riñones y en el músculo esquelético.

El gen CETP comprende 16 exones y se ha localizado en el brazo largo del cromosoma 16 (16q21) (Figura 1.8) adyacente al gen LCAT. Se han identificado varias mutaciones en el locus de CETP que provocan la ausencia de masa o actividad detectable de la CETP (Yamashita et al., 1997). Estas mutaciones son comunes en las poblaciones japonesas (Inazu et al., 1990) (Koizumi et al., 1991) (Takegoshi et al., 1992) (Inazu et al., 1994) aunque también se han informado en sujetos caucasianos (Hill et al., 1997) (Tamminen et al., 1996).

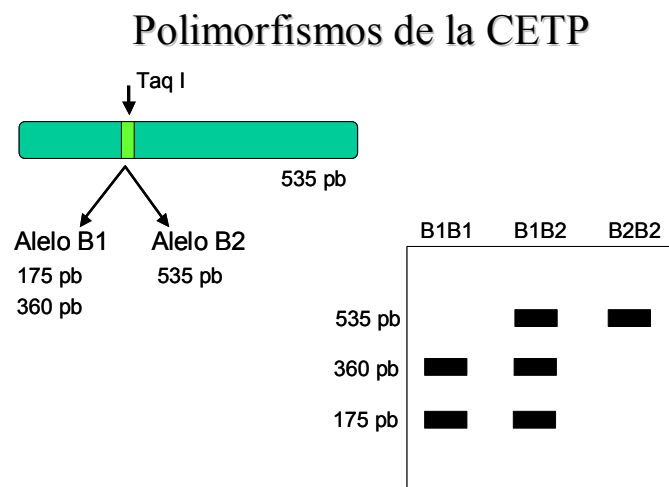


Figura 1.8. Estructura esquemática del gen CETP con sus polimorfismos

La deficiencia de CETP se asocia con la hiperalfalipoproteinemia cuya causa principal es el incremento de lipoproteínas de alta densidad y partículas de gran tamaño enriquecidas con el éster de la colessterina (abreviadas técnicamente como HDL). Por otra parte, las lipoproteínas ricas en triglicéridos y las lipoproteínas de

baja densidad (abreviadas técnicamente como LDL) son mas pequeñas y ricas en triglicéridos, lo cual muestra su papel en el intercambio neutro de lípidos (Yanashita et al., 1997).

Se han dado a conocer varios polimorfismos de longitud de fragmento de restricción (en inglés *restriction fragment length polymorphisms*, abreviado como RFLP) comunes en el locus del gen CETP (Drayna, Lawn et al., 1987) (Freeman et al. 1989) (Zuliani et al., 1990). El RFLP más estudiado hasta la fecha ha sido el TaqIB, que se ha presentado como un cambio de base silente que afecta al nucleótido número 277 en el primer intrón del gen (Drayna, Lawn et al., 1987) (Figura 1.8). El alelo B2 (ausencia del sitio de restricción de TaqIB) en este sitio polimórfico se ha asociado en sujetos normolipémicos con los incrementos de los niveles de HDL-C y el decremento de la actividad y los niveles de CETP (Kondo et al., 1989) (Freeman et al., 1994) (Hannuksela et al., 1994) (Kuivenhoven et al., 1997), mostrándose por tanto como una forma suave de deficiencia de CETP. Se ha sugerido que esta asociación pueda ser específica de la población (Tenkanen et al., 1991) (Mitchell et al., 1994) y que esté altamente influida por los factores ambientales tales como el consumo de alcohol y tabaco (Hannuksela et al., 1994) (Fumeron et al., 1995) (Kauma et al., 1996). Además, Kuivenhoven et al. (Kuivenhoven et al., 1998) han presentado una interacción entre el genotipo TaqIB y la progresión de las ECV bajo terapia. Estas observaciones podrían ser de relevancia significativa puesto que los niveles bajos de HDL plasmático se asocian con un incremento del riesgo de padecer enfermedad arterial coronaria (Gordon et al., 1989) (Gordon, et al., 1989 ). También la evidencia clínica sugiere que un incremento del 1% en los niveles de HDL-C se asocia con una reducción en la morbilidad y mortalidad cardiovascular de 2 a 3% (Manninen et al., 1988). Por lo tanto, el CETP podría tener un papel relevante en la aterogénesis a través de sus efectos en el metabolismo de HDL.

Entonces, se planteó un estudio con el propósito de determinar la frecuencia, la expresión fenotípica y la modulación potencial del riesgo de ECV en la población general por parte del RFLP de TaqIB en el primer intrón del gen CETP. En particular, se ha estudiado la variabilidad interindividual en los niveles de lípidos, en los perfiles de subclases de lipoproteínas y en el riesgo cardiovascular asociado con este polimorfismo de CETP entre los participantes del *Framingham Offspring Study*.

### 1.5.3.2 El gen APOE

El colesterol plasmático es uno de los principales factores de riesgo cardiovascular (Barter, 1994) (Sytkowski et al., 1990). Dado que las enfermedades cardiovasculares (ECV) son la primera causa de muerte en todo el mundo

civilizado, un buen conocimiento de los factores que determinan la colesterolemia, constituye una información muy relevante desde el punto de vista de la Salud Pública. Las concentraciones plasmáticas de colesterol están a su vez determinadas por factores genéticos y ambientales (Bodurtha et al. 1991) (Hegele, 1992). Aunque los factores ambientales, han sido ampliamente investigados en los estudios epidemiológicos (Bijnen et al., 1996) (Fujishima et al., 1992), la contribución de los determinantes genéticos a la colesterolemia en la población general está poco caracterizada (Ellsworth et al., 1999). En la última década se han descrito varias mutaciones asociadas al riesgo individual de hipercolesterolemia y de enfermedad cardiovascular (Talmud et al., 1991) (Bétard et al., 1996) (Connelly et al., 1990) (Reina et al., 1992), sin embargo, éstas se presentan con una frecuencia muy baja, y tienen por tanto un efecto muy limitado sobre el riesgo poblacional. Existen por otra parte una serie de genes que, si bien su efecto individual sobre la colesterolemia parece ser ligero, dado que sus polimorfismos son muy frecuentes, su impacto poblacional sería muy elevado (Ellsworth et al., 1999) (Kamboh et al., 1995). El más estudiado de estos genes, es el gen de la apolipoproteína E (apoE). Sus variaciones se han asociado con distintas concentraciones de lípidos plasmáticos y con un mayor riesgo de enfermedad cardiovascular (Davignon et al., 1988) (Hixson, 1991) (Kamboh, Aston et al., 1995) (Myers et al., 1996) (Wilson et al., 1994) (Lahoz et al., 1997). El gen estructural de la apoE (gen APOE) está localizado en el cromosoma 19 (19q13.2), siendo sus variantes más comunes las conocidas como apoE2 (alelo  $\epsilon$ 2), apoE3 (alelo  $\epsilon$ 3) y apoE4 (alelo  $\epsilon$ 4) (Figura 1.9).

### Polimorfismos en el gen de la apolipoproteína E

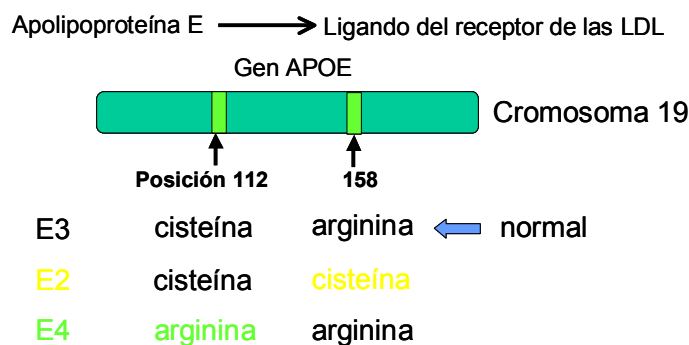


Figura 1.9. Estructura esquemática del gen APOE y sus polimorfismos

Estas variantes difieren una de otra según la presencia de arginina o de cisteína en las posiciones 112 y 158 (Davignon et al., 1988) (Wilson et al., 1994) (Breslow

et al., 1982) (Cumming et al., 1984). El alelo  $\epsilon 4$  se asocia, en general, a niveles más altos de colesterol total y de colesterol ligado a lipoproteínas de baja densidad (colesterol-LDL), mientras que el alelo  $\epsilon 2$  lo haría a concentraciones más bajas (Davignon et al., 1988) (Sing et al., 1985) (Utermann, 1987) (Davignon et al., 1999) (Tiret et al., 1994). Sin embargo, en algunos estudios no se han encontrado estas asociaciones (Aguilar et al., 1999) (Deiana et al., 1998) (Pérez-Estrada et al., 1990).

A escala mundial, la prevalencia de cada uno de los alelos difiere según el origen étnico-geográfico de la muestra (Hallman et al., 1991) (Gerdes et al., 1992). En Europa, la prevalencia del alelo  $\epsilon 4$  disminuye de Norte a Sur (Figura 1.10), al igual que lo hacen las tasas de mortalidad por ECV (Tiret et al., 1994) (Lucotte et al., 1997).

### Apolipoproteína E

Gradiente Norte-Sur en la frecuencia del alelo E4

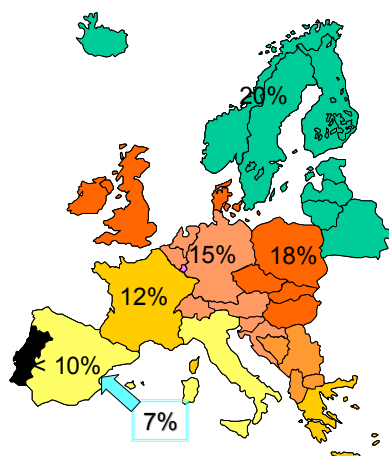


Figura 1.10. Disminución de Norte a Sur de la prevalencia del alelo  $\epsilon 4$  en Europa

En España, dos estudios realizados en población laboral (Muros et al., 1996) (Gómez-Coronado et al., 1999), han confirmado la baja prevalencia para el alelo  $\epsilon 4$  en nuestro país. Sin embargo, los resultados obtenidos en la asociación de este alelo con los niveles de lípidos plasmáticos han sido discordantes. A ello puede contribuir tanto la posible influencia de factores ambientales que modulen la expresión genotípica, como los sesgos en la estimación de la asociación resultado de los diseños transversales empleados. Por ello, el objetivo de este estudio es

estimar la asociación entre las variantes genéticas comunes en el gen de la apoE y el riesgo de hipercolesterolemia moderada en población laboral mediterránea, a través de un diseño de casos y controles, teniendo en cuenta la influencia de otros factores como el índice de masa corporal, el consumo de tabaco y de alcohol, el ejercicio físico y el nivel de estudios.

La variación genética del locus del gen Apolipoproteína E (abreviado técnicamente como APOE) es un importante determinante de las concentraciones plasmáticas del colesterol de baja densidad (en inglés *low-density lipoprotein cholesterol*, abreviado como LDL-C). En general, los portadores del alelo APOE2 ( $\epsilon$ 2) tienen menores concentraciones de LDL-C, y los portadores del alelo APOE4 ( $\epsilon$ 4) tienen mayores concentraciones de LDL-C, cuando se comparan con las de los homocigotos para el alelo APOE3 ( $\epsilon$ 3) (Eichner et al., 1990) (Boerwinkle et al., 1988) (Wilson et al., 1994) (Braeckman et al., 1996). Sin embargo, estos efectos se pueden modular por medio de factores como la edad, el sexo, raza y condiciones ambientales (Jarvik et al., 1994) (Pablos-Mendez et al., 1997) (Aguilar et al., 1999).

Se han presentado varias interacciones gen  $\times$  dieta (grasa y colesterol) que se asocian con la variabilidad en la respuesta del LDL-C a la intervención dietética (Lopez-Miranda et al., 1994) (Lehtimaki et al., 1997). Se han explorado en menor medida otras interacciones con factores de comportamiento, tal como el consumo de tabaco y alcohol. En cambio, se ha estudiado con extensión el efecto del consumo de alcohol en los lípidos plasmáticos y en las ECV. Los estudios epidemiológicos de casos y controles y de cohorte han descrito asociaciones en forma de “J” o de “U” entre el consumo de alcohol y las ECV (Kiechl et al., 1998) (McElduff et al., 1997), así como con todas las causas de mortalidad (Gronbaek et al., 1998). Los hombres y mujeres que beben una o dos veces por día (ingesta de alcohol moderada) muestran el menor riesgo de ECV (Renaud et al., 1998) (Keil et al., 1997) (Rimm et al., 1996). Los estudios experimentales y observacionales han informado que el consumo de alcohol moderado se asocia con incrementos en la concentración del HDL-C (Glueck et al., 1980) (Gaziano et al., 1993). Esta información se ha resumido en un metaanálisis que concluye que la ingesta de alcohol que concluye que la ingesta de alcohol está relacionada causalmente con bajos riesgos de ECV por medio de cambios en factores lipídicos y hemostáticos (Rimm et al., 1999). Sin embargo, el efecto del consumo de alcohol en LDL-C no está claro y no se ha presentado en el metaanálisis. En general, se han presentado asociaciones directas (Vasisht et al., 1992) o inversas (Langer et al., 1992) (Nakanishi et al., 1999) entre el alcohol y las concentraciones de LDL-C, así como la ausencia de efecto (McConnell et al., 1997) (Rakic et al., 1998). Como hecho a destacar, Hein et al. (1996) examinaron la conexión entre el alcohol, el LDL-C y el riesgo de ECV en un estudio de cohorte, y mostraron que el efecto del alcohol en el

riesgo de ECV depende de las concentraciones de LDL-C. La inconsistencia de los resultados presentados por varios estudios sugiere que otros factores, incluyendo la variabilidad genética, pueden modificar la asociación entre las concentraciones de LDL-L y el consumo de alcohol.

Entonces, se planteó un estudio con el propósito de examinar si la variación del locus del gen APOE modula la asociación entre la toma de alcohol y las concentraciones de LDL-C entre los participantes del *Framingham Offspring Study*.

### 1.5.3.3 El gen APOA1

Una de las recomendaciones de la Conferencia Científica en Nutrición Preventiva de 1999 (Deckelbaum et al., 1999), así como de varias organizaciones de salud (Ordovas et al., 1999), es investigar en la interacción de nutrientes con formas polimórficas específicas de genes, que juegan un papel en la causa de enfermedades graves, con el propósito de obtener datos para formular recomendaciones dietéticas más específicas en la prevención de la enfermedad.

La apolipoproteína A-I (abreviado apoA-I) tiene una larga historia en las ciencias médicas con respecto a su papel central en el metabolismo lipídico y en las ECV (Norum et al., 1982) (Rader et al., 1994). Es la proteína principal en HDL-C, constituyendo aproximadamente entre el 70 y 80% de la masa de la proteína, y el principal activador de la enzima transferasa acil colesterol lecitina (en inglés, abreviada como LCAT) (Fielding et al., 1972). El gen apoA-I (APOA1) corresponde al brazo largo del cromosoma 11 (Figura 1.11), agrupado con los loci estructurales para el apoC-III y el apoA-IV (Karathanasis, 1985). Varios estudios en las estructuras familiares de los pacientes han presentado asociaciones entre la variación genética en este locus y los niveles de lípidos y lipoproteínas plasmáticos (Ordovas et al., 1991) (Bruckert et al., 1997).

La transición común G hacia A (G/A) en la base -75 desde el sitio inicial de transcripción del gen APOA1 también ha sido una de las más estudiadas en la población general. En 1990, este polimorfismo fue en primer lugar asociado con apoA-I y el HDL-C (Jeenah et al., 1990) en los individuos portadores del raro alelo A que presentaban los mayores niveles comparados con los portadores del tipo natural G/G. Estudios posteriores que han examinado esta asociación han mostrado resultados contradictorios. Mientras algunos de ellos (Paul-Hayase et al., 1992) (Meng et al., 1997) estaban en concordancia con Jeenah et al. (1990), otros no podían detectar ninguna asociación (Civeira et al., 1993) (Barre et al., 1994) (Akita et al., 1995). Estas disparidades se pueden explicar por la influencia de los factores ambientales que modulan el efecto del polimorfismo genético. Aunque alguno de

estos estudios han investigado la posible interacción con el consumo de tabaco (Kamboh et al., 1996) (Talmud et al., 1994), ninguno de ellos han evaluado la influencia de la dieta en una muestra de población grande.

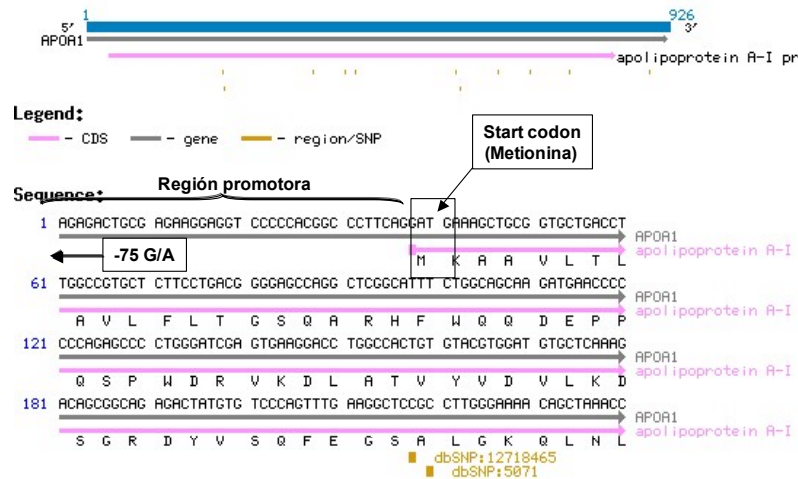


Figura 1.11. Estructura esquemática del gen APOA1  
(Fuente: GenBank NCBI Map Viewer)

En un estudio de intervención dietética, se ha informado de una respuesta diferente en lípidos plasmáticos de las variantes G/A de APOA1 dependiendo de la cantidad y el tipo de grasa consumida (Mata et al., 1998). Profundizando en el efecto de la grasa, se ha establecido claramente que los ácidos grasos, principalmente los ácidos grasos poliinsaturados (PUFA), pueden modular la expresión de muchos genes implicados en el metabolismo lipídico (Jump et al., 1997).

Considerando que el polimorfismo A/G en la posición -75 hacia arriba del sitio de iniciación de la transcripción se localiza en una región de ADN rica en pares GC, se ha propuesto que la presencia de A o G debería influir diferencialmente en la eficiencia de la transcripción del gen APOA1. Sin embargo, también se han publicado series de observaciones contradictorias sobre este efecto (Jump et al., 1997) (Smith et al., 1992), y los autores sugieren que los alelos A/G pueden responder diferentemente a los moduladores de la transcripción.

Entonces, se planteó un estudio con el propósito de investigar sobre la posible interacción de la ingesta de grasa en dieta, con los efectos del polimorfismo -75 bp G/A en el promotor del gen APOA1, sobre los niveles de HDL-C y apoA-I entre

los participantes del *Framingham Offspring Study*, para resolver las inconsistencias entre los informes antes mencionados.

#### 1.5.3.4 El gen LIPC

Las concentraciones del colesterol plasmático de alta densidad (HDL-C) son el mayor factor de riesgo para las ECV. En general, el HDL-C se incrementa con la actividad física, el consumo de alcohol y las dietas altas en ácidos grasos saturados (abreviado SFA, o SATFAT, aunque se toma en esta tesis la abreviatura SAFA, siguiendo con el mismo patrón que PUFA y MUFA). Sin embargo, es importante reducir la ingesta de SAFA ya que el LDL-C se incrementa bastante más. Por otra parte, el tabaco y la sustitución de la grasa en dieta por carbohidratos decreta el HDL-C (Mensink et al., 1992). Estudios experimentales adecuadamente controlados han demostrado que existe heterogeneidad de los lípidos plasmáticos a la grasa en dieta y sugieren la existencia de un componente genético (Berglund et al., 1999) (Dreon et al., 1998) (Katan et al., 1997). Entonces, las interacciones gen  $\times$  nutriente que afectan al metabolismo de HDL pueden contribuir a la variabilidad interindividual sustancial sobre los efectos de las dietas bajas en grasa con los lípidos plasmáticos (Berglund et al., 1999) (Dreon et al., 1998) (Katan et al., 1997).

Desde la perspectiva de la salud pública, considerando que la dieta es la piedra de toque de la prevención y el tratamiento de las ECV, e incluyendo los factores de decremento del LDL-C, es muy importante la comprensión de lo que hay implicado en la regulación dietética del metabolismo del HDL (Mensink et al., 1992).

La lipasa hepática (HL) es una enzima lipolítica que hidroliza triglicéridos (TG) y fosfolípidos en lipoproteínas plasmáticas (Thuren, 2000) y se ha sugerido que pudiera ser un nuevo factor de riesgo (Dugi et al., 2001). La HL también juega un papel como un ligando intermedio para el enlace y la absorción de las lipoproteínas vía proteoglicanos y/o rutas de receptores (Lambert et al., 2001). Por tanto, las funciones lipolíticas y no lipolíticas de la HL son importantes para el metabolismo de HDL y, finalmente, para su asociación con las ECV (Dugi et al., 2000). La sobreexpresión de HL disminuye las concentraciones de HDL-C, mientras que la deficiencia de HL las incrementa (Santamarina et al., 1998).

Se ha descrito una sustitución común de C a T en la posición -514 bp en la región promotora del gen de la HL, denominado LIPC (Figura 1.12). El alelo T se asocia con la disminución de la actividad de la HL plasmática, con el incremento de las concentraciones de HDL-C (Guerra et al., 1997) (Van't Hooft et al., 2000) (Vega et al., 1998), y con las subfracciones grandes de HDL (Couture et al., 2000). Aunque es conocido que la dieta es un factor importante que influye sobre las



concentraciones de HDL-C y las subclases de HDL, los estudios previos no evaluaron la interacción potencial gen  $\times$  dieta entre el polimorfismo  $-514C/T$  y la grasa dietética.

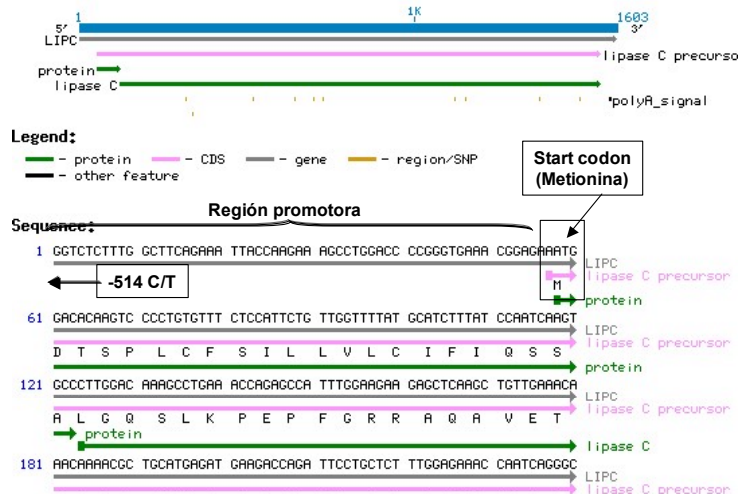


Figura 1.12. Estructura esquemática del gen LIPC (Fuente: GenBank NCBI Map Viewer)

Entonces, se planteó un estudio con el propósito de investigar cómo la ingesta de grasa en dieta, centrándose en los ácidos grasos específicos y los orígenes de la grasa, interactúa con el polimorfismo de LIPC para determinar las concentraciones y el tamaño de partícula de HDL-C entre los participantes del *Framingham Offspring Study*.

### 1.5.3.5 El gen SR-BI

En los sujetos diabéticos, una de las más comunes disfunciones lipídicas es la disminución del colesterol ligado a la lipoproteína plasmática de alta densidad (HDL-C) (Howard et al., 1998) (Taskinen, 2002) (Franceschini, 2001). Además, las concentraciones bajas de HDL-C se han definido como un aspecto significativo en el síndrome de la resistencia a la insulina, una condición metabólica común ligada a un riesgo cardiovascular alto, que se ha estimado que está presente en el 25% de la población estadounidense (Erkelens, 2001) (Ford et al., 2002).

Se cree que el HDL-C ejerce su papel antiaterogénico principalmente a través del proceso de distribución de colesterol desde los tejidos periféricos de vuelta al hígado como recolección del colesterol del cuerpo, denominado transporte inverso

del colesterol (Stein et al., 1999) (Silver et al., 2000). Un avance significativo en la comprensión de este proceso fue la identificación de la proteína receptor del recolector de clase B y tipo I, abreviado técnicamente SR-BI (Figura 1.13) (Acton et al., 1996) (NOTA: Realmente, el término inglés “*scavenger*” significa “basurero” o “carroñero”, por la función específica que realiza dicha proteína y, tradicionalmente, se ha mantenido la denominación inglesa de *scavenger receptor class B type I*). La SR-BI, una glicoproteína de la membrana celular, es el primer receptor de HDL que ha sido perfectamente definido y caracterizado en estudios in vitro y en animales.

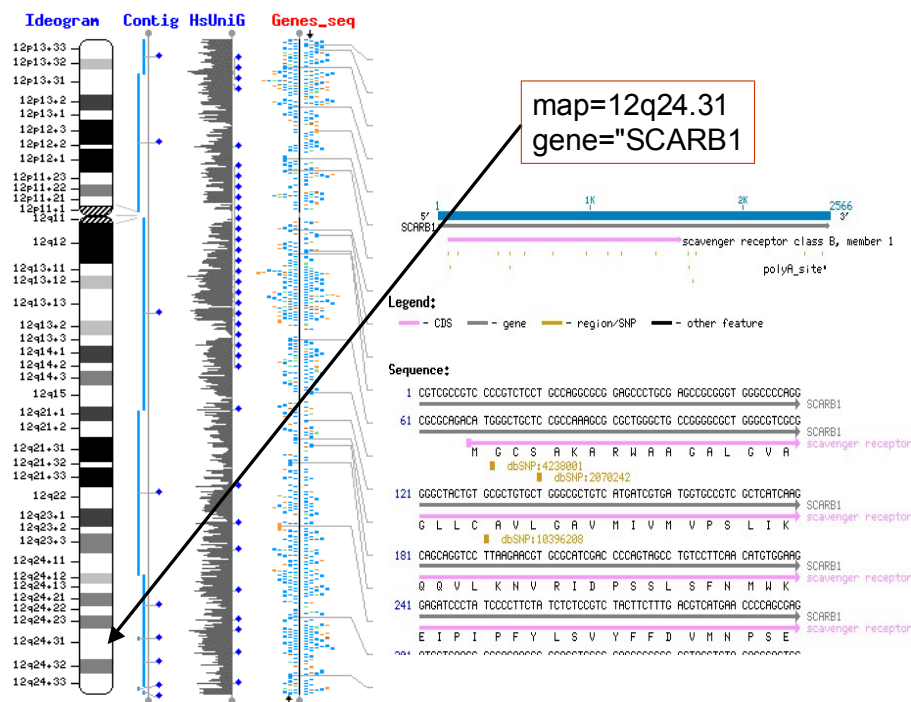


Figura 1.13. Estructura esquemática del gen SR-BI y situación en el cromosoma (Fuente: GenBank NCBI Map Viewer)

La mayor evidencia concerniente al papel de este receptor en el metabolismo de HDL proviene del ratón transgénico. La sobreexpresión del SR-BI en el ratón provoca drásticas reducciones en las concentraciones del HDL-C plasmática (Kozarsky et al., 1997) (Wang et al., 1998). Por el contrario, la subexpresión del SR-BI se ha asociado con un marcado incremento en el HDL-C plasmático, caracterizado por partículas de HDL altamente enriquecidas en colesterol y aclaración deficiente de HDL-C (Rigotti et al., 1997) (Varban et al., 1998) (Ueda et

al., 2000). Aunque el SR-BI se ha identificado como un receptor de HDL, es también un receptor multiligando (Acton et al., 1996) (Xu et al., 1997). Además del primer informe que identificaba el SR-BI como un receptor de HDL (Acton et al., 1996), algunos estudios en los animales han demostrado que el SR-BI puede influir en el LDL plasmático y en las concentraciones de apoB tal como se afirma en (Ueda et al., 1999) (Huszar et al., 2000), aunque no está claro si esto ocurre directamente o es un efecto secundario en el metabolismo del HDL.

En los humanos, la influencia del SR-BI en el HDL-C plasmático y en otros lípidos y lipoproteínas, no es muy conocido debido a la dificultad de determinar su actividad y función del SR-BI, y la ausencia de resultados en estudios epidemiológicos. En trabajos anteriores se ha caracterizado el gen SR-BI humano y se ha investigado su variación genética en población del sur de Europa (Acton et al., 1999).

Se han descrito tres variantes comunes del gen, que son todas polimorfismos de único nucleótido (en inglés *single nucleotide polymorphisms*, abreviado como SNP, y siendo este acrónimo ampliamente utilizado en la bibliografía, tanto en español como en inglés), localizados en el exón 1 (G→A), el exón 8 (C→T) y el intrón 5 (C→T) del gen SR-BI. Se han encontrado asociaciones con HDL-C y LDL-C, sugiriendo que el SR-BI puede jugar un papel en el metabolismo de ambos tipos de lipoproteínas en humanos. Además, se han detectado asociaciones con triglicéridos e IMC. En conjunto, estos datos permiten formular la hipótesis de que el SR-BI debería estar implicado en la determinación de algunas características del síndrome metabólico.

Por otra parte, el gen SR-BI está localizado en la región cromosómica 12q24 (Cao et al., 1997), que se ha relacionado repetidamente con la diabetes en varios estudios de exploración del genoma, y en los que se han identificado algunos genes candidatos (por ejemplo, MODY3 y NIDDM2) (Vaxillaire et al., 1995) (Mahtani et al., 1996) (Bektas et al., 1996) (Lindgren et al., 2002).

Por lo tanto, teniendo en cuenta las siguientes consideraciones:

1. La necesidad de más estudios epidemiológicos genéticos que analicen el impacto del SR-BI en el metabolismo humano de lipoproteínas.
2. La evidencia que implica al gen SR-BI en la variación de concentraciones de lípidos plasmáticos en las poblaciones de estudios previos.
3. Que el curso de la diabetes de tipo II se caracteriza por los cambios en la concentración de lípidos plasmáticos, modificaciones en la composición y tamaño de lipoproteínas (Taskinen, 2002) (Golf et al., 2000) (Tan et al., 1999) que pueden ser moduladores importantes de la expresión del SR-BI.

Entonces, se planteó un estudio con el propósito de estudiar lo siguiente:

4. Evaluar el efecto fenotípico en concentraciones de lípidos, lipoproteínas y apolipoproteínas, así como en los perfiles de las subclases de lipoproteínas, de las variantes genéticas comunes del gen SR-BI en la cohorte de los participantes del *Framingham Offspring Study*.
5. Examinar la asociación de la variación genética en el locus del gen SR-BI con el curso de la diabetes de tipo II en la cohorte de los participantes del *Framingham Offspring Study*.
6. Estimar si la diabetes modifica la asociación entre la variación del gen SR-BI y los rasgos fenotípicos examinados.

#### 1.5.3.6 El gen PLIN

La obesidad está caracterizada por el almacenamiento excesivo de triacilglicéridos (TAG o TG, que son equivalentes, aunque se emplea TG en esta tesis) en el tejido adiposo. Los mecanismos que controlan el almacenamiento y liberación de TG en gotitas de lípidos se conocen muy poco, si bien son cruciales para comprender la regulación del metabolismo de la energía y el IMC (Frayn et al., 2003) (Nielsen et al., 2003). Por otra parte, las anomalías en el almacenamiento y lipólisis de TG son manifestaciones tempranas de la resistencia a la insulina y de la diabetes de tipo II (revisado en (Lewis et al., 2002)).

Las perilipinas son fosfoproteínas que cubren específicamente las gotitas de lípidos intracelulares (Greenberg et al., 1991) (Greenberg et al., 1993) (Londos et al., 1996). Estudios experimentales en células cultivadas y en modelos animales han mostrado que estas proteínas son esenciales en la regulación de la deposición y movilización de los TG (Servetnick et al., 1995) (Londos et al., 1996) (Blanchette-Mackie et al., 1995) (Brasaemle et al., 2000) (Sztalryd et al., 2002) (Martinez-Botas et al., 2000) (Tansey et al., 2001). En los humanos, el gen de la perilipina (abreviado como PLIN) se ha asignado a la localización cromosómica 15q26.1 (Figura 1.14) (Nishiu et al., 1998), cerca de los anteriormente publicados loci de susceptibilidad para la obesidad, la diabetes y la hipertrigliceridemia (Mori et al., 2002) (Duggirala et al., 2000). Este gen genera cuatro productos distintos, Perilipina A, B, C y D, debido a las diferencias en el mecanismo de splicing.

La perilipina A es el isomorfo más común en el tejido adiposo humano (Greenberg et al., 1993) (Souza et al., 1998) (Garcia et al., 2003). La perilipina es una diana de la proteína kinasa A (PKA) y la perilipina no fosforilada puede actuar como barrera a la lipólisis de los TG, con la mediación de la lipasa sensitiva a hormonas (HSL), en gotitas de lípidos.

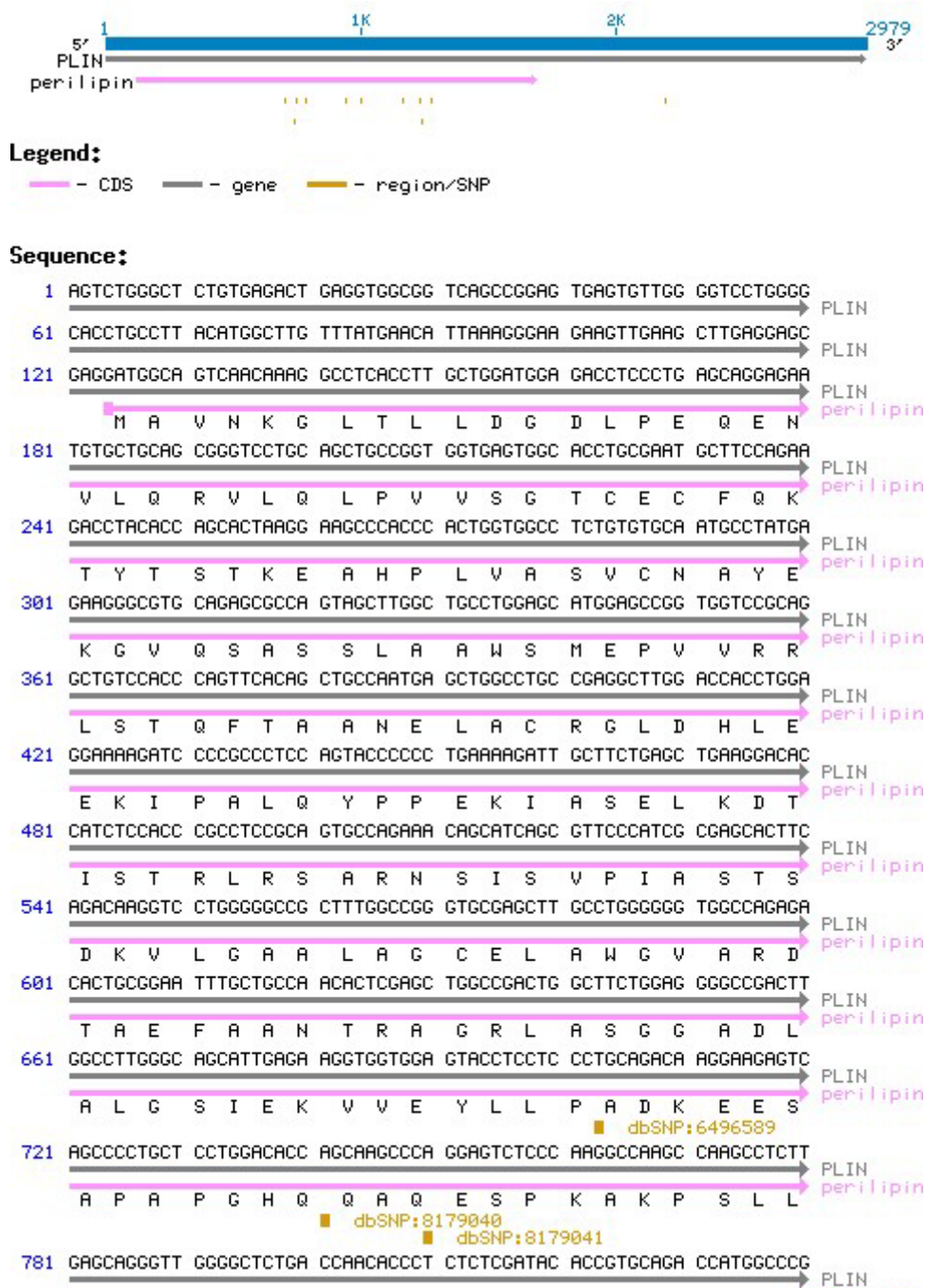


Figura 1.14. Estructura esquemática del gen PLIN (Fuente: GenBank NCBI Map Viewer)

Sin embargo, siguiendo la fosforilación, la perilipina puede facilitar la acción de la HSL (Sztalryd et al., 2002) (Tansey et al., 2003). Por tanto, la perilipina A funciona para aumentar el almacenamiento celular de TG por medio de la disminución de la tasa de hidrólisis de TG y desempeña un papel adicional en el control de la liberación de TG en el momento en que se necesita. Adicionalmente, la regulación de este proceso es clave en la homeostasis del metabolismo energético y su alteración puede contribuir a la obesidad y la disfunción del metabolismo lipídico.

De acuerdo con los estudios funcionales en células cultivadas, los datos de modelos animales presentados por dos grupos investigadores en PLIN del ratón muestran que la ausencia de perilipina provoca delgadez, incremento de la lipólisis basal, resistencia a la obesidad inducida por la dieta y obesidad inversa en los ratones *Lepr* (db/db) (un modelo de ratón experimental) (Martinez-Botas et al., 2000) (Tansey et al., 2001). Todo ello sugiere que el gen PLIN puede tener también un papel principal en la etiología de la obesidad en humanos. Sin embargo, hasta la fecha no existen estudios epidemiológicos amplios que hayan examinado esta asociación.

Por lo tanto, puesto que la evidencia actual abunda en la noción de que las variaciones en la actividad o en las concentraciones de perilipina están asociadas con la deficiente regulación de la lipólisis adiposa, se formula la hipótesis de que las variantes genéticas del gen PLIN, que podrían afectar a la expresión génica, el splicing de ARN mensajero, o la actividad de la proteína perilipina en humanos, se podría asociar con la variabilidad en las medidas antropométricas y las concentraciones de lípidos plasmáticos, estando ambos aspectos implicados en el riesgo del síndrome metabólico.

Entonces, se planteó un estudio con el propósito de examinar la asociación entre los polimorfismos comunes del gen PLIN y los fenotipos de obesidad en una amplia muestra de sujetos caucásicos extraídos de la población general así como de pacientes con obesidad severa. Este enfoque basado en la utilización de la combinación de población general y pacientes de hospital permite cubrir un amplio rango de fenotipos de obesidad y prevenir las limitaciones de diseños experimentales por separado (Little et al., 2002).

## 1.6 La simbiosis entre Bioinformática y Epidemiología Cardiovascular

En esta sección se describen el marco de trabajo principal de esta tesis, el Framingham Study, los conceptos y mecanismos para la definición y cálculo del riesgo cardiovascular, así como la importancia de la Bioinformática para el análisis genético. Finalmente, se exponen los principios para establecer un marco de integración entre la Bioinformática y la Epidemiología Genómica.

### 1.6.1 El Framingham Heart Study

El *Framingham Heart Study* es un estudio epidemiológico, iniciado en 1947, con el objeto de conocer los factores de riesgo implicados en las ECV. Gracias a él se han identificado muchos factores ambientales como determinantes de las ECV. Actualmente, dicho Estudio se ha centrado en la búsqueda de genes responsables de las mismas, así como en conocer la posible interacción genes-ambiente (Ordovás et al., 1995).

#### 1.6.1.1 Orígenes y justificación

En 1948, bajo la dirección del *National Heart Institute* (conocido en la actualidad como el *National Heart, Lung, and Blood Institute*, NHLBI) se dio comienzo a un proyecto destinado a investigar los orígenes de la enfermedad cardiovascular, cuyas tasas de prevalencia e incidencia habían venido incrementándose en los Estados Unidos desde comienzos del siglo XX hasta constituir una verdadera epidemia. Esto llegó a tal grado que, a finales de los años 40, las enfermedades del corazón se habían convertido en la primera causa de muerte de los norteamericanos. El objetivo principal de este proyecto, denominado el *Framingham Heart Study*, fue identificar las características individuales que contribuían a la aparición posterior de enfermedades cardiovasculares, mediante la estrategia de seguir su desarrollo a largo plazo en un grupo grande de personas que hasta el momento no hubieran tenido infarto ni eventos cerebrovasculares y que tampoco hubieran presentado síntomas de enfermedades cardiovasculares ni cerebrovasculares.

El enfoque, aunque ya había sido aplicado a otros problemas en salud, quizás desde la época de las investigaciones sobre la transmisión del cólera, era absolutamente novedoso con respecto a las enfermedades cardiovasculares: establecer de qué manera y por qué las personas que desarrollaban enfermedades cardiovasculares y cerebrovasculares eran diferentes de aquellas que escapaban a ellas.

### 1.6.1.2 Sujetos de estudio

Los investigadores reunieron una cohorte de 5209 hombres y mujeres de 30 a 62 años de edad, quienes aceptaron participar en el estudio de manera completamente voluntaria y gratuita. Estos participantes eran residentes en el municipio de Framingham, situado en la costa este de los Estados Unidos. En el estado de Massachussets. Según Daniel Levy, director del Framingham, se eligió la ciudad de Framingham, con la que se denominó al estudio porque, como estaba situada a unos 30 kilómetros de Boston, los participantes tenían acceso a uno de los centros médicos más importantes de Estados Unidos cuando se inició el estudio y porque los individuos también participaban en un estudio financiado por el Gobierno sobre la tuberculosis (Figura 1.15).

## Estudio Framingham

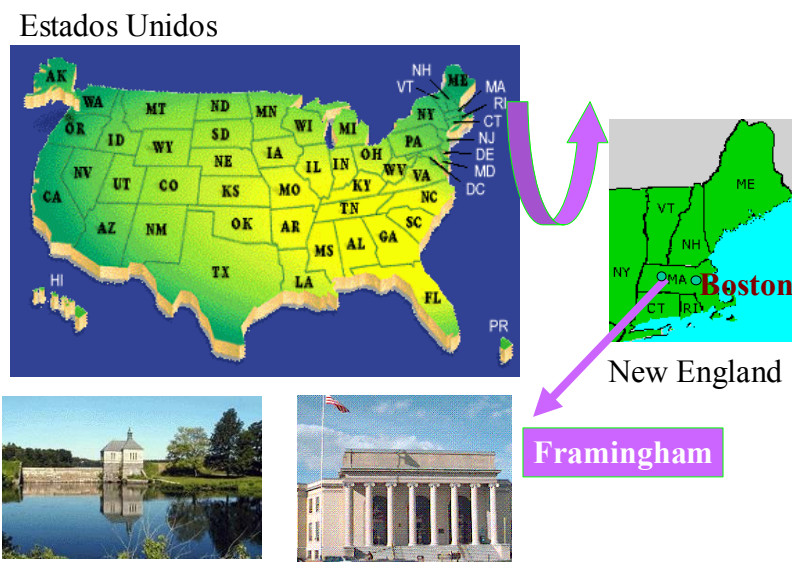


Figura 1.15. Localización geográfica del Estudio Framingham

Cada dos años los participantes en la cohorte son sometidos a evaluación integral: historia clínica detallada, examen físico, examen oftalmológico, electrocardiograma, radiografía de tórax, cuestionarios sobre estilo de vida, exámenes de laboratorio y, en épocas más recientes, otras pruebas de complejidad creciente tales como ecocardiograma, gammagrafía ósea, osteodensitometría, que



permiten evaluar múltiples aspectos de su estado actual de salud. A partir de 1971, el estudio reclutó una nueva cohorte, la segunda generación, de 5124 personas seleccionadas entre los hijos (y sus esposas) de los miembros originales de la cohorte original de Framingham: el *Offspring Study*. Estos son los actuales supervivientes y han constituido los sujetos de estudio de esta tesis. Las características basales de estos sujetos se presentan en la Tabla 1.2.

**Tabla 1.2. Distribución por edad y sexo de los participantes del Framingham Offspring Study al inicio del estudio (1971)**

| Edad    | GRUPOS DE EDAD |       |       |       |       |       |       | Total |
|---------|----------------|-------|-------|-------|-------|-------|-------|-------|
|         | < 10           | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-70 |       |
| Hombres | 0              | 126   | 544   | 789   | 694   | 298   | 38    | 2489  |
| Mujeres | 6              | 113   | 692   | 836   | 739   | 246   | 14    | 2646  |

Desde 1995, 500 personas de los grupos minoritarios de la comunidad de Framingham (hispanos, afroamericanos, asiáticos) vienen participando en el *Omni Study*, cuyo objetivo es evaluar si los factores implícitos en el desarrollo de la enfermedad cardiovascular son los mismos en estas personas que en las otras dos cohortes.

Después de 53 años de investigaciones, se ha iniciado otra fase del famoso estudio, con la tercera generación de los participantes originarios. El Instituto Nacional de Sangre, Pulmón y Corazón y la Universidad de Boston, que se asociaron en 1971 para codirigir el estudio Framingham, han anunciado recientemente la puesta en marcha de la nueva fase, incluyendo unas directrices y unas guías sobre la investigación. La información completa del diseño del estudio desde sus inicios hasta la actualidad se encuentra en la siguiente URL:

<http://www.nhlbi.nih.gov/about/framingham/index.html>.

El objetivo de la nueva fase es identificar los genes que contribuyen a un buen estado de salud y los que se asocian al desarrollo de enfermedad cardiovascular, hematológica y pulmonar. Con el desarrollo de las nuevas técnicas de imagen se podrá detectar la aterosclerosis en las primeras fases y compararla con las arterias de los sujetos sanos. Esta nueva fase contribuirá a consolidar los conocimientos iniciales sobre susceptibilidad genética cuyo estudio ya fue abordado en el *Framingham Offspring Cohort*, sobre la que se está centrando esta tesis.

Los encargados de la nueva fase piensan reclutar aproximadamente unos 3.500 nietos de los sujetos seleccionados en 1948 y puede ser una oportunidad para analizar las nuevas cuestiones planteadas en la enfermedad cardiovascular, sobre

todo los riesgos, prestando particular atención a la transmisión familiar de los factores genéticos.

**Tabla 1.3. Principales hitos alcanzados en la cohorte del Framingham Offspring Study (Fuente: Institutos Nacionales de Salud de Estados Unidos)**

| <b>Año</b> | <b>Hito</b>  |
|------------|--|
| 1948       | Comienzo del Estudio Framingham.   |
| 1956       | Resultados en la progresión de la enfermedad cardíaca reumática.   |
| 1959       | Encuentran factores que aumentan la probabilidad de la enfermedad cardíaca. Algunos ataques del corazón descubiertos son silentes.   |
| 1960       | El tabaquismo aumenta el riesgo de la enfermedad cardíaca.   |
| 1961       | Los niveles de colesterol, tensión arterial y anomalías en el electrocardiograma elevan el riesgo de la enfermedad cardíaca.   |
| 1965       | Primer informe de Framingham sobre el ictus.   |
| 1967       | Actividad física reduce el riesgo de la enfermedad cardíaca y la obesidad lo aumenta.  |
| 1970       | La tensión arterial alta eleva el riesgo de ictus.   |
| 1971       | Comienza el estudio para describir la progresión de la enfermedad cardíaca.  |
| 1974       | Descripción de la diabetes y de sus complicaciones.  |
| 1976       | La menopausia aumenta el riesgo de enfermedad cardíaca.  |
| 1977       | Se describen los efectos de los triglicéridos y del colesterol LDL y HDL.  |
| 1978       | Los factores psicosociales afectan a la enfermedad cardíaca. La fibrilación atrial también eleva el riesgo.  |
| 1981       | El filtro de los cigarrillos no protege frente a la enfermedad cardíaca coronaria. El informe publica la relación entre dieta y enfermedad cardíaca.   |
| 1983       | Informes sobre el prolapso de la válvula mitral.   |
| 1986       | Primer informe sobre demencia.   |
| 1987       | Los niveles altos del colesterol se correlacionan directamente con el riesgo de muerte en hombres jóvenes. El fibrinógeno eleva el riesgo de la enfermedad cardíaca. La terapia hormonal con estrógenos reduce el riesgo de las fracturas de la cadera en mujeres postmenopáusicas.                |
| 1988       | Los niveles altos del colesterol HDL reducen el riesgo de muerte. La hipertensión sistólica aislada también eleva el riesgo de enfermedad cardíaca y el tabaquismo aumenta el riesgo de ictus.   |
| 1990       | La homocisteína se perfila como posible factor de riesgo.  |
| 1991       | Se establecen los modelos de la predicción del riesgo de la enfermedad cardíaca.   |
| 1993       | La hipertensión sistólica aislada moderada eleva el riesgo de enfermedad cardíaca. Un informe predice supervivencia después de un paro cardíaco.   |
| 1994       | La hipertrofia del ventrículo izquierdo eleva el riesgo de ictus. La lipoproteína (a) se confirma como posible factor de riesgo de enfermedad cardíaca. Se describen los factores de riesgo para la fibrilación atrial. La apolipoproteína E se considera factor de riesgo de enfermedad cardíaca. |
| 1995       | Primer informe Framingham sobre insuficiencia cardíaca diastólica. Comienza el estudio de OMNI para minorías.  |
| 1996       | Describen la progresión de la hipertensión a la insuficiencia cardíaca.  |
| 1997       | Publicación de los efectos acumulativos del tabaco y del colesterol alto en el riesgo de aterosclerosis. Investigación del impacto de la hipertrofia del ventrículo izquierdo en la insuficiencia cardíaca en sujetos asintomáticos.   |

### 1.6.1.3 Hitos principales alcanzados en esta cohorte

Si se hiciera una búsqueda exhaustiva en Google o en PubMed, basada en el vocablo “Framingham”, sin restricciones de campos o contenidos, la cantidad de información sería enorme. No hay editorial, revisión de tema, guía de tratamiento, sección de correspondencia que al hablar de enfermedades cardiovasculares no tenga que referirse obligatoriamente a Framingham. En la Tabla 1.3 se resumen los principales hitos alcanzados en la investigación cardiovascular alcanzados en esta cohorte.

## 1.6.2 Caracterización del riesgo cardiovascular

El concepto de “factor de riesgo cardiovascular” fue acuñado en la década de los años cincuenta por los investigadores del *Framingham Heart Study* y se utiliza para definir a un con-junto de características, genéticas o adquiridas, que se asocian con un riesgo elevado de presentar enfermedades cardiovasculares. La gran ventaja de la caracterización de los denominados factores de riesgo, es que se puede actuar sobre ellos para minimizarlos y prevenir así la enfermedad. Por ejemplo, una vez establecido que el consumo de tabaco es un factor de riesgo cardiovascular, se puede minimizar dicho riesgo disminuyendo o eliminando el consumo de tabaco. Tradicionalmente, los factores de riesgo cardiovascular se han clasificado en modificables (factores ambientales) y en no modificables (factores biológicos), argumentando que un individuo no podía actuar modificando su “carga” genética para reducir el riesgo. Sin embargo, los datos de los primeros estudios de epidemiología genómica abordando las interacciones gen-ambiente han puesto de manifiesto que en la mayoría de las ocasiones no existe un determinismo genético, sino que son muy importantes las denominadas modulaciones ambientales. Es decir, una determinada variación genética en el genoma no se asocia de manera determinista con un mayor riesgo de enfermedad, sino que en función del comportamiento ambiental, se puede incluso anular dicho riesgo. La dieta, junto con el ejercicio físico y el consumo de tabaco y de alcohol son los principales moduladores de los factores genéticos cuyo análisis conjunto es necesario evaluar en los nuevos diseños de epidemiología genómica de las ECV e incorporar en los cálculos integrados del riesgo cardiovascular (Ordovás et al, 2002).

Además la interpretación del riesgo estimado de padecer una enfermedad cardiovascular requiere una definición precisa de las ECV y de los distintos tipos de riesgo que pueden calcularse. Las ECV están definidas como enfermedades coronarias, incluyendo infartos de miocardio, enfermedades vasculares de la periferia y fallos cardíacos. Las distintas estimaciones de riesgo que se dan en la actualidad son el riesgo absoluto y el riesgo relativo. El *Riesgo absoluto* se define

como la probabilidad de padecer enfermedades cardiovasculares durante un periodo de tiempo. El *Framingham Heart Study* sitúa ese periodo en 10 años. El *Riesgo relativo* se define como la probabilidad de padecer enfermedades cardiovasculares basándose en los factores de riesgo individuales.

Si las tendencias actuales se mantienen en los avances científicos y en los resultados de las investigaciones en curso, se llegará a una mayor comprensión de la contribución de los factores de riesgo sobre las enfermedades cardiovasculares, lo que permitirá que se realice un control todavía más adecuado y con éxito de la mayor causa de mortalidad que sufren actual-mente los países desarrollados.

En (Wilson et al., 1998) se da un modelo predictivo del riesgo cardiovascular, basado en el estudio de una parte de la población de muestra del *Framingham Heart Study*, que se denomina habitualmente la Ecuación de Framingham. En el Anexo 10.1 se desarrolla este modelo.

### **1.6.3 La Bioinformática para el análisis genético**

Teniendo presentes, las limitaciones de los modelos clásicos para el cálculo de riesgo cardiovascular, los axiomas de la Epidemiología Genómica de las ECV, la finalización de la secuenciación del Genoma Humano y la potencialidad de la Bioinformática, el gran reto planteado en esta tesis, como consecuencia de la línea de investigación del grupo en el que se encuadra el doctorando, es el desarrollo de un modelo bioinformático integrado para abordar este reto. Seguidamente se resumen las principales aproximaciones de este modelo.

#### **1.6.3.1. Conceptos básicos**

Se parte de la asunción general de que el fenotipo (fenotipo final y fenotipos intermedios) se compone del genotipo (conjunto de factores genéticos), y de un conjunto de factores ambientales (cuya denominación se propone que sea “ambientotipo”). Dado que este modelo incorpora datos genéticos y ambientales hasta ahora no considerados por los modelos clásicos, es imprescindible, como etapa inicial en la construcción de este modelo, la generación de nuevo conocimiento acerca de la contribución relativa y posible interacción de los múltiples factores genéticos y ambientales implicados. Son necesarias pues nuevas investigaciones que generen este conocimiento, así como la incorporación de los datos que consistentemente ya se han ido generando.

De manera inicial, para la generación de este conocimiento se tienen en cuenta las consideraciones generales de Andrieu et al. (1998) en su revisión de los

métodos que aplican la aproximación genética al estudio de la interacción gen-ambiente: consideración de dos categorías según la situación de la información inicial (1. Cuando se conoce la susceptibilidad genética y se dispone de un indicador o medida actual; y 2. Cuando no ocurre nada de esto). Se prefiere por su flexibilidad la aproximación al riesgo relativo que se realiza a través de la odds ratio (OR) estimada a partir de un diseño de casos y controles. En este tipo de diseño se consideran casos los individuos que presentan el fenotipo cardiovascular de interés, y controles aquellos individuos con ausencia de dicho fenotipo, calculando la contribución de los factores genéticos y ambientales considerados globalmente como “exposición”. Estas estimaciones del riesgo pueden ajustarse simultáneamente por un amplio conjunto de potenciales variables de confusión, así como realizar las pertinentes estratificaciones. De manera clásica se han utilizado dos aproximaciones para el cálculo del riesgo: el modelo multiplicativo y el modelo aditivo.

#### Modelo de riesgo multiplicativo:

$$OR_{interacción} = \frac{OR_{exposición,genotipo}}{OR_{exposición} * OR_{genotipo}} \quad (E: 1.1)$$

donde

$OR_{exposición}$ : Odd Ratio del efecto de la exposición al factor ambiental aislado.

$OR_{genotipo}$ : Odd Ratio del efecto del genotipo aislado.

$OR_{exposición,genotipo}$ : Odd Ratio del efecto combinado del genotipo y de la exposición al factor ambiental.

#### Modelo de riesgo multiplicativo:

$$OR_{interacción} = \frac{OR_{exposición,genotipo}}{OR_{exposición} + OR_{genotipo} - 1} \quad (E: 1.2)$$

Donde los elementos de la ecuación tienen el mismo significado que en el modelo anterior. En ambos modelos, el valor obtenido del OR sobre el efecto de interacción se puede interpretar de la forma siguiente:

$$OR_{\text{interacción}}: \begin{cases} > 1 & \text{efecto significativo} \\ = 0 & \text{no procede} \\ < 1 & \text{efecto significativo} \end{cases} \quad (\text{E: 1.3})$$

Este método permite determinar la interacción entre factores de riesgo no específicos, y luego concentrarse en la interacción gen-ambiente. Las frecuencias de los distintos factores a incluir en el estudio determina principalmente el tamaño de la muestra.

Para la obtención de las funciones que posibiliten los cálculos del riesgo cardiovascular con la universalidad y especificidad que se pretende, es necesario disponer de datos sobre estas variables en poblaciones de diferentes lugares geográficos a escala mundial.

### 1.6.3.2 Fuentes de datos

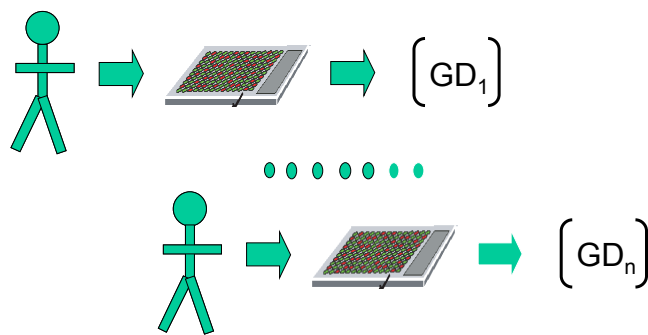
Las fuentes de datos que provienen de los estudios epidemiológicos son diversas y los datos son de diversos tipos. Sin embargo, para caracterizar con precisión el modelo, se asume que los factores de riesgo están representados en un modelo por variables que pertenecen a unos dominios de datos determinados como los siguientes:

- *Dominio de datos binarios*: los valores válidos son únicamente “existe” y “no existe”.
- *Dominio de datos discretos infinitamente numerables*: los valores válidos son valores enteros discretos sin limitación (ni valor ínfimo, ni valor supremo).
- *Dominio de datos discretos finitamente numerables (categóricos)*: los valores válidos son valores enteros discretos con limitación (conjunto con valores ínfimo y supremo) y cardinalidad reducida.
- *Dominio de datos continuos infinitamente numerables*: los valores válidos son valores decimales (números racionales) sin limitación y con la precisión (número de cifras decimales) conveniente.

Una de las principales fuentes de datos son las micromatrices genómicas, tanto de genotipado masivo, como de perfiles de expresión para cada individuo. Aunque la tecnología de micromatrices de ADN ha experimentado un espectacular desarrollo desde que en 1994 Affymetrix construyó la primera micromatriz de

ADN de alta densidad, la principal limitación todavía reside en el análisis e interpretación de los datos (Cox et al., 2003). Actualmente, se pueden analizar en el laboratorio miles de genes en poco tiempo y a un coste muy reducido. Sin embargo, los datos generados pueden contener importantes sesgos que le restan toda utilidad potencial. Por ello, uno de los principales retos actuales de la Bioinformática es el establecimiento de estándares para la normalización y el filtrado de los datos que se generan con estas micromatrices, así como el desarrollo de nuevas metodologías de análisis que supere los algoritmos de racimos (clusters) y los tests estadísticos tradicionales.

Superadas estas limitaciones, la información proveniente de las micromatrices refleja de manera no sesgada los datos del genotipo de cada individuo, que se representan como  $GD_i$ , siendo  $i = 1, \dots, n$ ; y  $n$  es el tamaño de la muestra de individuos (Figura 1.16).



$n$ : tamaño de la muestra

Figura 1.16. Las matrices genómicas como fuentes de datos del modelo de riesgo

Por otra parte, los análisis clínicos en el laboratorio producen datos bioquímicos de cada individuo de la muestra. Cada conjunto de datos bioquímicos procedentes de cada individuo se representa como  $BD_i$ , siendo  $i = 1, \dots, n$ ; y  $n$  es el tamaño de la muestra de individuos (Figura 1.17).

Finalmente, las exploraciones clínicas producen datos clínicos, antropométricos y ambientales de cada individuo de la muestra. Cada conjunto de datos clínicos, antropométricos y ambientales procedentes de cada individuo se representa respectivamente como  $CD_i$ ,  $AD_i$  y  $ED_i$ , siendo  $i = 1, \dots, n$ ; y  $n$  es el tamaño de la muestra de individuos (Figura 1.18).

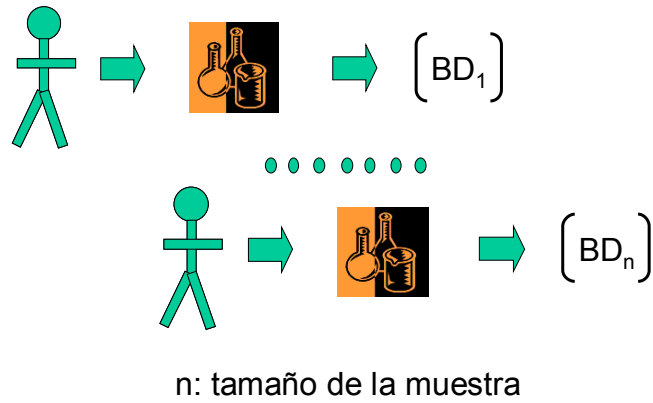


Figura 1.17. Los análisis clínicos como fuentes de datos del modelo de riesgo

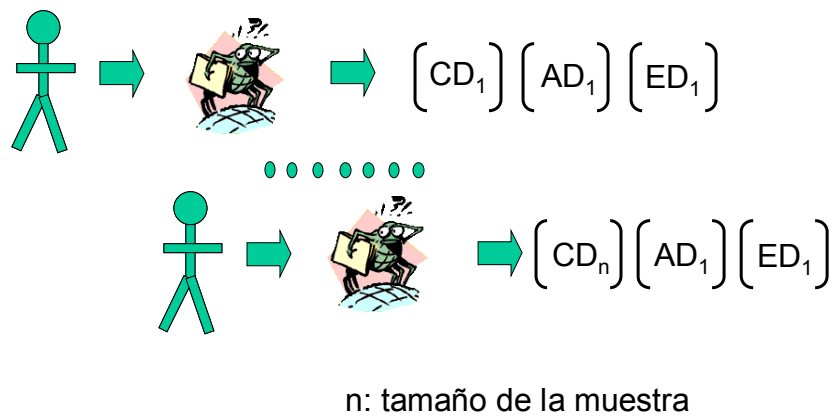


Figura 1.18. Las exploraciones clínicas como fuentes de datos del modelo de riesgo

Por lo tanto, los datos del fenotipo de cada individuo de la muestra,  $PD_i$ , es a priori la unión de los conjuntos anteriores:

$$\left[ PD_i \right] = \left[ GD_i \right] \cup \left[ BD_i \right] \cup \left[ CD_i \right] \cup \left[ AD_i \right] \cup \left[ ED_i \right] \quad (E: 1.4)$$

Normalmente, los datos bioquímicos, clínicos y antropométricos se consideran como expresiones observables del genotipo, o fenotipos intermedios. Así, se puede simplificar la expresión algebraica anterior agrupando los conjuntos de datos que



no son datos ambientales, es decir, agrupando los datos genotípicos y fenotípicos intermedios de la forma siguiente:

$$\left( PD_i \right) = \left\{ \left( GD_i \right) \cup \left( BD_i \right) \cup \left( CD_i \right) \cup \left( AD_i \right) \right\} \cup \left( ED_i \right) \quad (E: 1.5)$$

obteniendo la siguiente expresión donde  $G'D_i$  es la unión de los datos genotípicos y fenotípicos intermedios:

$$\left( PD_i \right) = \left( G'D_i \right) \cup \left( ED_i \right) \quad (E: 1.6)$$

Esta expresión general es la que indica, desde el punto de vista de los dominios de datos, que los datos genéticos y ambientales deben tratarse de forma integrada en la medida de lo posible.

### 1.6.3.3 Un modelo conceptual para el cálculo del riesgo cardiovascular

Es sabido que la contribución de cada factor al fenotipo es un efecto complejo compuesto por los efectos de los factores y los efectos de la interacción factor-factor. En primer lugar, se establece un conjunto de símbolos para distinguir los tipos de efectos:

- Interacción gen-gen: “gg”.
- Interacción gen-ambiente: “ge”.
- Interacción ambiente-ambiente: “ee”.

A continuación se definen las funciones que explican los distintos tipos de efectos:

- Función  $\alpha(x_i)$ : efectos del factor  $x_i$  aislado.
- Función  $\beta(x_i, y_j)$ : efectos de la interacción entre los factores  $x_i$  e  $y_j$ .
- Función  $\gamma^{gg}(x_i, y_j)$ : efectos de cada elemento  $(i, j)$  en la rejilla de interacción.
- Función  $\Gamma^{gg}(n, m)$ : efectos globales de la rejilla de interacción de dimensión  $n \times m$ .

Se hace uso de los símbolos  $\Sigma$  (sumatorio de todos los términos) y  $\Pi$  (productorio de todos los términos) para las expresiones matemáticas de estas funciones que se desarrollan a continuación.

### Interacción gen-gen: "gg"

En la Figura 1.19 se muestra la primera de las vistas parciales del modelo conceptual, la matriz que representa las posibles interacciones gen-gen entre todos los genes candidatos por medio de los polimorfismos identificados. La matriz es parcialmente simétrica con respecto a la diagonal principal (i,i).

En esta vista parcial, los efectos de cada elemento, función  $\gamma^{gg}(x_i, y_j)$ , y globales, función  $\Gamma^{gg}(n, m)$ , se expresan de la forma siguiente:

$$\text{Función } \gamma^{gg}(g_i, g_j) = \alpha(g_i) + \alpha(g_j) + \beta(g_i, g_j) \quad (\text{E: 1.7})$$

$$\text{Función } \Gamma^{gg}(n, m) = \frac{1}{2} \sum_i \sum_j [\alpha(g_i) + \alpha(g_j) + \beta(g_i, g_j)] - \sum_i [\alpha(g_i) + \beta(g_i, g_i)] \quad (\text{E: 1.8})$$

|   |   | i              |                |                |                |                |
|---|---|----------------|----------------|----------------|----------------|----------------|
|   |   | →              |                |                |                |                |
|   |   |                | g <sub>1</sub> | g <sub>2</sub> | g <sub>3</sub> | g <sub>4</sub> |
| j | ↓ |                |                |                |                |                |
|   |   | g <sub>1</sub> | *              |                | #              |                |
|   |   | g <sub>2</sub> |                | *              |                | #              |
|   |   | g <sub>3</sub> | #              |                | *              |                |
|   |   | g <sub>4</sub> |                | #              |                | *              |

Figura 1.19. Modelo de interacciones: matriz gen-gen. \*: interacción por defecto de cada par de genes idénticos. #: interacción efectiva de cada par de genes distintos

**Interacción gen-ambiente: “ge”**

En la Figura 1.20 se muestra la segunda de las vistas parciales del modelo conceptual, la matriz que representa las posibles interacciones gen-ambiente entre todos los genes candidatos por medio de los polimorfismos identificados y los factores ambientales caracterizados. La matriz no es simétrica con respecto a la diagonal principal (i,i). En este caso, no hay interacción por defecto de cada par de factores idénticos.

En esta vista parcial, los efectos de cada elemento, función  $\gamma^{ge}(x_i, y_j)$ , y globales, función  $\Gamma^{ge}(n, m)$ , se expresan de la forma siguiente:

$$\text{Función } \gamma^{ge}(g_j, e_k) = \alpha(g_j) + \alpha(e_k) + \beta(g_j, e_k) \quad (\text{E: 1.9})$$

$$\text{Función } \Gamma^{ge}(m, r) = \frac{1}{2} \sum_j \sum_k [\alpha(g_j) + \alpha(e_k) + \beta(g_j, e_k)] - \sum_i [\alpha(g_j) + \beta(g_j, e_k)] \quad (\text{E: 1.10})$$

|    |   | e1 | e2 | e3 | e4 |
|----|---|----|----|----|----|
| g1 | # |    |    | #  |    |
| g2 |   |    |    |    | #  |
| g3 |   |    |    |    |    |
| g4 |   |    | #  |    | #  |

Figura 1.20. Modelo de interacciones: matriz gen-ambiente. #: interacción efectiva de cada par de factores distintos

**Interacción ambiente-ambiente: “ee”**

En la Figura 1.21 se muestra la tercera de las vistas parciales del modelo conceptual, la matriz que representa las posibles interacciones ambiente-ambiente

entre todos los factores ambientales caracterizados. La matriz es parcialmente simétrica con respecto a la diagonal principal (k,k).

En esta vista parcial, los efectos de cada elemento, función  $\gamma^{eg}(x_i, y_j)$ , y globales, función  $\Gamma^{eg}(n, m)$ , se expresan de la forma siguiente:

$$\text{Función } \gamma^{ee}(e_k, e_l) = \alpha(e_k) + \alpha(e_l) + \beta(e_k, e_l) \quad (\text{E: 1.11})$$

$$\text{Función } \Gamma^{ee}(r, s) = \frac{1}{2} \sum_k \sum_l [\alpha(e_k) + \alpha(e_l) + \beta(e_k, e_l)] - \sum_k [\alpha(e_k) + \beta(e_k, e_l)] \quad (\text{E: 1.12})$$

|     | k → |    |    |    |
|-----|-----|----|----|----|
|     | e1  | e2 | e3 | e4 |
| i ↓ | e1  | *  |    | #  |
|     | e2  |    | *  | #  |
|     | e3  | #  |    | *  |
|     | e4  |    | #  |    |

*Figura 1.21. Modelo de interacciones: matriz ambiente-ambiente. \*: interacción por defecto de cada par de factores idénticos. #: interacción efectiva de cada par de factores distintos*

### Integración de las vistas

El siguiente paso en la construcción del modelo es la integración paulatina de las tres vistas. En primer lugar, se combinan las vistas primera (matriz gen-gen) y segunda (matriz gen-ambiente) dando el resultado expuesto en la Figura 1.22.

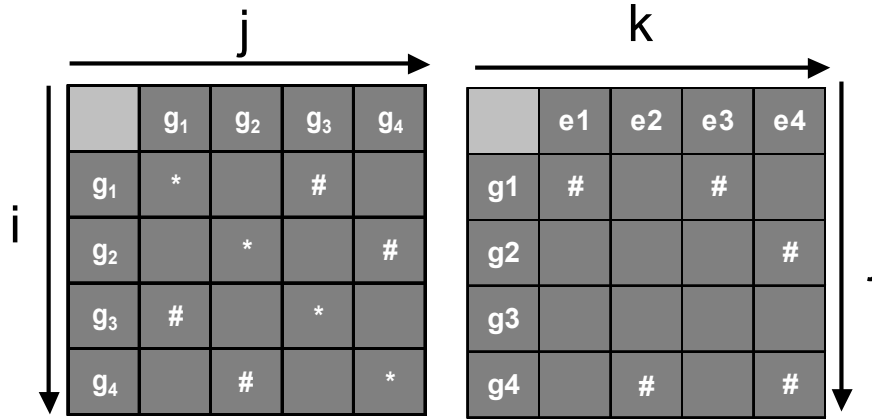


Figura 1.22. Modelo de interacciones: combinación de matrices gen-gen y gen-ambiente

En esta vista combinada, los efectos de cada elemento, función  $\varphi^{gg,ge}(x_i, y_j, z_k)$ , y globales, función  $\Gamma^{gg,ge}(n, m, r)$ , se expresan de la forma siguiente:

$$\text{Función } \varphi^{gg,ge}(g_i, g_j, e_k) = \gamma^{gg}(g_i, g_j) \bullet \gamma^{ge}(g_j, e_k) \tag{E: 1.13}$$

$$\text{Función } \Gamma^{gg,ge}(n, m, r) = \Gamma^{gg}(n, m) \bullet \Gamma^{ge}(m, r) = \Pi_{n,r} [\dots] \tag{E: 1.14}$$

En segundo lugar, se combina la vista tercera (matriz ambiente-ambiente) con las dos anteriores dando el resultado expuesto en la Figura 1.23.

En esta vista combinada, los efectos de cada elemento, función  $\varphi^{gg,ge,ee}(x_i, y_j, z_k)$ , y globales, función  $\Gamma^{gg,ge,ee}(n, m, r, s)$ , se expresan de la forma siguiente:

$$\text{Función } \varphi^{gg,ge,ee}(g_i, g_j, e_k) = \gamma^{gg}(g_i, g_j) \bullet \gamma^{ge}(g_j, e_k) \bullet \gamma^{ee}(e_k, e_j) \tag{E: 1.15}$$

$$\text{Función } \Gamma^{gg,ge,ee}(n, m, r, s) = \Gamma^{gg}(n, m) \bullet \Gamma^{ge}(m, r) \bullet \Gamma^{ee}(r, s) \tag{E: 1.16}$$

Con lo cual se tiene

$$\Gamma^{gg}(n, m) \bullet \Gamma^{ge}(m, r) \bullet \Gamma^{ee}(r, s) = \Pi_{n,r} [\dots] \Pi_{m,s} [\dots] \tag{E: 1.17}$$

$$\text{Función } \Gamma^{gg,ge,ee}(n, m, r, s) = \Pi_{n,r} [\dots] \Pi_{m,s} [\dots] \tag{E: 1.18}$$

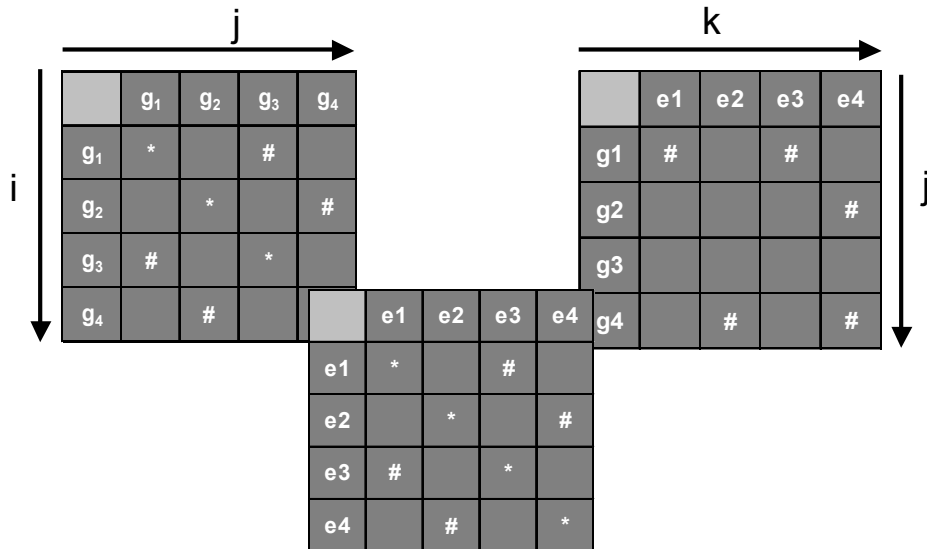


Figura 1.23. Modelo de interacciones: combinación de matrices gen-gen, gen-ambiente y ambiente-ambiente

La Fórmula (E: 1.18) es la expresión final del modelo formal global para el cálculo de los efectos de las interacciones combinadas entre los factores genéticos y ambientales que componen el riesgo cardiovascular. En esta expresión se han obtenido *tres* dimensiones distintas de pares de interacciones y *cinco* dimensiones distintas de efectos de distintos tipos.

Con esta formulación conceptual, se debe estudiar si se pueden analizar los datos producidos con este modelo con las herramientas conceptuales y técnicas disponibles en la actualidad. De entrada, las funciones  $\alpha(x_i)$  (efectos del factor  $x_i$  aislado) y  $\beta(x_i, y_j)$  (efectos de la interacción entre los factores  $x_i$  e  $y_j$ ), pueden ser de distinta naturaleza, como, por ejemplo, funciones matemáticas, funciones estadísticas, técnicas heurísticas, algoritmos de minería de datos, etc. Es decir, no se debe plantear la aplicación del modelo basándose solamente en la utilización de herramientas matemáticas.

Este es uno de los axiomas fundamentales de la Epidemiología Genómica, es decir, buscar las herramientas necesarias e intentar combinarlas para encontrar soluciones que, aún no siendo perfectas, puedan ir aproximándose a la mejor interpretación. Entonces, en esta tesis se adopta este enfoque cuando se tratan y procesan los resultados en el ámbito de la Epidemiología Genómica de las Enfermedades Cardiovasculares.

### 1.6.4 Un marco de integración para la Bioinformática y la Epidemiología Genómica

Esta tesis se ha desarrollado en un marco general que se va conformando día a día mediante la consolidación de un gran conjunto de disciplinas relacionadas con las ciencias de la vida, las ciencias de la salud y las ciencias de la computación, que se denomina Medicina Genómica (Figura 1.24).

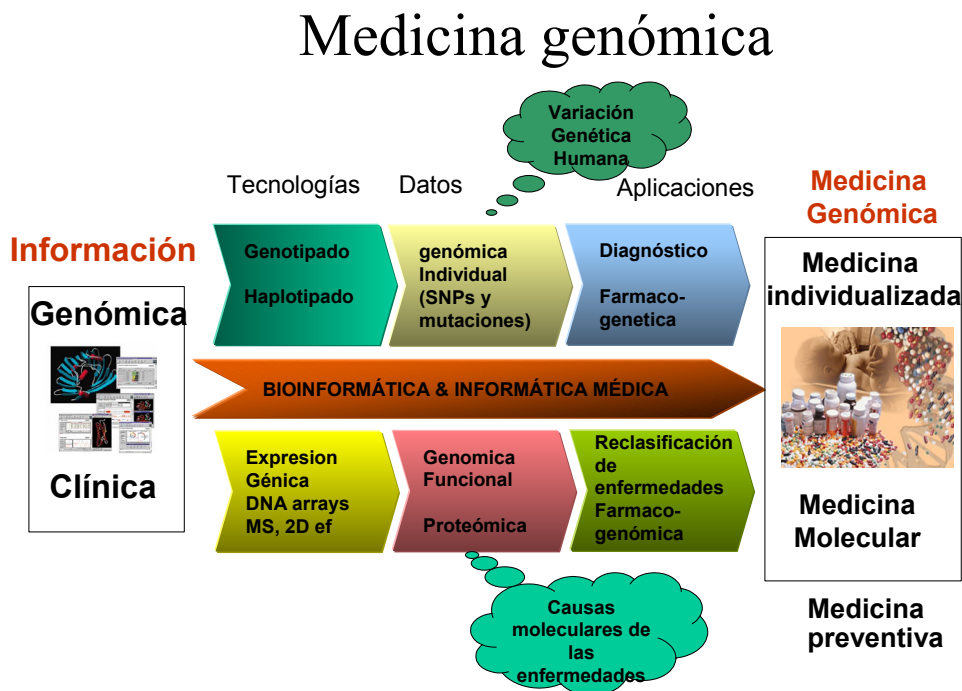


Figura 1.24. Arquitectura de La Medicina Genómica (Fuente: Fernando Martín, SemBioGen de Zaragoza 2003)

Como se puede apreciar en la Figura 1.24 el fin último de la Medicina Genómica es la aplicación de la medicina individualizada teniendo en cuenta el genoma de cada sujeto, sea para prevenirle de las enfermedades (Medicina genómica preventiva), sea para tratarle de las que tiene (Medicina genómica clínica). En la misma Figura se puede observar también el papel que está desempeñando la Bioinformática, junto con la Informática Médica, en la evolución de esta estrategia. Sin embargo, por mor de simplicidad, no aparecen dos de las

disciplinas que también son importantes, la Epidemiología Genómica y la Informática Biomédica. La primera está integrada en las sucesivas etapas de tecnologías, datos y aplicaciones. La segunda está a caballo de la Bioinformática y la Informática Médica.

Actualmente, la Medicina Genómica sigue siendo un enfoque conceptual de muy alto nivel y no es posible obtener consecuencias prácticas ni procesos concretos de él. Por lo tanto, ha sido necesario delimitar el campo de actuación de las investigaciones de esta tesis en un marco mucho más concreto y que servía para los propósitos específicos de la misma. Así, se diseñó y estableció un marco de trabajo consistente en la integración de la Bioinformática como disciplina complementaria de la Epidemiología Genómica Cardiovascular de forma que, una sin la otra no pueden llegar a cumplir los objetivos planteados según se detallan en el Capítulo 2 (Figura 1.25).

Partiendo de la información sobre los sujetos del Estudio Framingham, los procesos principales de la Epidemiología Genómica Cardiovascular aplicados en el laboratorio de referencia son la extracción y tratamiento de ADN de las muestras de sangre, la obtención de expresión génica en las muestras analizadas según los genes seleccionados, el genotipado de regiones en las que se han encontrado mutaciones pero no están identificadas, el haplotipado con los alelos correspondientes a los sujetos del estudio, y la búsqueda de interacciones gen\*gen, gen\*ambiente y sus variantes (gen\*dieta, gen\*sexo, etc.) para poder establecer asociaciones con los fenotipos de disfuncionalidades o enfermedades cardiovasculares identificadas. Todo ello se hace siguiendo lo más rigurosamente posible las teorías científicas y las normas técnicas (y éticas y legales si se trata de información de sujetos humanos y/o seres vivos), y aplicando las técnicas de laboratorio y análisis según las indicaciones de los protocolos científicos.

Por otra parte, la Bioinformática va proporcionando enfoques metodológicos para el desarrollo de sistemas de tratamiento de la información o para su correcta organización, herramientas bioinformáticas para solucionar problemas de gestión y búsqueda de secuencias o similares, y soporte en la aplicación de las diversas tecnologías necesarias, así como en el análisis de los datos generados. Todo ello se hace en función de los principios, reglas y recomendaciones de la Ingeniería del Software como aproximación formal y también práctica en el desarrollo de sistemas bioinformáticos.



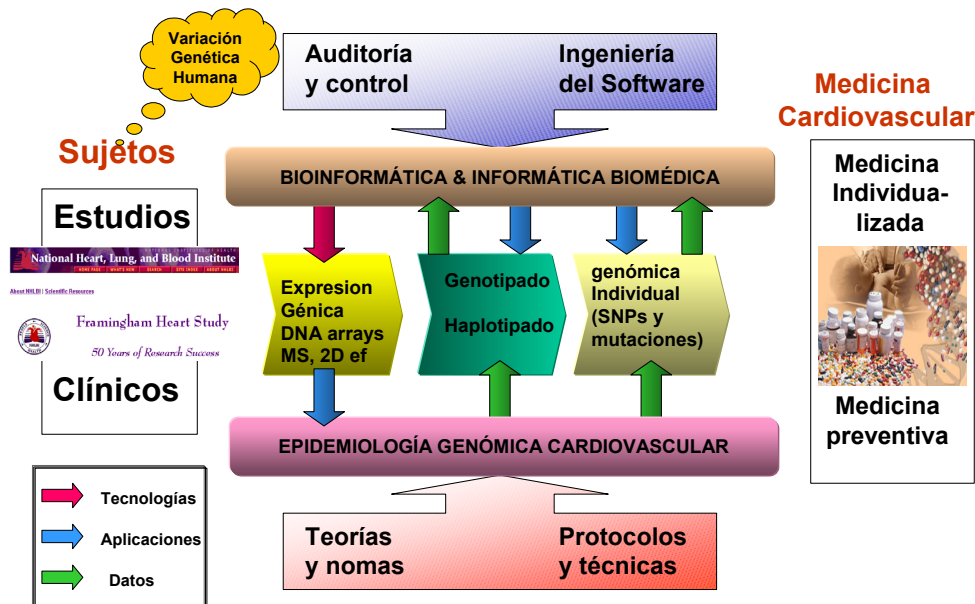


Figura 1.25. Arquitectura de integración de la Bioinformática como soporte de la Epidemiología Genómica

Además, para garantizar la calidad de los datos y los productos, así como la seguridad de la información y de los recursos, aparte del control de la privacidad y uso ético de los datos personales, se aplican los principios, normas y técnicas de la Auditoría de Sistemas de Información adaptadas en un Modelo de Auditoría Informática.

## 1.7 Problemas y proyectos planteados en los procesos de investigación

Desde que se iniciaron los trabajos de investigación planteados en el marco de esta tesis, en la primera estancia como investigador invitado en el HNRC de la Tufts University, en 1997, surgieron algunos problemas y conflictos provocados por el choque de dos formas de trabajar distintas. Una de ellas, la del grupo receptor, era la tradicional en Biología Molecular y Epidemiología Genética, centrada en los

procedimientos de laboratorio (fundamentalmente, manejo de muestras, extracción de ADN, amplificación de ADN, genotipado y análisis estadísticos). La otra era la visión del doctorando, centrada en la Ciencia de la Computación y los Sistemas de Información, donde se priman los aspectos de tratamiento, almacenamiento, recuperación y salvaguarda de la información; así como la aplicación metódica de los procesos y herramientas necesarios para dichas actividades con eficiencia y calidad.

En otras palabras, se encontraron en el mismo territorio la visión científica basada en la muestra biológica, frente a la visión científico-ingeniera basada en la información. En lugar de producirse una divergencia y un fracaso en el proyecto, ambas partes se dieron cuenta de que sus formas de trabajar podían conjugarse y encontrar un mecanismo de acople para trabajar en equipo. Esto fue reforzado por el gran auge que empezó a cobrar el Proyecto del Genoma Humano (Collins et al., 2003) y sus grandes necesidades de tratamiento de información, dando lugar a disciplinas como la propia Epidemiología Genómica, la Bioinformática, y todo el rosario de las denominada comúnmente “ómicas”: Genómica, Proteómica, etc.

Así pues, a lo largo de la colaboración con el grupo del Dr. Ordovás, han ido surgiendo necesidades de tratamiento y control de la información, generada por los sucesivos experimentos realizados en el ámbito del Framingham Heart Study, que el doctorando ha ido abordando sucesivamente. Las soluciones aportadas se han supeditado al marco de integración de Bioinformática descrito anteriormente. En esta sección se exponen los subproyectos bioinformáticos y similares concluidos. Al final de la memoria se mencionarán los proyectos en curso o simplemente diseñados y pendientes de emprender en un futuro cercano.

### **1.7.1 Soporte de comunicación instrumental (PLATEX)**

La Genética Funcional es una disciplina relativamente nueva que ha experimentado un espectacular desarrollo a raíz de los avances en el Proyecto del Genoma Humano y con la ayuda de la Ingeniería Médica. La investigación en esta disciplina, así como en otras muchas relacionadas con la Medicina, se basa en el manejo eficiente de instrumentos de laboratorio que ejecutan amplificaciones y cuantificaciones de ADN y análisis de secuencias y polimorfismos genéticos. Actualmente, estos instrumentos tienen un rendimiento tan alto y manejan tal cantidad de información, que deben estar controlados por software incrustado que asuma las funciones principales para la gestión de la información necesaria y producida por los procesos químicos y físicos que tienen lugar en el laboratorio. Normalmente, los instrumentos de laboratorio vienen de proveedores distintos y sus sistemas de software de control se han desarrollado para diferentes plataformas

estándar. Pero puede ocurrir que, incluso en el mismo instrumento, no todos sus procesos estén interconectados de forma que se puedan trasvasar los datos.

Por esto, los investigadores en Genómica piden soluciones rápidas y simples que ayuden a perfeccionar la gestión de la información y su transvase entre procesos e instrumentos. Hay varias alternativas. Algunas de ellas son demasiado complejas o costosas de desarrollar. Otras son difíciles de manejar. Pero aquí es donde entra en liza la Bioinformática, puesto que permite, al haber avanzado en la introducción de nuevas tecnologías y metodologías, la construcción de herramientas simples y potentes basadas en productos de software estándar.

#### 1.7.1,1 El sistema instrumental

El *ABI PRISM® 3100 Genetic Analyzer*, es un sistema automático de electroforesis capilar que puede separar, descartar y analizar en una sola aplicación hasta 16 capilares de fragmentos de DNA etiquetados con fluorescencia<sup>1</sup>. La operación de este instrumento está dirigida por una estación de trabajo que tiene instalado el sistema operativo Microsoft® Windows NT™. Una cámara CCD convierte la información de fluorescencia en información digital que se procesa en el software *3100 Data Collection* del ordenador. Posteriormente, los datos producidos mediante la electroforesis y procesados por el software, se guardan en la base de datos del instrumento y se muestran como electroferogramas. A continuación, se extraen de la base de datos y se analizan. Y una vez analizados, se almacenan en forma de ficheros de texto en el disco duro de la estación de trabajo.

Finalmente, estos ficheros deben procesarse para crear los registros de placa de fragmentos que son las entradas para el análisis en los programas *Gene Scan Analysis* y *DNA Sequencing Analysis*. El software del Analizador Genético utiliza un registro de placa para asignar a cada placa códigos de muestra (ID), colores del conjunto de muestras (*Dye*), módulos de ejecución y módulos de análisis (Figura 1.26).

Por otra parte, los capilares de fragmentos de DNA etiquetados con fluorescencia se recogen en una estructura que se denomina Rejilla PCR y se maneja mediante una hoja de cálculo (Figura 1.27).

|    | A       | B           | C            | D            | E       | F          | G             | H               | I                  | J                       | K                 |
|----|---------|-------------|--------------|--------------|---------|------------|---------------|-----------------|--------------------|-------------------------|-------------------|
| 10 |         |             |              |              |         |            |               |                 |                    |                         |                   |
| 2  | NHS88_1 | GS          | 96-Well      |              |         |            |               |                 |                    |                         |                   |
| 3  | Well    | Sample Name | Color Number | Standard Dye | Dye Set | Color Info | Color Comment | BioLIMS Project | Sample Tracking Id | Run Module              | Analysis Module   |
| 4  | A1      |             | 1            | 1            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 5  | A1      |             | 1            | 2            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 6  | A1      |             | 1            | 3            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 7  | A1      |             | 1            | 4            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 8  | A1      |             | 1            | 5            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 9  | B1      |             | 1            | 1            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 10 | B1      |             | 1            | 2            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 11 | B1      |             | 1            | 3            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 12 | B1      |             | 1            | 4            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 13 | B1      |             | 1            | 5            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 14 | C1      |             | 1            | 1            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 15 | C1      |             | 1            | 2            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 16 | C1      |             | 1            | 3            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 17 | C1      |             | 1            | 4            | 5       | E5         |               | 3100_Project1   |                    | SNP36_POP4DefaultModule | GS120Analysis.gsp |
| 18 |         |             |              |              |         |            |               |                 |                    |                         |                   |

Figura 1.26. Estructura de datos del registro de placa de fragmentos de ADN. En la columna "Well", cada letra (A, B, C, etc.) designa uno de los pocillos de la placa. En la columna "Color Number", se incluyen los colores del conjunto Dye (hasta 5). Las columnas "Sample Name" y "Color Info" tienen información inespecífica

|    | A | B | C     | D     | E     | F     | G     | H     | I     | J     | K     | L     | M     | N     |
|----|---|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1  |   |   |       |       |       |       |       |       |       |       |       |       |       |       |
| 2  | 1 |   | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |
| 3  |   | A | —     | F3985 | F3739 | F3990 | F3986 | F0331 | F3951 | F0353 | T0087 | F0362 | F0371 | T0095 |
| 4  |   | B | F0354 | F0376 | F0121 | F0374 | F0360 | F0372 | F0351 | F3293 | F0373 | F3726 | F0361 | F3263 |
| 5  |   | C | F2329 | F0615 | F0305 | F0364 | F0363 | D0376 | F0365 | —     | F0571 | F0066 | F0062 | F0027 |
| 6  |   | D | F0030 | F0355 | F0573 | F0099 | F0083 | F0572 | F0605 | F0124 | F0461 | F0316 | F0569 | F0032 |
| 7  |   | E | F0089 | F0570 | F0526 | F0537 | F0084 | F0568 | F0086 | F0087 | F0466 | F0346 | F0464 | F0453 |
| 8  |   | F | F0037 | F0356 | F0462 | F0367 | —     | F0325 | F0463 | F0454 | F0315 | F0081 | F0107 | F0085 |
| 9  |   | G | F0366 | F0345 | F0335 | F0336 | F0090 | F0134 | F0133 | F0131 | F0135 | F0100 | F0101 | F0136 |
| 10 |   | H | F0341 | F0590 | F0580 | F0459 | F0472 | F0311 | F0102 | F0470 | F0324 | F0620 | F0098 | —     |
| 11 |   |   |       |       |       |       |       |       |       |       |       |       |       |       |

Figura 1.27. Estructura de datos de los fragmentos de ADN etiquetados químicamente que contiene la hoja de cálculo de la rejilla de PCR (Son 8 filas de 12 columnas, es decir, 96 muestras como máximo, aunque puede haber "faltas" representadas por "—")

Los datos producidos por electroforesis se muestran en forma de electroferograma, y a continuación se analizan y se guardan como ficheros de texto simples mediante el subsistema de Auto-Extracción. A cada conjunto de muestras

(Dye), que puede tener hasta cinco colores, se le asigna un código de muestra basado en el registro de placa de fragmentos. Estos ficheros son las entradas para el software de análisis como *Gene Scan Analysis*, *Sequence Analysis*, y *Genotyper*. Igual que antes, los códigos de muestra se basan en el registro de placa de fragmentos (Figura 1.28).

| N       | O           | P            | Q            | R       | S          |       |
|---------|-------------|--------------|--------------|---------|------------|-------|
| 1       |             |              |              |         |            |       |
| NHS98_1 | GS          | 96-Well      |              |         |            |       |
| Well    | Sample Name | Color Number | Standard Dye | Dye Set | Color Info | Color |
| A1      | 99999       | 1            | 5            | E5      | 99999      |       |
| A1      | 99999       | 2            | 5            | E5      | 99999      |       |
| A1      | 99999       | 3            | 5            | E5      | 99999      |       |
| A1      | 99999       | 4            | 5            | E5      | 99999      |       |
| A1      | 99999       | 5            | 5            | E5      | 99999      |       |
| B1      | F0354       | 1            | 5            | E5      | F0354      |       |
| B1      | F0354       | 2            | 5            | E5      | F0354      |       |
| B1      | F0354       | 3            | 5            | E5      | F0354      |       |
| B1      | F0354       | 4            | 5            | E5      | F0354      |       |
| B1      | F0354       | 5            | 5            | E5      | F0354      |       |
| C1      | F2329       | 1            | 5            | E5      | F2329      |       |
| C1      | F2329       | 2            | 5            | E5      | F2329      |       |
| C1      | F2329       | 3            | 5            | E5      | F2329      |       |
| C1      | F2329       | 4            | 5            | E5      | F2329      |       |
| C1      | F2329       | 5            | 5            | E5      | F2329      |       |
| D1      | F0030       | 1            | 5            | E5      | F0030      |       |

Figura 1.28. Estructura de datos del fichero de placa que es leído por los programas de secuenciación de ADN y análisis de polimorfismos. Las columnas "Sample Name" y "Color Info" tienen información relativa a las etiquetas de cada muestra. El valor "99999" indica "faltas" en los pocillos

### 1.7.1.2 El tratamiento de los registros de placa

En el caso concreto del instrumento ABI PRISM® 3100 Genetic Analyzer, su Manual de Usuario (Applied Biosystems, 2001) recomienda varios métodos para crear registros de placa de fragmentos. Algunos de ellos son automáticos pero están basados en software o bases de datos obsoletas. El resto no son automáticos e implican el uso de hojas de cálculo y ficheros de texto libre o delimitados por tabuladores. En particular, uno de los más recomendados es el Método 3 descrito de la forma siguiente (Figura 1.29): (1) Ábrase el fichero de texto delimitado por tabuladores provisto mediante Microsoft® Excel™ como una nueva hoja de cálculo; (2) modifíquense los datos de placa de fragmentos y de muestras; y (3) ciérrase el fichero guardando los cambios realizados como un fichero de texto delimitado por tabuladores.

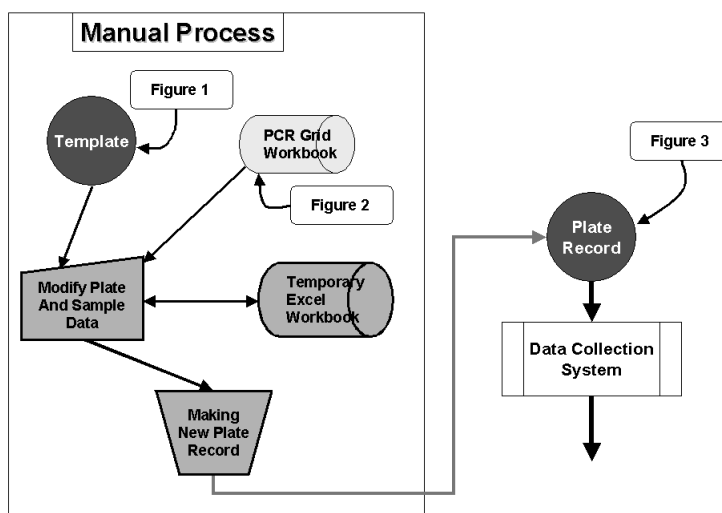


Figura 1.29. El proceso de conversión manual que produce los ficheros de datos (plate record) para la secuenciación y análisis de ADN

Los analizadores genéticos utilizan a menudo etiquetas de registros de muestras correspondientes a placas de 96 o 384 pocillos en otras hojas de cálculo de Microsoft® Excel™. Sin embargo, la estructura del registro de datos de estas hojas de cálculo es completamente distinta de la estructura de los ficheros de texto requeridos por el software de captación de datos para crear el registro de placa de fragmentos. La modificación de los datos de placa y de muestra comportan una gran cantidad de operaciones manuales repetitivas debido al enorme volumen de datos para buscar, copiar, pegar y revisar por cada una de las etiquetas de placa. Y por tanto, el riesgo de error es bastante alto.

### 1.7.2 Manejo y visualización de secuencias (SeqPacker)

Uno de los problemas comunes en el laboratorio del HNRCA, y paralelamente en los trabajos en colaboración con el grupo EPIGEM de la U. de Valencia, está en relación con la manipulación de información sobre secuencias de ADN en los procesos de laboratorio. Sobre todo, a la hora de visualizar las cadenas obtenidas y de explorarlas para poder obtener los primers correspondientes. Esto ocurre porque el software que acompaña a los instrumentos de laboratorio para análisis genético, por ejemplo, ABI PRISM, Quiagen, Affymetrix, etc., es muy rígido y no permite

hacer otras cosas que las marcadas en los requisitos de gestión y control de los correspondientes instrumentos.

### 1.7.2.1 El problema de la manipulación de secuencias

En la realidad de la investigación genómica y en Biología Molecular, la manipulación de secuencias de ADN y ARN es una tarea ordinaria desde ha ce mucho tiempo. Normalmente se quiere hacer una extracción de subsecuencias de primers desde una cadena larga de ADN o ARN, o se necesita buscar una subcadena en una cadena en sentido directo o inverso. El doctorando ha llegado a ver a compañeros de su equipo de investigadores visualizando secuencias de nucleótidos en líneas sin formato (y también en formatos FASTA y GenBank) utilizando el procesador de textos Microsoft Word o incluso el Microsoft Notepad con ficheros de texto normal. Por lo tanto, resulta muy duro intentar hacer alineamientos simples o buscar subsecuencias sin tener las herramientas adecuadas.

Una forma de paliar esta dificultad es la de visualizar las secuencias en bloques de 5 o 10 nucleótidos, formando columnas y con la numeración de la posición de cada base al final y al principio de cada línea (en el Anexo 12.4 se muestran las secuencias de los genes sometidos a estudio en este formato), que es el formato FASTA. Además, si se utiliza el código de colores estándar que asigna un color distinto a cada nucleótido (A: verde; C: negro; G: azul; T o U: rojo) la tarea resulta más fácil. Pero si además se puede cambiar la asignación de colores para, por ejemplo, los usuarios daltónicos, o se puede cambiar el tipo y tamaño de letra para, por ejemplo, los que tienen capacidad visual disminuida, se permite el acceso a un número de usuarios mayor que antes no se lo planteaban (Fisher et al., 1997).

Una de las investigaciones concretas en que se ha detectado esta necesidad de herramientas específicas ha sido en el marco de la colaboración nodo a nodo de la red INBIOMED entre el grupo IRIS (UJI), al cual pertenece el doctorando, y el grupo EPIGEM (U. Valencia). Se trata de la manipulación de secuencias de ADN, extraído de muestras de sangre de sujetos participantes en los estudios de la Comunidad Valenciana, para identificar nuevos polimorfismos asociados a anormalidades lipídicas y cardiovasculares (Ordovas et al., 2002). El grupo EPIGEM, enviaba las muestras a analizar al Servicio de Bioinformática y éste le devolvía los resultados del análisis que se podían visualizar con el programa Chromas.

### 1.7.2.2 La secuenciación de las muestras

El proceso de análisis de secuenciación de las muestras, consiste en irradiarlas con marcadores fluorescentes para luego someterlas a lecturas con un rayo láser que detecta la reflexión de la luz y la separa en cuatro canales según el tipo de nucleótido leído (hay una absorción y reflexión distinta para A, C, G y T, respectivamente). Las lecturas representan la probabilidad de que cada nucleótido sea de uno de los cuatro tipos. Y a medida que la secuenciación avanza, el dispositivo de secuenciación aplica un algoritmo propio a los canales que calcula cuál de los cuatro nucleótidos tiene la mayor probabilidad en cada lectura, asignando dicho nucleótido como resultado. Al final del proceso, la cadena resultante se añade como parte de la salida. Si resulta que las probabilidades son bastante parecidas o no llegan a un umbral mínimo, el instrumento lo toma como ruido o base desconocida y no le asigna ninguna de las letras, sino el comodín “N” o “K”. La salida es un fichero en formato ABI que se describe en el siguiente apartado.

Se ha efectuado un cálculo de la tasa de error de las secuencias manejadas obteniendo un valor de 5 bases equivocadas por cada 1.000 leídas, es decir, de 0,5%. Puesto que las secuencias a analizar tenían, en ese caso, entre 100 y 300 bases, la tasa de fallos esperada era de 0,5 a 1,5 bases respectivamente, no era necesario mejorar la precisión del algoritmo mencionado. Sin embargo, lo que dejaba que desear era la capacidad de representación de este producto, Chromas, cuya interfaz es bastante rígida y no permitía hacer operaciones que eran absolutamente necesarias como, por ejemplo, poder adaptar la interfaz (colores, tipos y tamaños de letras, organización de la zona de trabajo, etc.), poder leer secuencias de ficheros con varios formatos y compararlas entre sí, poder cambiar el sentido de las cadenas (directo/inverso), poder hacer comparaciones y búsquedas de las cadenas leídas, e incluso tener acceso al contenido del portapapeles para obtener fácilmente cadenas o trozos de páginas de Web.

### 1.7.2.3 El formato ABI para los ficheros de nucleótidos

El formato ABI de los ficheros de nucleótidos es uno de los formatos más confusos que se hayan podido diseñar y constituye todo un reto su decodificación. Este formato se compone de un conjunto de registros heterogéneos y de longitud variable que contienen, entre otros tipos de información, la lista completa de los nucleótidos secuenciados según han sido producidos y asignados por el dispositivo secuenciador (usualmente un *ABI PRISM sequencer*). En la Figura 1.30 se muestra el submodelo de objetos en notación UML (Booch et al., 1999) (Rumbaugh et al., 2000) que representa la estructura de información del formato ABI.



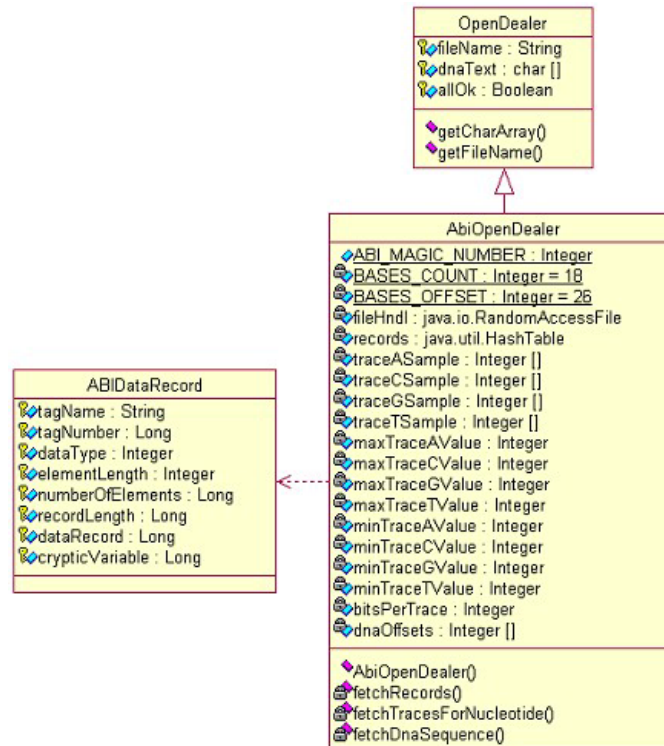


Figura 1.30. Submodelo de objetos del formato ABI de los registros que almacenan las secuencias de nucleótido producidos por el algoritmo de lectura y asignación del secuenciador

Por lo tanto, era necesario confeccionar un algoritmo que permitiera acceder al fichero y leer el contenido de los registros extrayendo únicamente la secuencia de nucleótidos, así como que permitirá guardar dicha secuencia en los formatos normalizados, como FASTA y GenBank (GenBank, 2004). Por otra parte, también se necesitaba una herramienta con gran capacidad de visualización y manipulación de estas secuencias.

En el Capítulo 4 se describen los resultados de este proyecto cuyo producto final es una herramienta bioinformática integrada para la manipulación de secuencias de nucleótidos (ADN y ARN) como una especie de editor con una interfaz gráfica muy amigable. Esta herramienta se ha denominado “SeqPacker”.

Con el propósito de tener un conjunto de herramientas que puedan ejecutarse en distintas plataformas (UNIX, linux, Windows, Appel, etc.) uno de los requisitos

técnicos fundamentales (salvo casos excepcionales) que se fijó inicialmente, ha sido el de utilizar el lenguaje de programación estándar Java para todos los desarrollos de herramientas bioinformáticas. Esto permite además, aplicar con facilidad los principios de la Ingeniería del Software (los que correspondan) ente los cuales figura el de la reutilización del software, y también usar entornos de modelado y desarrollo potentes y asequibles.

### **1.7.3 Búsqueda y detección de polimorfismos (PREDICARD)**

El proyecto que se describe en esta sección es la descripción de parte de la investigación bioinformática, realizada bajo el marco de una beca de movilidad para profesores de universidad financiada por el Ministerio de Educación, Cultura y Deporte, (código PR2003-0063), en el Nutrition and Genomics Laboratory de la Tufts University-United States Department of Agriculture Human Nutrition Research Center on Aging, dirigido por el Dr. José M. Ordovás. El proyecto planteado tenía el título “Sistema bioinformático de soporte en el diseño y análisis de micromatrices de ADN basadas en SNP: aplicación en estudios de genómica cardiovascular”.

#### **1.7.3.1 Introducción**

El impacto de las enfermedades cardiovasculares en la sociedad es bien patente, tanto desde el punto de vista económico, como desde el punto de vista social con sus correspondientes costes sociales. Según la OMS, las enfermedades cardiovasculares son ya la primera causa de mortalidad en el ámbito mundial, de forma que 16,6 millones de personas mueren cada año en el mundo como resultado de las enfermedades cardiovasculares. Esto representa una de cada tres muertes ocurridas. Por otra parte, la atención clínica del paciente con enfermedad cardiovascular es cara y además, dado su carácter de enfermedad crónica, requiere tratamiento prolongado, normalmente de por vida. Estos costos desvían los recursos económicos de la familia y de la sociedad hacia la atención médica y también acortan los años más productivos de los individuos, socavando el desarrollo de la sociedad.

En el ámbito de la salud pública, la mayor parte de los eventos cardiovasculares podrían evitarse o retrasarse significativamente mediante una prevención primaria apropiada y con una relación muy positiva de beneficio/costo. Para ello, la detección precoz del riesgo individual y la implementación de las medidas preventivas más adecuadas constituyen los bastiones principales de la lucha contra esta plaga de la sociedad moderna. Dado el componente genético de estas enfermedades que marcan la predisposición del individuo, es imprescindible

incorporar este conocimiento y las técnicas asociadas en la lucha contra las enfermedades cardiovasculares.

En la actualidad hay una fuerte inversión pública y privada en la Epidemiología Genómica Cardiovascular, tal que cualquier aportación científica y técnica es bien recibida, provenga de donde provenga. Además, hay proyectos en esta disciplina que se realizan por medio de colaboraciones internacionales, sean permanentes o puntuales.

### 1.7.3.2 Planteamiento del problema

El proyecto tiene como objetivo la elaboración de un Modelo Genómico Cardiovascular para recoger y documentar los polimorfismos de los genes relacionados directamente con manifestaciones genotípicas y fenotípicas de las patologías cardiovasculares, los factores de protección y la eficacia a la intervención terapéutica. Para cumplir este objetivo el proyecto se basa en la concurrencia de varias disciplinas científicas y técnicas, así como en la utilización de diversas tecnologías, todas ellas relacionadas con la Genética, la Genómica, la Bioinformática, la Ingeniería del Software, la Gestión de Proyectos, la Inteligencia Artificial, la Medicina Cardiovascular y la Epidemiología Genómica. En esta subsección se describen los aspectos más generales de este proyecto ya que se realizó un diseño detallado en cuanto a tareas, recursos y costes.

### 1.7.4.2.1 Fundamentos del proyecto

Los inputs del proyecto son los datos experimentales, la evidencia publicada y el conocimiento de expertos en la disciplina. Con esta información, se trataba de identificar los genes asociados a estas patologías y los polimorfismos que las determinan. Para ello era necesario consultar la información de las bases de datos genómicas, principalmente las de GenBank del NCBI (Nucleotide, SNP, Protein, etc.) (NCBI, 2003) Con todo ello, se construyó una base de datos de polimorfismos que contenía, entre otras cosas, una secuencia de algunos cientos de nucleótidos por cada polimorfismo identificado. Se buscaba principalmente polimorfismos que estuvieran asociados a un solo nucleótido, denominados SNP (*Single Nucleotide Polymorphism* - Polimorfismos de Nucleótido Único). En estas secuencias, el SNP correspondiente al polimorfismo se dispone en el medio de la secuencia, y así, las ramas izquierda y derecha son las subsecuencias que vienen antes y después de dicho SNP en el gen correspondiente (Figura 1.31).

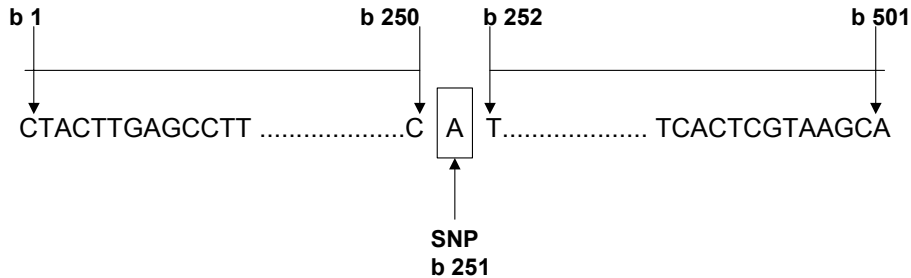


Figura 1.31. Estructura de la secuencia correspondiente a cada polimorfismo

Una de las decisiones tomadas en el proyecto que no estaban completamente fundamentadas en los principios científicos de la Biología Molecular era sobre el tamaño final de la secuencia obtenida. Como se muestra en la Figura 1.31, el tamaño elegido fue de 500 bases más el SNP, es decir, 501 bases en total distribuidas en una subsecuencia inicial de 250 bases, el SNP, y una subsecuencia final de otras 250 bases. Sin merma de la validez de los resultados, igualmente se podría haber elegido un tamaño de 100 o de 1000 bases.

Dado que una de las posibles aplicaciones de estas secuencias es la de servir de base para el diseño de micromatrices de ADN, de cada una de las secuencias se pueden sacar oligonucleótidos de 20, 25 o más nucleótidos. Para evitar las hibridaciones naturales en los pozillos de las sondas, es recomendable que los oligonucleótidos se tomen de zonas que no se solapen, o que los oligonucleótidos no compartan excesivo número de bases. Si se considera que cada probando necesita de 20 a 25 pocillos conteniendo un oligonucleótido distinto extraído de la misma cadena, el número mínimo de bases necesarias sin solapamiento es de:

$$\begin{aligned} \text{número mínimo de bases} &= 20 \text{ oligonucleótidos} \times 20 \text{ bases/oligonucleótido} \\ &= 400 \text{ oligonucleótidos} \end{aligned}$$

Y el número máximo de bases, tomando valores estándares de Affymetrix es de:

$$\begin{aligned} \text{número máximo de bases} &= 25 \text{ oligonucleótidos} \times 25 \text{ bases/oligonucleótido} \\ &= 625 \text{ oligonucleótidos} \end{aligned}$$

Teniendo en cuenta que se pueden dar solapamientos y, a veces, se quiere tener solapamientos en los oligonucleótidos generados, el número de 500 o 501 bases resulta adecuado. Además, con esta longitud, el tamaño de los ficheros que contienen las secuencias es relativamente pequeño y manejable.

#### 1.7.4.2.2 Estructura funcional del proyecto

El modelo de proceso del proyecto se muestra en la Figura 1.32. Las actividades del proceso que corresponden a funciones del proyecto son las siguientes (se mantienen los nombres en inglés ya que así se diseñó el proyecto):

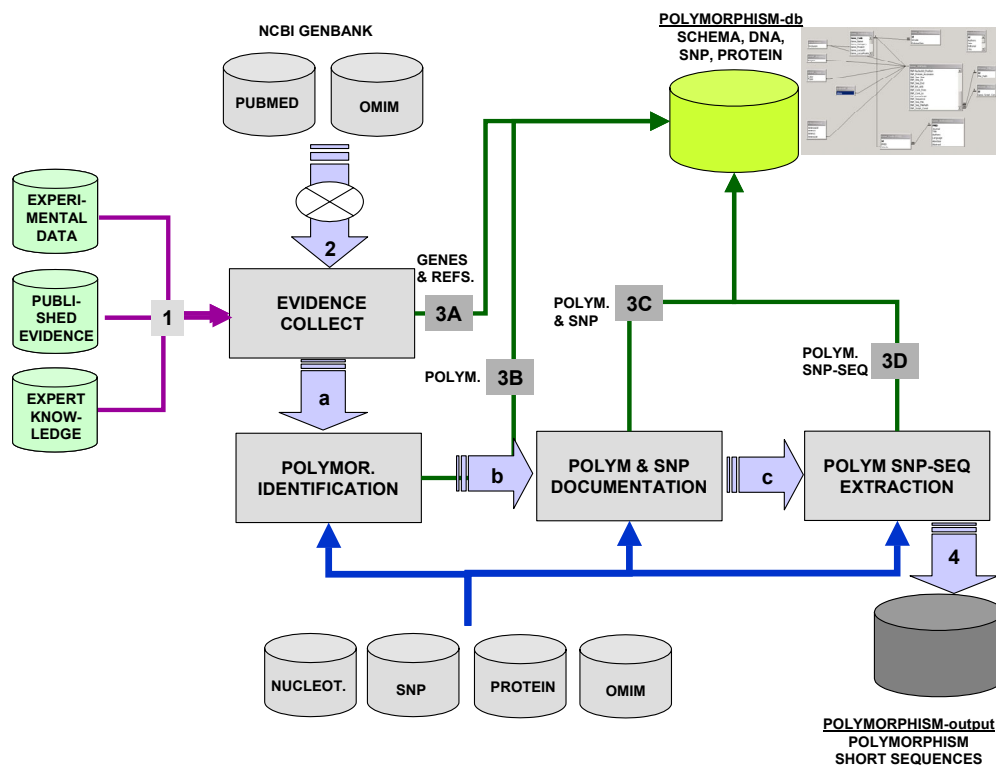


Figura 1.32. Estructura funcional del proyecto

1. **EVIDENCE COLLECT (Recolección de evidencias):** Recolección de las evidencias experimentales y publicadas según el criterio de expertos en el campo científico. Se parte de la literatura publicada, los resultados de experimentos, el conocimiento experto y la información almacenada en las

bases de datos de NCBI GenBank: PubMed y OMIM. Los resultados parciales se guardan en la base de datos.

2. ***POLYMORPHISM IDENTIFICATION (Identificación de polimorfismos)***: Partiendo de la información de la Fase 1, se identifican los polimorfismos en las bases de datos de NCBI GenBank: Nucleotide, SNP, Protein, OMIM, etc. Los resultados parciales se guardan en la base de datos.
3. ***POLYMORPHISM & SNP DOCUMENTATION (Documentación de polimorfismos y SNP)***: Partiendo de la información de la Fase 2, se documentan los polimorfismos y los SNP asociados mediante las bases de datos de NCBI GenBank: Nucleotide, SNP, Protein, OMIM, etc. Los resultados parciales se guardan en la base de datos.
4. ***POLYMORPHISM SNP-SEQ EXTRACTION (Extracción de las secuencias de los polimorfismos y SNP asociados)***: Partiendo de la lista de SNP que se obtiene de la Función 3, se obtienen las subcadenas mediante la aplicación del NCBI Sequence Viewer sobre las bases de datos de NCBI GenBank LocusLink. Las secuencias y sus descriptores se guardan en la base de datos y al mismo tiempo se generan los ficheros de texto para cada secuencia.

En el Capítulo 3 se describe la metodología empleada así como los recursos que fueron necesarios. En el Capítulo 4 se describe el proceso realizado como resultado del proyecto. Y en el Capítulo 5 se incluye una breve discusión sobre el proceso descrito.

#### **1.7.4 Manejo y compresión de secuencias (Squeezeer)**

Otro de los problemas comunes en el laboratorio del HNRCA, y paralelamente en los trabajos en colaboración con el grupo EPIGEM de la U. de Valencia, está en relación con la obtención de información sobre secuencias de ADN desde las bases de datos genómicas como GenBank. Dado que la carga de peticiones y consultas a estos portales cada día es mayor, los tiempos de acceso y respuesta, sobre todo si se trata de bajar secuencias de ADN (genes y cromosomas), se dilatan excesivamente. Esto hace que no se pueda diseñar un proceso bioinformático que trabaje con secuencias y pueda garantizar la obtención de las necesarias desde GenBank o similares.

Por otra parte, la transmisión de ficheros de secuencias entre investigadores o desde grupos a instituciones y viceversa, cada vez está más popularizada. Pero esto ha hecho que, junto con otras cosas, se cargue excesivamente el tráfico de la red. Los investigadores de laboratorios genómicos están acostumbrados a manejar ficheros de texto con secuencias de nucleótidos que visualizan, en los que aplican

búsquedas o que envían a otro colega. Sin embargo, los ficheros de texto no son eficientes, aunque mucho más que ficheros de aplicaciones ofimáticas y pueden ser leídos por cualquiera que los intercepte. Cuando se trata de transmitir secuencias recién genotipadas y no publicadas, la confidencialidad es un requisito absoluto.

Entonces, es en el contexto descrito antes en el que se sitúa otra de las soluciones bioinformáticas derivadas de la investigación realizada en esta tesis: la compresión de secuencias de ADN, ARN o AA para poder manipular y transmitir ficheros de grandes secuencias, como cromosomas y genes, por medios electrónicos y, además, incorporando un cierto grado de confidencialidad.

#### **1.7.4.1 Obtención de ficheros de secuencias**

Muchas veces resulta interesante disponer en local de las secuencias de ADN y ARN de los genomas, en particular, del genoma humano, si tener que estar accediendo constantemente a las bases de datos en Internet (GenBank, etc.) que están bastante saturadas. En el HNRC, el Dr. Larry Parnell se dedica a realizar procesos de genómica comparativa entre genes de distintos organismos buscando las similitudes regionales y funcionales y, para ello, necesita disponer de los genomas completos de los organismos comparados (homo sapiens, mosca, ratón, rata, etc.)

La solución está en bajarse el genoma completo de cada organismo, agrupado en cromosomas, y almacenarlo en el puesto local. Pero el problema está en que los ficheros en cada cromosoma, sobre todo los grandes, son de un gran tamaño. Por ejemplo, desde el servidor FTP de GenBank (GenBank-Ftp, 2004) se pueden bajar los ficheros comprimidos de los cromosomas del genoma humano (y de todos los organismos ya genotipados). Estos ficheros, pese a estar comprimidos con el algoritmo gzip, ocupan aproximadamente entre 6 MB del cromosoma sexual Y y 70 MB del cromosoma 1. Cuando se descomprimen, los ficheros de texto correspondientes ocupan entre 11 y 330 MB.

Por supuesto, no hay utilidad o herramienta informática estándar en una plataforma ordinaria con MS Windows (Notepad, Wordpad, MS Word, otros editores de texto, etc.) que pueda abrir semejante fichero. Esto es debido a la estrategia ordinaria en aplicaciones en MS Windows, que es la de cargar en memoria todo el fichero y trabajar con accesos directos a memoria principal y así evitar los accesos a memoria secundaria, mucho más lentos. Por lo tanto, la necesidad planteada es doble: disponer de ficheros de secuencias de bases con la mayor compresión posible, y tener una herramienta que permita visualizar grandes ficheros e incluso seleccionar y copiar pequeños trozos de secuencias para pasarlas a otras aplicaciones y/o guardarlas en otros ficheros distintos.

#### 1.7.4.2 Obtención de ficheros de secuencias

Para poder hacer las pruebas correspondientes, se han obtenido los ficheros originales en formato FASTA de los cromosomas en el servidor de FTP de GenBank (GenBank-Ftp, 2004). Estos ficheros están comprimidos con el algoritmo gzip para que el tiempo de transmisión sea menor y, además, se controle la validez del fichero bajado desde la base de datos por el propio algoritmo unzip que aplica un control de código CRC.

Los ficheros originales de los cromosomas son una secuencia de una o mas secciones en formato FASTA. Este formato presenta una cabecera que empieza con el carácter “>” y termina con el carácter “\n” (carácter de nueva línea). Entre estos caracteres puede aparecer cualquier cadena de texto, incluyendo “T” y “U”, estos últimos se pueden confundir con los nucleótidos Timina o Uracilo. Así, un fichero de cromosoma está estructurado de la siguiente manera, donde cada línea es una sección:

```
CabeceraFASTA_1 Código_genético_1
CabeceraFASTA_2 Código_genético_2
...
CabeceraFASTA_i Código_Genético_i
...
CabeceraFASTA_n Código_Genético_n
```

Esta estructura de fichero puede resultar muy útil para establecer correspondencias con los gráficos que muestran los cromosomas en el *NCBI Map Viewer* del GenBank, pero es poco eficiente a la hora de aplicar un tratamiento de análisis, exploración o compresión con eficiencia.

#### 1.7.4.3 Codificación de caracteres

El principio básico del que se parte es que, desde el punto de vista del tratamiento de la información genómica, una cadena de ADN (a, c, g, t) o ARN (a, c, g, u) es una cadena de caracteres almacenada en un fichero de texto. En las arquitecturas y plataformas informáticas estándar, los caracteres se codifican internamente en una tira de un número fijo de bits, el byte, cuyo valor binario correspondiente está determinado por el sistema de codificación estándar empleado, ASCII o EBCDIC.



Por otra parte, el tamaño del byte se ha fijado en 8 bits (antes había de 9 y otros) de forma estándar para casi todas las arquitecturas de procesadores. Por lo tanto, cada uno de los caracteres que corresponde a una base ocupa un byte en un fichero de texto ordinario de forma que, si hay  $N$  bases en una secuencia, el fichero tiene un tamaño mínimo de  $N$  bytes (se deben sumar caracteres de control como, por ejemplo, el de fin de fichero y otros). Esto se muestra en la Figura 1.33 para el caso del sistema de codificación ASCII.

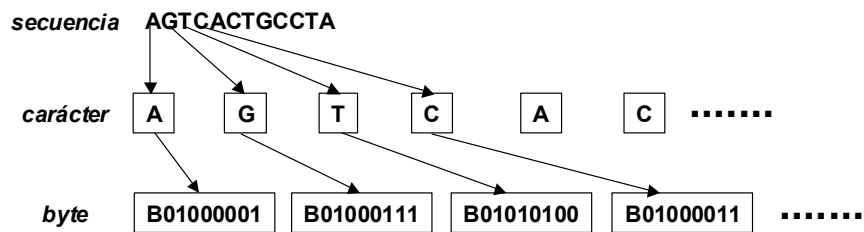


Figura 1.33. Codificación y almacenamiento ordinarios de las secuencias de bases

Este sistema es bastante eficiente cuando se maneja el conjunto tradicional de 256 caracteres ( $2^8$  combinaciones posibles de 8 bits que pueden tomar los valores “0” y “1”). Sin embargo, ¿qué ocurre si solamente se manejan cuatro caracteres sin distinguir mayúsculas o minúsculas: “a”, “c”, “g” “t” (o “u”)?

La respuesta es obvia: se desperdicia el 75% de cada byte porque solamente se necesitan dos posiciones para representar una base ( $2^2$  combinaciones posibles de 2 bits que pueden tomar los valores “0” y “1”). Entonces, con el mecanismo adecuado, un lenguaje de programación, estructuras de datos e instrucciones específicas, puede ser posible almacenar solamente dos bits por cada base obteniendo una compactación de cuatro bases por byte. Por ejemplo, se supone que se asignan los valores siguientes: “00” para “a” o “A”, “01” para “g” o “G”, “10” para “c” o “C”, “11” para “t” o “T” (“u” o “U”). Dada una secuencia determinada, “AGTCACTGCCTA” (o indistintamente “agtcactgcta”), el resultado de la compactación, donde se aplica un sencillo diccionario de conversión, se muestra en la Figura 1.34.

A priori, el tamaño del fichero resultante es de  $N/4$ , siendo  $N$  el tamaño en bases de la secuencia. Y la tasa de compactación o compresión es lógicamente de un 75% ( $((N - N/4)/N) \times 100$ ). Si además, se procesa la cadena resultante con un algoritmo *gzip*, que es uno de los algoritmos estándar de compresión de ficheros de caracteres, la tasa de compresión de la secuencia de nucleótidos y, del fichero que la contiene, puede ser mucho más alta en función de la repetición de bytes

compactados, es decir, grupos de cuatro bases (“ACTG”, “AGTC”, “AAGG”, etc.) en la secuencia original.

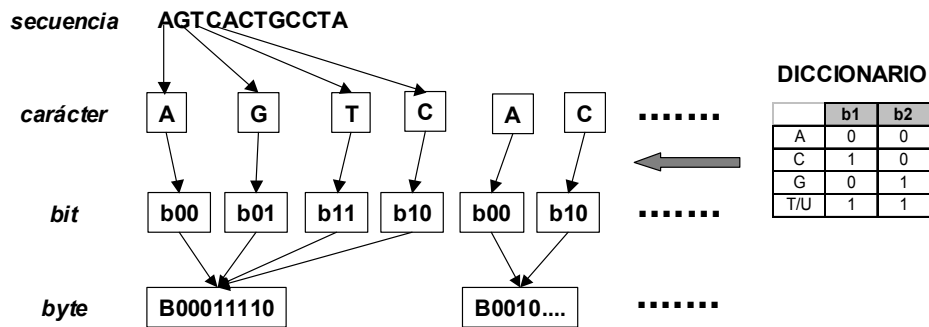


Figura 1.34. Compactación y almacenamiento de las secuencias de bases por cada byte

Con el diccionario utilizado se plantea otro pequeño problema: ¿qué ocurre con las secuencias no perfectas, resultantes de procesos de genotipado u otros, que tienen posiciones con bases sin identificar codificadas con los caracteres “K” o “N” (o indistintamente “k” o “n”)? Además, si se introduce un quinto carácter, ya no sirve el diccionario con dos bits y hay que ampliarlo a tres bits, donde se tiene un conjunto de ocho posibles combinaciones ( $2^3$  combinaciones posibles de 3 bits que pueden tomar los valores “0” y “1”). Esto lleva al planteamiento de un diccionario ampliado de tres bits con un mecanismo específico para tratarlo (Figura 1.35).

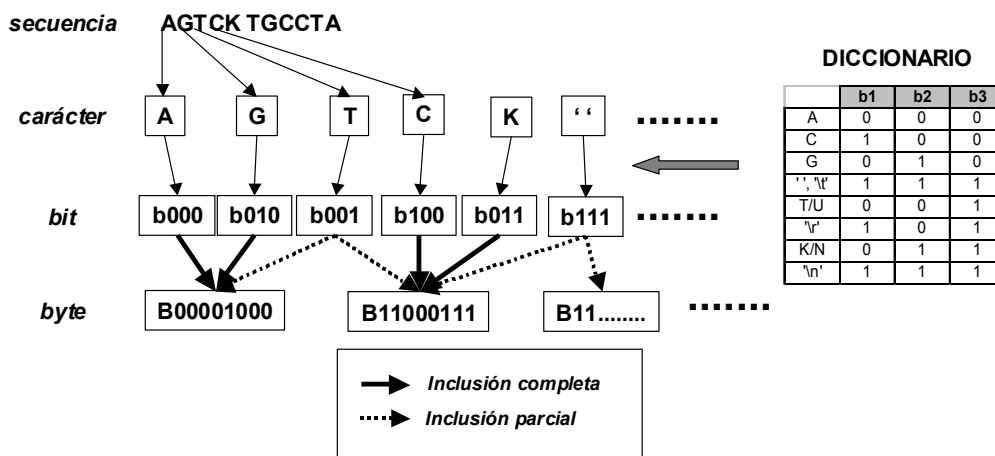


Figura 1.35. Compactación en cada byte de las secuencias de bases con tres bits por base

Por ello se ha propuesto la elaboración de unas utilidades de compresión, denominadas “Squeezer2B” y “Squeezer3B” que manejan bits en lugar de bytes. En el Capítulo 4 de Resultados se describe la solución propuesta en esta tesis y la comparación con algoritmos estándar.

### 1.7.5 Calidad y seguridad de los procesos

Hay que considerar un enfoque distinto de la disciplina de la Bioinformática con respecto a los aspectos de aseguramiento de la calidad y la seguridad (Cass et al., 2000). No se puede olvidar que la Bioinformática es una disciplina en parte tecnológica y en parte científica (Science News, 2001), con lo que se puede exigir que sus procesos y resultados, por una parte cumplan determinados requisitos de seguridad y fiabilidad, y por la otra sigan rigurosamente el método científico (Shortlife et al., 2002).

Existen instrumentos de control de las funciones de Tecnología de Información que normalmente son examinados y verificados por auditores informáticos. Estos aplican principios y normas de auditoría generales, pero técnicas y herramientas específicas. Por la otra parte, existen procedimientos formales para verificar si en una disciplina se sigue o no el método científico. De hecho, este es uno de los requisitos básicos que debe cumplir cualquier trabajo que se envíe a publicar en las revistas científicas. Entonces, la situación actual observada es que no existen, o son escasos y no están homologados, instrumentos de control de las funciones biomédicas en general y bioinformáticas en particular, dentro de unas disciplinas clave en múltiples proyectos relacionados con investigaciones médicas, agronómicas, alimentarias, fitosanitarias y zoonosanitarias, etc. Dichos instrumentos deberían tener como requisito principal la capacidad de analizar disciplinas que suelen combinar habitualmente enfoques tecnológicos y científicos. Además, tampoco existe una metodología de auditoría que permita inspeccionar y verificar, tanto las funciones bioinformáticas, como sus instrumentos de control.

Con estas consideraciones, la visión de la Bioinformática que tienen sus primeros creadores e impulsores, los biólogos moleculares con aptitudes para desarrollar algoritmos y programas de ordenador (Science News, 2001), debe ser sustituida por una visión más formal como la que aportan los ingenieros informáticos que lentamente van asumiendo cada vez más responsabilidades bioinformáticas (Hass et al., 2000). A pesar de ello, se debe llegar a un equilibrio en el grado de colaboración entre bioinformáticos, cuya formación básica es la Biología Molecular o similar, o los que provienen de la Ingeniería Informática. De todas formas, cuando la formación especializada en Bioinformática como titulación

propia esté consolidada, los problemas de competencia quedarán sensiblemente reducidos.

#### 1.7.5.1 Auditoría y control en Bioinformática

Tomando el caso específico de la Bioinformática, como disciplina de apoyo en las investigaciones biomédicas, se ha detectado la enorme necesidad de la implantación de sistemas de control y supervisión con la misma seriedad que en cualquier otro campo de aplicación de la Tecnología de Información (Sackman, 1997). Hay que destacar que en los proyectos en que interviene la Bioinformática se están invirtiendo ingentes cantidades de fondos públicos y privados, se manejan equipos e instrumentos de alto coste, se emplean especialistas con sólida formación y, además, se trata información sensible.

En este sentido, la investigación en Biomedicina en la actualidad exige la concurrencia de recursos humanos y tecnológicos específicos (Rondel et al., 1993) (Sackman, 1997). La formación de los recursos humanos y la complejidad y sofisticación del resto hace que los presupuestos manejados por los laboratorios de investigación sean muy altos (10) y los problemas muy diversos (Cass et al., 2002) (Rindfleisch, 1997). Sin embargo, los instrumentos de control y optimización de dichos recursos no están muy desarrollados en este campo (Rindfleisch, 1997).

En otro orden de cosas, existe la *Information Systems Audit and Control Association* (ISACA), que es una asociación profesional de ámbito mundial para la regulación de la práctica de la Auditoría de Sistemas de Información (ISACA, 2004). ISACA ha publicado en 1997, y mantiene un procedimiento de revisión y renovación, un marco metodológico formal denominado COBIT (*Control Objectives for Information and related Technology* - Objetivos de Control para la Información y Tecnologías Afines), que está ampliamente aceptado por la comunidad internacional de auditores de sistemas de información como una norma estándar.

Dado el carácter multidisciplinario de la Bioinformática y en función de la agrupación de problemas que se muestra en la Tabla 1.4, en la auditoría de esta disciplina concurren distintos aspectos clásicos a estudiar: organización y planificación, diseño y desarrollo de sistemas de información, adquisición de recursos, seguridad física y lógica, protección de datos, calidad de servicio, rendimiento y eficiencia, etc.

**Tabla 1.4A. Elementos del Espacio del Problema de Auditoría y su relación con las Áreas y Problemas de Bioinformática**

| Elemento del Espacio del Problema de Auditoría | Código | Áreas y Problemas de Bioinformática |                         |                       |                                   |                                   |                                    |                                     |                              |                                 |                    |                                |                        |
|--|--------|-------------------------------------|-------------------------|-----------------------|-----------------------------------|-----------------------------------|------------------------------------|-------------------------------------|------------------------------|---------------------------------|--------------------|--------------------------------|------------------------|
|  |        | Análisis de secuencias              | Análisis de estructuras | Análisis de funciones | Simulación de procesos biológicos | Almacenamiento y gestión de datos | Gestión del conocimiento biológico | Adquisición en intercambio de datos | Análisis e interpretación de | Seguridad y protección de datos | Nuevas tecnologías | Nuevos enfoques de computación | Organización y Gestión |
| Teoría científica biológica                    | EAB01  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   | ✓                                  |                                     | ✓                            |                                 |                    |                                |                        |
| Teoría científica informática                  | EAB02  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   | ✓                                  |                                     | ✓                            |                                 | ✓                  | ✓                              |                        |
| Metodología informática                        | EAB03  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            |                                 | ✓                  |                                |                        |
| Técnica experimental                           | EAB04  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   |                                    | ✓                                   |                              |                                 |                    |                                |                        |
| Protocolo experimental                         | EAB05  | ✓                                   | ✓                       | ✓                     |                                   |                                   |                                    | ✓                                   |                              |                                 |                    |                                |                        |
| Formalización de datos                         | EAB06  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            |                                 |                    |                                | ✓                      |
| Interpretación de datos                        | EAB07  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   |                                    |                                     | ✓                            |                                 |                    |                                |                        |
| Datos sensibles                                | EAB08  |                                     |                         |                       |                                   | ✓                                 | ✓                                  | ✓                                   |                              | ✓                               |                    |                                | ✓                      |
| Uso de los datos                               | EAB09  |                                     |                         |                       |                                   | ✓                                 | ✓                                  | ✓                                   |                              | ✓                               |                    |                                | ✓                      |
| Bioética                                       | EAB10  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            | ✓                               |                    |                                | ✓                      |
| Seguridad                                      | EAB11  |                                     |                         |                       |                                   | ✓                                 | ✓                                  | ✓                                   | ✓                            | ✓                               | ✓                  |                                | ✓                      |
| Eficiencia                                     | EAB12  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            | ✓                               | ✓                  |                                | ✓                      |
| Calidad  | EAB13  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            | ✓                               | ✓                  | ✓                              | ✓                      |
| Sistemas SW                                    | EAB14  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            | ✓                               | ✓                  |                                |                        |
| Sistemas HW                                    | EAB15  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                   | ✓                            | ✓                               | ✓                  |                                |                        |
| Sistemas instrumentales                        | EAB16  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   |                                    |                                     | ✓                            |                                 |                    |                                |                        |
| Políticas de investigación                     | EAB17  |                                     |                         |                       |                                   |                                   |                                    |                                     |                              |                                 |                    |                                | ✓                      |
| Sistemas de investigación                      | EAB18  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   | ✓                                  |                                     |                              |                                 | ✓                  |                                | ✓                      |
| Políticas de gestión                           | EAB19  |                                     |                         |                       |                                   |                                   | ✓                                  |                                     |                              |                                 |                    |                                | ✓                      |

**Tabla 1.4B. Elementos del Espacio del Problema de Auditoría y su relación con las Áreas y Problemas de Bioinformática (continuación)**

| Elemento del Espacio del Problema de Auditoría | Código | Áreas y Problemas de Bioinformática |                         |                       |                                   |                                   |                                    |                                    |                                    |                                 |                    |                                |                        |
|--|--------|-------------------------------------|-------------------------|-----------------------|-----------------------------------|-----------------------------------|------------------------------------|------------------------------------|------------------------------------|---------------------------------|--------------------|--------------------------------|------------------------|
|  |        | Análisis de secuencias              | Análisis de estructuras | Análisis de funciones | Simulación de procesos biológicos | Almacenamiento y gestión de datos | Gestión del conocimiento biológico | Adquisición e intercambio de datos | Análisis e interpretación de datos | Seguridad y protección de datos | Nuevas tecnologías | Nuevos enfoques de computación | Organización y Gestión |
| Sistemas de gestión                            | EAB20  |                                     |                         |                       |                                   | ✓                                 | ✓                                  |                                    |                                    |                                 |                    |                                | ✓                      |
| Tecnologías biológicas                         | EAB21  | ✓                                   | ✓                       | ✓                     | ✓                                 |                                   |                                    |                                    |                                    |                                 | ✓                  |                                |                        |
| Tecnologías de la información                  | EAB22  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                  | ✓                                  | ✓                               | ✓                  |                                | ✓                      |
| Capacidad profesional                          | EAB23  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                  | ✓                                  | ✓                               | ✓                  | ✓                              | ✓                      |
| Formación específica                           | EAB24  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                  | ✓                                  | ✓                               | ✓                  | ✓                              | ✓                      |
| Capacidad de coordinación                      | EAB25  | ✓                                   | ✓                       | ✓                     | ✓                                 | ✓                                 | ✓                                  | ✓                                  | ✓                                  | ✓                               | ✓                  | ✓                              | ✓                      |

Sin embargo, también hay otros aspectos que normalmente no se tratan en una auditoría tradicional: bioética, metodología científica, protocolos experimentales, estructuración y análisis de datos, etc. Por lo tanto, dos de los objetivos planteados en esta tesis, correspondientes a esta parte, han sido los siguientes:

- Encontrar a través de COBIT instrumentos de auditoría sobre las funciones bioinformáticas para confeccionar una propuesta de marco metodológico adaptado para la Auditoría Bioinformática, teniendo en cuenta los aspectos científicos, tecnológicos y éticos de la disciplina.
- La realización de una auditoría de sistemas de información en un laboratorio de investigación biomédica para obtener el diseño de un instrumento de control en la organización y aplicación de los sistemas informáticos como soporte de los procesos científicos.

#### 1.7.5.2 Planteamiento inicial de la auditoría

Para establecer el punto de partida de la auditoría sobre la Bioinformática, es necesario determinar el contexto del problema y el alcance del estudio. Esto ayuda a eliminar los aspectos irrelevantes para el estudio. A continuación, se deben

identificar y caracterizar los elementos del problema a auditar, que se agrupan en el Espacio del Problema de Auditoría. A partir de este espacio, se debe determinar si todos los elementos son elementos de riesgo o no y los riesgos asociados a los mismos. Esto se hace mediante un análisis preliminar de riesgos. Finalmente, aplicando un análisis de riesgos más detallado, se deben obtener las posibles soluciones de control y objetivos de auditoría. La realización de este proceso debe mostrar si los instrumentos de ASI son adecuados o no para el problema establecido. Dado que este documento debe respetar una extensión establecida por el contexto científico en el que se enmarca, no se describen exhaustivamente todos los pasos y aspectos del proceso, sino los más destacables.

El contexto del problema es la aplicación de una auditoría de sistemas de información global a la propia disciplina de Bioinformática en la forma que se ha caracterizado en la sección primera, teniendo en cuenta que ha surgido como la concurrencia de diversas disciplinas científicas y tecnológicas. Sin embargo, dado que su contribución es mayor que el resto, se restringe el contexto a las disciplinas Biología Molecular y Genética, Ciencia de la Computación e Ingeniería Informática. La auditoría global se desglosa en auditorías sobre áreas y en auditorías tecnológicas (Bernal et al., 1999) (Piattini et al., 1998).

El alcance del estudio consiste en la auditoría sobre la aplicación de teorías y metodologías científicas, técnicas y protocolos experimentales, metodologías y técnicas de ingeniería, tecnologías y políticas e instrumentos de investigación y gestión. Todo ello restringido a la Bioinformática relacionada con el estudio de individuos de la especie humana, que se podría denominar Bioinformática Humana.

Por tanto, el objetivo de la auditoría es la caracterización de un proceso de auditoría arquetípico sobre la Bioinformática Humana para descubrir si el marco metodológico es o no adecuado.

La Tabla 1.4 establece la relación entre elementos del Espacio y las categorías o subcategorías de problemas identificados en la Bioinformática. A cada uno de los elementos se le ha asignado un código que lo representará en los siguientes pasos. El sistema de codificación aplicado toma las iniciales “EAB”, acrónimo de “Espacio de Auditoría Bioinformática”, y añade un número secuencial de dos cifras.

## 1.8 Bibliografía

### 1.8.1 Bibliografía de Bioinformática y disciplinas asociadas

- Andrieu N., Goldstein AM. "Epidemiologic and genetic Approaches in the Study of Gene-environment Interaction: an Overview of Available Methods". *Epidemiol Rev*, 1998; 20, 2: 137-147.
- Backofen R., Gilbert D. "Bioinformatics and Constraints". *Constraints*, 6; 2001: 141- 156.
- Collins F. S., Green E. D., Guttmacher A. E., Guyer M. S. "A vision for the future of genomics research A blueprint for the genomic era". *Nature* 2003, 422: 835-847.
- Collins F. S., Morgan M., Patrinos A. "The Human Genome Project: Lessons from Large-Scale Biology". *Science*, 2003, 300: 286-290.
- Coltell O., Corella D. "Bioinformática y Epidemiología Genómica". *I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud*; 2002, 37: 15-26.
- Cook S. A., Rosenzweig A. "DNA Microarrays. Implications for Cardiovascular Medicine". *Circ Res*. 2002;91:559-564.
- INBIOMED. Red Temática de Investigación Cooperativa de Informática Biomédica. <http://www.inbiomed.retics.net>.
- Jackson D.G., Healy M.D., Davison D.B. Bioinformatics: not just for sequences anymore. *BIOSILICO* 2003; I(3): 103-111.
- Jones R.L. "The Internet and Healthcare Information Systems: How Safe Will Patient Data Be?". *IS Audit & Control Journal*, I; 1998: 25-30.
- Khoury M. "Genetic Epidemiology". Rothman K.J., Greenland S. (eds.) *Modern Epidemiology* (2nd ed.) Little, Brown and Co., Boston, MA (USA), 1997.
- Khoury M., Beaty T.H., Cohen B.H. *Fundamentals of Genetic Epidemiology*. Oxford University Press, New York, 1993.
- Marrugat J., Solanas P., D'Agostino R., Sullivan L., Ordovás J., et al. "Coronary risk estimation in Spain using a calibrated framingham function". *Rev Esp Cardiol* 2003; 56: 253-261.
- Martín Sánchez F., López V., Sánchez J. P., Liébana I. Red Temática de Investigación Cooperativa en Informática Biomédica, INBIOMED. *Revista I+S. Informática y Salud*, 2004, 46: 7-13.
- Martín-Sánchez F. et al. Facilitating Genomic Medicine for Future Healthcare. *J. Biomed. Inform.* Artículo en prensa, doi:10.1016/j.jbi.2003.09.003 .
- McKusick V.A. "The Human Genome Organisation: History, Purposes and Membership". *Genomics*, 5; 1989: 385-387.



- Ordovás J. M. et al. "Identification and molecular analysis of two apob gene mutations causing low plasma cholesterol levels". *Circulation*, 92, 1995: 2036-2040.
- Ordovás JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on HDL metabolism: Evidence of a strong dose-effect in this gene-nutrient interaction in the Framingham Study". *Circulation*, 2002; 18: 2315-2321.
- Roberts R. "Bioinformatics Analysis of Gene Banks Provides a Treasure Trove for the Functional Genomist". *J. Mol. Cell. Cardiol.*; 32; 2000: 1917-1919.
- Sackman H. *Biomedical Information Technology. Global Social Responsibilities for the Democratic Age.* Academic Press, San Diego, CA (USA), 1997.
- Science's News staff. "A History of the Human Genome Project". *Science Magazine*; 291 (5507); 2001: 1195-1261.
- Van Bemmelen J.H., Musen M.A.(eds.) *Handbook of Medical Informatics.* Springer-Verlag, Heidelberg, 1997.
- Wilson PWF., D'Agostino RB., Levy D., Belanger A., Silbershatz H., Kannel W. "Prediction of Coronary Heart Disease Using Risk Factor Categories". *Circulation*, 1998; 97,18: 1837-1847.

## 1.8.2 Bibliografía de CETP

- Drayna D, Jamagin AS, McLean J, Henzel W, Kohr W, Fielding C, et al. Cloning and sequencing of human cholesteryl ester transfer protein cDNA. *Nature* 1987; 327:632-634.
- Drayna D, Lawn R. Multiple RFLP's at the human cholesteryl ester transfer protein (CETP) locus. *Nucleic Acids Res* 1987; 15:4698.
- Freeman D, Shepherd J, Packard CJ, Humphries SE, Gaffney D. An *StuI* RFLP at the human cholesteryl ester transfer protein (CETP) locus. *Nucleic Acids Res* 1989; 17:2880.
- Freeman DJ, Griffin BA, Holmes AP, Lindsay GM, Gaffney D, Packard CJ, et al. Regulation of plasma HDL cholesterol and subfraction distribution by genetic and environmental factors: Associations between the *TaqI* B RFLP in the CETP gene and smoking and obesity. *Arterioscler Thromb* 1994; 14:336-344.
- Fumeron F, Betoulle D, Luc G, Behague I, Ricard B, Poirier O, et al. Alcohol intake modulates the effect of a polymorphism of the cholesteryl ester transfer protein gene on plasma high density lipoprotein and the risk of myocardial infarction. *J Clin Invest* 1995; 96:1664-1671.

- Gordon DJ, Rifkind BM. High-density lipoprotein. The clinical implications of recent studies. *N Engl J Med* 1989; 321:1311-1316.
- Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* 1977; 62:707-714.
- Hannuksela ML, Liinamaa MJ, Kesäniemi YA, Savolainen MJ. Relation of polymorphisms in the cholesteryl ester transfer protein gene to transfer protein activity and plasma lipoprotein levels in alcohol drinkers. *Atherosclerosis* 1994; 110:35-44.
- Hill SA, Nazir DJ, Jayaratne P, Bamford KS, McQueen MJ. Mutations in cholesteryl ester transfer protein and hepatic lipase in a North American population. *Clin Biochem* 1997; 30:413-418.
- Inazu A, Brown ML, Hesler CB, Agellon LB, Koizumi J, Takata K, et al. Increased high-density lipoprotein levels caused by a common cholesteryl ester transfer protein gene mutation. *N Engl J Med* 1990; 323:1234-1238.
- Inazu A, Jiang X-C, Haraki T, Yagi K, Kamon N, Koizumi J, et al. Genetic cholesteryl ester transfer protein deficiency caused by two prevalent mutations as a major determinant of increased levels of high density lipoprotein cholesterol. *J Clin Invest* 1994; 94:1872-1882.
- Kauma H, Savolainen MJ, Heikkilä R, Rantala AO, Lilja M, Reunanen A, et al. Sex difference in the regulation of plasma high density lipoprotein cholesterol by genetic and environmental factors. *Hum Genet* 1996; 97:156-162.
- Koizumi J, Inazu A, Yagi K, Koizumi I, Uno Y, Kajinami K, et al. Serum lipoprotein lipid concentration and composition in homozygous and heterozygous patients with cholesteryl ester transfer protein deficiency. *Atherosclerosis* 1991; 90:189-196.
- Kondo I, Berg K, Drayna DT, Lawn RM. DNA polymorphism at the locus for human cholesteryl ester transfer protein (CETP) is associated with high density lipoprotein cholesterol and apolipoprotein levels. *Clin Genet* 1989; 35:49-56.
- Kuivenhoven JA, de Knijff P, Boer JMA, Smalheer HA, Botma GJ, Seidell JC, et al. Heterogeneity at the CETP gene locus - Influence on plasma CETP concentrations and HDL cholesterol levels. *Arterioscler Thromb Vasc Biol* 1997; 17:560-568.
- Kuivenhoven JA, Jukema JW, Zwinderman AH, de Knijff P, McPherson R, Bruschke AV, et al. The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. The Regression Growth Evaluation Statin Study Group. *N Engl J Med* 1998; 338:86-93.
- Manninen V, Elo MO, Frick MH, Haapa K, Heinonen OP, Heinsalmi P, et al. Lipid alterations and decline in the incidence of coronary heart disease in the Helsinki Heart Study. *JAMA* 1988; 260:641-651.
- Mitchell RJ, Earl L, Williams J, Bisucci T, Gasiamis H. Polymorphisms of the gene coding for the cholesteryl ester transfer protein and plasma lipid levels in Italian and Greek migrants to Australia. *Human Biology* 1994; 66:13-25.

- Takegoshi T, Haba T, Kitoh C, Inazu A, Koizumi J, Mabuchi H, et al. Compound heterozygote of cholesteryl ester transfer protein deficiency in a patient with hyperalphalipoproteinemia. *Atherosclerosis* 1992; 96:83-85.
- Tamminen M, Kakko S, Kesäniemi YA, Savolainen MJ. A polymorphic site in the 3' untranslated region of the cholesteryl ester transfer protein (CETP) gene is associated with low CETP activity. *Atherosclerosis* 1996; 124:237-247.
- Tenkanen H, Koskinen P, Kontula K, Aalto-Setälä K, Mänttari M, Manninen V, et al. Polymorphisms of the gene encoding cholesterol ester transfer protein and serum lipoprotein levels in subjects with and without coronary heart disease. *Hum Genet* 1991; 87:574-578.
- Yamashita S, Sakai N, Hirano K, Arai T, Ishigami M, Maruyama T, et al. Molecular genetics of plasma cholesteryl ester transfer protein. *Curr Opin Lipidol* 1997; 8:101-110.
- Zuliani G, Hobbs HH. EcoNI polymorphism in the human cholesteryl ester transfer protein (CETP) gene. *Nucleic Acids Res* 1990; 18:2834.

### 1.8.3 Bibliografía de APOE

- Aguilar CA, Talavera G, Ordovas JM, et al. The apolipoprotein E4 allele is not associated with an abnormal lipid profile in a Native American population following its traditional lifestyle. *Atherosclerosis* 1999;142:409-14.
- Barter PJ. Cholesterol and cardiovascular disease: Basic science. *Aust N Z J Med* 1994;24:83-88.
- Bétard C, Kessling AM, Roy M, Davignon J. Influence of genetic variability in the nondeletion LDL- receptor allele on phenotypic variation in French-Canadian familial hypercholesterolemia heterozygotes sharing a 'null' LDL-receptor gene defect. *Athero* 1996;119:43-55.
- Bijnen FCH, Feskens EJM, Caspersen CJ, Giampaoli S, Nissinen AM, Menotti A, Mosterd WL, Kromhout D. Physical activity and cardiovascular risk factors among elderly men in Finland, Italy, and the Netherlands. *Am J Epidemiol* 1996;143:553-561.
- Bodurtha JN, Chen CW, Mosteller M, Nance WE, Schieken RM, Segrest J. Genetic and environmental contributions to cholesterol and its subfractions in 11-year-old twins: The Medical College of Virginia twin study. *Arterioscler Thromb* 1991;11:844-850.
- Boerwinkle E, Utermann G. Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am J Hum Genet* 1988;42:104-12.
- Braeckman L, De Bacquer D, Rosseneu M, De Backer G. Apolipoprotein E polymorphism in middle-aged Belgian men: Phenotype distribution and relation to serum lipids and lipoproteins. *Atherosclerosis* 1996;120:67-73.

- Breslow JL, McPherson J, Nussbaum AL, Williams HW, Lofquist-Kahl F, Karathanasis SK, Zannis VI. Identification and DNA sequence of a human apolipoprotein E cDNA clone. *J Biol Chem* 1982;257:14639-14641.
- Connelly PW, Maguire GF, Lee M, Little JA. Plasma lipoproteins in familial hepatic lipase deficiency. *Arterio* 1990;10:40-48.
- Cumming AM, Robertson FW. Polymorphism at the apoprotein-E locus in relation to risk of coronary disease. *Clin Genet* 1984;25:310-313.
- Davignon J, Cohn JS, Mabile L, Bernier L. Apolipoprotein E and atherosclerosis: insight from animal and human studies. *Clin Chim Acta* 1999;286:115-143.
- Davignon J, Gregg RE, Sing CF. Apolipoprotein E polymorphism and atherosclerosis. *Arterio* 1988;8:1-21.
- Deiana L, Pes GM, Carru C, Errigo A, Pettinato S, Carcassi C, Baggio G, Contu L. Lack of influence of apolipoprotein E4 on lipoprotein levels in the island population of Sardinia. *Eur J Clin Invest* 1998;28:290-294.
- Eichner JE, Kuller LH, Ferrell RE, Meilahn EN, Kamboh MI. Phenotypic effects of apolipoprotein structural variation on lipid profiles.III. Contribution of apolipoprotein E phenotype to prediction of total cholesterol, apolipoprotein B, and low density lipoprotein cholesterol in the healthy women study. *Arteriosclerosis* 1990; 10: 379-85.
- Ellsworth DL, Sholinsky P, Jaquish C, Fabsitz RR, Manolio TA. Coronary heart disease. At the interface of molecular genetics and preventive medicine. *Am J Prev Med* 1999;16:122-133.
- Fujishima M, Kiyohara Y, Ueda K, Hasuo Y, Kato I, Iwamoto H. Smoking as cardiovascular risk factor in low cholesterol population: The Hisayama Study. *Clin Exp Hypertens [A]* 1992;14A:99-108.
- Gaziano JM, Buring JE, Breslow JL, et al. Moderate alcohol intake, increased levels of high-density lipoprotein and its subfractions, and decreased risk of myocardial infarction. *N Engl J Med* 1993;329:1829-34.
- Gerdes LU, Klausen IC, Sihtm I, Faergeman O. Apolipoprotein E polymorphism in a Danish population compared to findings in 45 other study populations around the World. *Genet Epidemiol* 1992;9:155-167.
- Glueck CJ, Hogg E, Allen C, Gartside PS. Effects of alcohol ingestion on lipids and lipoproteins in normal men: isocaloric metabolic studies. *Am J Clin Nutr* 1980;33:2287-93.
- Gómez-Coronado D, Alvarez JJ, Entrala A, Olmos JM, Herrera E, Lasuncion MA. Apolipoprotein E in men and women from a Spanish population: allele frequencies and influence on plasma lipids and apolipoproteins. *Athero* 1999;147:167-176.
- Gronbaek M, Deis A, Becker U, et al. Alcohol and mortality: is there a U-shaped relation in elderly people? *Age & Ageing* 1998;27:739-44.

- Hallman DM, Boerwinkle E, Saha N, Sandholzer C, Jürgen Menzel H, Császár A, Utermann G. The apolipoprotein E polymorphism: A comparison of allele frequencies and effects in nine populations. *Am J Hum Genet* 1991;49:338-349.
- Hegele RA. Gene-environment interactions in atherosclerosis. *Mol Cell Biochem* 1992;113:177-186.
- Hein HO, Suadicani P, Gyntelberg F. Alcohol consumption, serum low density lipoprotein cholesterol concentration, and risk of ischaemic heart disease: Six year follow up in the Copenhagen male study. *Brit Med J* 1996;312:736-41.
- Hixson JE, PDAY Research Group. Apolipoprotein E polymorphisms affect atherosclerosis in young males. *Arterioscler Thromb* 1991;11:1237-1244.
- Jarvik GP, Austin MA, Fabsitz RR, et al. Genetic influences on age-related change in total cholesterol, low density lipoprotein-cholesterol, and triglyceride levels: Longitudinal apolipoprotein E genotype effects. *Genet Epidemiol* 1994;11:375-84.
- Kamboh MI, Aston CE, Hamman RF. The relationship of APOE polymorphism and cholesterol levels in normoglycemic and diabetic subjects in a biethnic population from the San Luis Valley, Colorado. *Athero* 1995;112:145-159.
- Kamboh MI, Evans RW, Aston CE. Genetic effect of apolipoprotein(a) and apolipoprotein E polymorphisms on plasma quantitative risk factors for coronary heart disease in American black women. *Athero* 1995;117:73-81.
- Keil U, Chambless LE, Döring A, Filipiak B, Stieber J. The relation of alcohol intake to coronary heart disease and all-cause mortality in a beer-drinking population. *Epidemiology* 1997;8:150-6.
- Kiechl S, Willeit J, Rungger G, Egger G, Oberhollenzer F, Bonora E. Alcohol consumption and atherosclerosis: what is the relation? Prospective results from the Bruneck Study. *Stroke* 1998;29:900-7.
- Lahoz C, Ordovás JM. ApoE: lípidos plasmáticos, cardiopatía isquémica y enfermedad de Alzheimer. *Med Clin (Barc)* 1997;109:31-36.
- Langer RD, Criqui MH, Reed DM. Lipoproteins and blood pressure as biological pathways for effect of moderate alcohol consumption on coronary heart disease. *Circulation* 1992;85:910-5.
- Lehtimäki T, Frankberg-Lakkala H, Solakivi T, et al. The effect of short-term fasting, apolipoprotein E gene polymorphism, and sex on plasma lipids. *Am J Clin Nutr* 1997;66:599-605.
- Lopez-Miranda J, Ordovas JM, Mata P, et al. Effect of apolipoprotein E phenotype on diet-induced lowering of plasma low density lipoprotein cholesterol. *J Lipid Res* 1994;35:1965-75.
- Lucotte G, Loirat F, Hazout S. Pattern of gradient of apolipoprotein E allele \*4 frequencies in western Europe. *Human Biology* 1997;69:253-262.

- McConnell MW, Vavouranakis I, Wu LL, Vaughan DE, Ridker PM. Effects of a single, daily alcoholic beverage on lipid and hemostatic markers of cardiovascular risk. *Am J Cardiol* 1997;80:1226-28.
- McElduff P, Dobson AJ. How much alcohol and how often? population based case-control study of alcohol consumption and risk of major coronary event. *Brit Med J* 1997;314:1159-64.
- Muros M, Rodríguez-Ferrer C. Apolipoprotein E polymorphism influence on lipids, apolipoproteins and Lp(a) in a Spanish population underexpressing apo E4. *Athero* 1996;121:13-21.
- Myers RH, Schaefer EJ, Wilson PWF, D'Agostino R, Ordovas JM, Espino A, Au R, White RF, Knoefel JE, Cobb JL, McNulty KA, Beiser A, Wolf PA. Apolipoprotein E e4 association with dementia in a population-based study: The Framingham study. *Neurology* 1996;46:673-677.
- Nakanishi N, Nakamura K, Ichikawa S, Suzuki K, Tatara K. Relationship between lifestyle and serum lipid and lipoprotein levels in middle-aged Japanese men. *Eur J Epidem* 1999;15:341-8.
- Pablos-Mendez A, Mayeux R, Ngai C, Shea S, Berglund L. Association of apo E polymorphism with plasma lipid levels in a multiethnic elderly population. *Arterioscler Thromb Vasc Biol* 1997;17:3534-41.
- Pérez-Estrada E, Escolar JL, Valdivielso P, Andrade R, Pérez-Estrada M, González-Santos P. Fenotipos de la apoproteína E. Un estudio de la población de Málaga. *Rev Clin Esp* 1990;187:3-5.
- Rakic V, Puddey IB, Dimmitt SB, Burke V, Beilin LJ. A controlled trial of the effects of pattern of alcohol intake on serum lipid levels in regular drinkers. *Atherosclerosis* 1998;137:243-52.
- Reina M, Brunzell JD, Deeb SS. Molecular basis of familial chylomicronemia: Mutations in the lipoprotein lipase and apolipoprotein C-II genes. *J Lipid Res* 1992;33:1823-1832.
- Renaud SC, Gueguen R, Schenker J, d'Houtaud A. Alcohol and mortality in middle-aged men from eastern France. *Epidemiology* 1998;9:184-8.
- Rimm EB, Klatsky A, Grobbee D, Stampfer MJ. Review of moderate alcohol consumption and reduced risk of coronary heart disease: Is the effect due to beer, wine, or spirits. *Brit Med J* 1996;312:731-6.
- Rimm EB, Williams P, Criqui M, Stampfer MJ. Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. *Brit Med J* 1999;319:1523-8.
- Sing CF, Davignon J. Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 1985;37:268-285.

- Sytkowski PA, Kannel WB, D'Agostino RB. Changes in risk factors and the decline in mortality from cardiovascular disease. The Framingham Heart Study. *N Engl J Med* 1990;322:1635-1641.
- Talmud P, Tybjaerg-Hansen A, Bhatnagar D, Mbewu A, Miller JP, Durrington P, Humphries S. Rapid screening for specific mutations in patients with a clinical diagnosis of familial hypercholesterolaemia. *Athero* 1991;89:137-141.
- Tiret L, de Knijff P, Menzel H-J, Ehnholm C, Nicaud V, Havekes LM. ApoE polymorphism and predisposition to coronary heart disease in youths of different European populations: The EARS Study. *Arterioscler Thromb* 1994;14:1617-1624.
- Utermann G. Apolipoprotein E polymorphism in health and disease. *Am Heart J* 1987;113:433-440.
- Vasisth S, Pant MC, Srivastava LM. Effect of alcohol on serum lipids & lipoproteins in male drinkers. *Indian J Med Res* 1992;96:333-7.
- Wilson PWF, Myers RH, Larson MG, Ordovas JM, Wolf PA, Schaefer EJ. Apolipoprotein E alleles, dyslipidemia, and coronary heart disease: The Framingham Offspring Study. *JAMA* 1994;272:1666-1671.

#### 1.8.4 Bibliografía de APOA1

- Akita H, Chiba H, Tsuji M, Hui SP, Takahashi Y, Matsuno K, Kobayashi K. Evaluation of G-to-A substitution in the apolipoprotein A-I gene promoter as a determinant of high-density lipoprotein cholesterol level in subjects with and without cholesteryl ester transfer protein deficiency. *Hum Genet* 1995;96:521-526.
- Barre DE, Guerra R, Verstraete R, Wang Z, Grundy SM, Cohen JC. Genetic analysis of a polymorphism in the human apolipoprotein A-I gene promoter: Effect on plasma HDL-cholesterol levels. *J Lipid Res* 1994;35:1292-1296.
- Bruckert E, Von Eckardstein A, Funke H, Beucler I, Wiebusch H, Turpin G, Assmann G. The replacement of arginine by cysteine at residue 151 in apolipoprotein A-I produces a phenotype similar to that of apolipoprotein A-IMilano. *Athero* 1997;128:121-128.
- Civeira F, Pocovi M, Cenarro A, Garces C, Ordovas JM. Adenine for guanine substitution - 78 base pairs to the apolipoprotein(APO) A-I gene: relation with high density lipoprotein cholesterol and apoA-I concentrations. *Clin Genet* 1993;44:307-312.
- Deckelbaum RJ, Fisher EA, Winston M, Kumanyika S, Laurer RM, Pi-Sunyer FX, Jeor S, Schaefer EJ, Weinstein IB. Summary of a scientific conference on preventive nutrition: pediatrics to geriatrics. *Circ* 1999;100:450-456.
- Fielding CJ, Shore VG, Fielding PE. A protein co-factor of lecithin:cholesterol acyltransferase. *Biochem Biophys Res Comm* 1972;46:1493-1498.

- Jeenah M, Kessling A, Miller N, Humphries SE. G to A substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI and high density lipoprotein cholesterol concentrations. *Mol Biol Med* 1990;7:233-241.
- Jump DB, Clarke SD, Thelen A, Liimatta M, Ren B, Badin M. Dietary fat, genes, and human health. *Adv Exp Med Biol* 1997;422:176.
- Kamboh MI, Aston CE, Nestlerode CM, McAllister AE, Hamman RF. Haplotype analysis of two APOA1/MspI polymorphisms in relation to plasma levels of apo A-I and HDL-cholesterol. *Athero* 1996;127:255-262.
- Karathanasis SK. Apolipoprotein multigene family: tandem organization of human apolipoprotein A-I, C-III and A-IV genes. *Proc Natl Acad Sci USA* 1985;82:6374-6378.
- Mata P, Lopez-Miranda J, Pocovi M, Alonso R, Lahoz C, Marin C, Garces C, Cenarro A, Perez-Jimenez F, De Oya M, Ordovas JM. Human apolipoprotein A-I gene promoter mutation influences plasma low density lipoprotein cholesterol response to dietary fat saturation. *Athero* 1998;137:367-376.
- Meng QH, Pajukanta P, Valsta L, Aro A, Pietinen P, Tikkanen MJ. Influence of apolipoprotein A-I promoter polymorphism on lipid levels and responses to dietary change in Finnish adults. *J Intern Med* 1997;241:373-378.
- Norum RA, Lakier JB, Goldstein S, Angel A, Goldberg RB, Black WD, Noffze DK, Dolphin PJ, Edelglass J, Borograd DD, Alaupovic P. Familial deficiency of apolipoprotein A-I and C-III and precocious coronary artery disease. *N Engl J Med* 1982;306:1513-1519.
- Ordovas JM, Civeira F, Genest JJ, Craig S, Robbins AH, Meade T, Pocovi M, Frossard P, Masharani U, Wilson PWF, Salem D, Ward RH, Schaefer EJ. Restriction fragment length polymorphisms of the apolipoprotein A-I, C-III, A-IV gene locus: Relationships with lipids, apolipoproteins, and premature coronary artery disease. *Athero* 1991.
- Ordovas JM. The genetics of serum lipid responsiveness to dietary interventions. *Proc Nutr Soc* 1999;58:171-187.
- Paul-Hayase H, Rosseneu M, Robinson D, Van Bervliet JP, Deslypere JP, Humphries SE. Polymorphisms in the apolipoprotein (apo) AI-CIII-AIV gene cluster: Detection of genetic variation determining plasma apo AI, apo CIII and apo AIV concentrations. *Hum Genet* 1992;88:439-446.
- Rader DJ, Hoeg JM, Brewer HB Jr. Quantitation of plasma apolipoproteins in the primary and secondary prevention of coronary artery disease. *Ann Intern Med* 1994;102:1012-1025.
- Smith JD, Brinton EA, Breslow JL. Polymorphism in the human apolipoprotein A-I gene promoter region. Association of the minor allele with decreased production rate in vivo and promoter activity in vitro. *J Clin Invest* 1992;89:1796-1800.



Talmud P, Ye S, Humphries SE. Polymorphism in the promoter region of the apolipoprotein AI gene associated with differences in apolipoprotein AI levels: the European Atherosclerosis Research Study. *Genet Epidemiol* 1994;11:265-280.

### 1.8.5 Bibliografía de LIPC

Berglund L, Oliver EH, Fontanez N, et al. HDL-subpopulation patterns in response to reduction in dietary total and saturated fat intakes in healthy subjects. *Am J Clin Nutr* 1999;70:992-1000.

Couture P, Otvos JD, Cupples LA, et al. Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles. The Framingham Offspring Study. *Arterioscler Thromb Vasc Biol* 2000;20:815-822.

Dreon DM, Fernstrom HA, Campos H, et al. Change in dietary saturated fat intake is correlated with change in mass of large-low-density-lipoprotein particles in men. *Am J Clin Nutr* 1998;67:828-836.

Dugi KA, Amar MJA, Haudenschield CC, et al. In vivo evidence for both lipolytic and nonlipolytic function of hepatic lipase in the metabolism of HDL. *Arterioscler Thromb Vasc Biol* 2000;20:793-800.

Dugi KA, Brandauer K, Schmidt N, et al. Low hepatic activity is a novel risk factor for coronary artery disease. *Circulation* 2001;104:3057-3062.

Guerra R, Wang J, Grundy SM, Cohen JC. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A*. 1997;94:4532-4537.

Guerra R, Wang J, Grundy SM, et al. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *PNAS (USA)* 1997;94:4532-4537.

Katan MB, Grundy SM, Willett WC. Should a low-fat, high-carbohydrate diet be recommended for everyone? Beyond low-fat diets. *N Engl J Med* 1997;337:563-566.

Lambert G, Amar MJA, Martin P, et al. Hepatic lipase deficiency decreases the selective uptake of HDL-cholesterol esters in vivo. *J Lipid Res* 2001;41:667-672.

Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins: A meta-analysis of 27 trials. *Arterioscler Thromb* 1992;12:911-919.

Santamarina-Fojo S, Haudenschield CC, Amar M. The role of hepatic lipase in lipoprotein metabolism and atherosclerosis. *Curr Opin Lipidol* 1998;9:211-219.

Thuren T. Hepatic lipase and HDL metabolism. *Curr Opin Lipidol* 2000;11:277-283.

- Van't Hooft FM, Lundahl B, Ragogna F, et al. Functional characterization of 4 polymorphisms in promoter region of hepatic lipase gene. *Arterioscler Thromb Vasc Bio* 2000; 20:1335-1339.
- Vega GL, Clark LT, Tang A, et al. Hepatic lipase activity is lower in African American than in white American men: effects of 5' flanking polymorphism in the hepatic lipase gene. *J Lipid Res* 1998;39, 228-232.

### 1.8.6 Bibliografía de SR-BI

- Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler Thromb Vasc Biol* 1999; 19:1734-1743.
- Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, Krieger M. Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* 1996; 271:518-520.
- Bektas A, Suprenant ME, Wogan LT, Plengvidhya N, Rich SS, Warram JH, Krolewski AS, Doria A. Evidence of a novel type 2 diabetes locus 50 cM centromeric to NIDDM2 on chromosome 12q. *Diabetes* 1996; 48:2246-2251.
- Cao G, Garcia CK, Wyne KL, Schultz RA, Parker KL, Hobbs HH. Structure and localization of the human gene encoding SR-BI/CLA-1. Evidence for transcriptional control by steroidogenic factor 1. *J Biol Chem* 1997; 272:33068-33076.
- Erkelens DW. Insulin resistance syndrome and type 2 diabetes mellitus. *Am J Cardiol* 2001; 88:38-42.
- Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA* 2002; 287:356-359.
- Franceschini G. Epidemiologic evidence for high-density lipoprotein cholesterol as a risk factor for coronary artery disease. *Am J Cardiol* 2001; 88:9-13.
- Goff DC Jr, D'Agostino RB Jr, Haffner SM, Saad MF, Wagenknecht LE. Lipoprotein concentrations and carotid atherosclerosis by diabetes status: results from the Insulin Resistance Atherosclerosis Study. *Diabetes Care* 2000; 23:1006-1011.
- Howard BV, Cowan LD, Go O, Welty TK, Robbins DC, Lee ET. Adverse effects of diabetes on multiple cardiovascular disease risk factors in women. The Strong Heart Study. *Diabetes Care* 1998; 21:1258-1265.
- Huszar D, Varban ML, Rinninger F, Feeley R, Arai T, Fairchild-Huntress V, Donovan MJ, Tall AR. Increased LDL cholesterol and atherosclerosis in LDL receptor-deficient mice with attenuated expression of scavenger receptor B1. *Arterioscler Thromb Vasc Biol* 2000; 20:1068-1073.

- Kozarsky KF, Donahee MH, Rigotti A, Iqbal SN, Edelman ER, Krieger M. Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels. *Nature* 1997; 387:414-417.
- Lindgren CM, Mahtani MM, Widen E, McCarthy MI, Daly MJ, Kirby A, Reeve MP, Kruglyak L, Parker A, Meyer J, Almgren P, Lehto M, Kanninen T, Tuomi T, Groop LC, Lander ES. Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am J Hum Genet* 2002; 70:509-516.
- Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, Brayer J, Bryant B, Chan G, Daly M, Forsblom C, Kanninen T, Kirby A, Kruglyak L, Munnely K, Parkkonen M, Reeve-Daly MP, Weaver A, Brettin T, Duyk G, Lander ES, Groop LC. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat Genet* 1996; 14:90-94.
- Rigotti A, Trigatti BL, Penman M, Rayburn H, Herz J, Krieger M. A targeted mutation in the murine gene encoding the high density lipoprotein (HDL) receptor scavenger receptor class B type I reveals its key role in HDL metabolism. *Proc Natl Acad Sci U S A* 1997; 94:12610-12615.
- Silver DL, Jiang XC, Arai T, Bruce C, Tall AR. Receptors and lipid transfer proteins in HDL metabolism. *Ann N Y Acad Sci* 2000; 902:103-111.
- Stein O, Stein Y. Atheroprotective mechanisms of HDL. *Atherosclerosis* 1999; 144:285-301.
- Tan KC, Ai VH, Chow WS, Chau MT, Leong L, Lam KS; Influence of low density lipoprotein (LDL) subfraction profile and LDL oxidation on endothelium-dependent and independent vasodilation in patients with type 2 diabetes. *J Clin Endocrinol Metab* 1999; 84:3212-3216.
- Taskinen MR. Diabetic dyslipidemia. *Atherosclerosis*; 2002; S 3:47-51.
- Ueda Y, Gong E, Royer L, Cooper PN, Francone OL, Rubin EM. Relationship between expression levels and atherogenesis in scavenger receptor class B, type I transgenics. *J Biol Chem* 2000; 275:20368-20373.
- Ueda Y, Royer L, Gong E, Zhang J, Cooper PN, Francone O, Rubin EM. Lower plasma levels and accelerated clearance of high density lipoprotein (HDL) and non-HDL cholesterol in scavenger receptor class B type I transgenic mice. *J Biol Chem* 1999; 274:7165-7171.
- Varban ML, Rinninger F, Wang N, Fairchild-Huntress V, Dunmore JH, Fang Q, Gosselin ML, Dixon KL, Deeds JD, Acton SL, Tall AR, Huszar D. Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A* 1998; 95:4619-4624.
- Vaxillaire M, Boccio V, Philippi A, Vigouroux C, Terwilliger J, Passa P, Beckmann JS, Velho G, Lathrop GM, Froguel P. A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nat Genet* 1995; 9:418-423.

- Wang N, Arai T, Ji Y, Rinninger F, Tall AR. Liver-specific overexpression of scavenger receptor BI decreases levels of very low density lipoprotein ApoB, low density lipoprotein ApoB, and high density lipoprotein in transgenic mice. *Biol Chem* 1998; 273:32920-32926.
- Xu S, Laccotripe M, Huang X, Rigotti A, Zannis VI, Krieger M. Apolipoproteins of HDL can directly mediate binding to the scavenger receptor SR-BI, an HDL receptor that mediates selective lipid uptake. *J Lipid Res* 1997;38:1289-1298.

### 1.8.7 Bibliografía de PLIN

- Blanchette-Mackie EJ, Dwyer NK, Barber T, Coxey RA, Takeda T, Rondinone CM, Theodorakis JL, Greenberg AS, Londos C. Perilipin is located on the surface layer of intracellular lipid droplets in adipocytes. *J Lipid Res* 1995; 36:1211-1226.
- Brasaemle DL, Rubin B, Harten IA, Gruia-Gray J, Kimmel AR, Londos C. Perilipin A increases triacylglycerol storage by decreasing the rate of triacylglycerol hydrolysis. *J Biol Chem* 2000; 275:38486-38493.
- Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'connell P, Stern MP. A major susceptibility locus influencing plasma triglyceride concentrations is located on chromosome 15q in mexican americans. *Am J Hum Genet* 2000; 66:1237-1245.
- Frayn KN, Karpe F, Fielding BA, Macdonald IA, Coppack SW. Integrative physiology of human adipose tissue. *Int J Obes Relat Metab Disord* 2003; 27:875-888.
- Garcia A, Sekowski A, Subramanian V, Brasaemle DL. The central domain is required to target and anchor perilipin A to lipid droplets. *J Biol Chem* 2003; 278:625-635.
- Greenberg AS, Egan JJ, Wek SA, Garty NB, Blanchette-Mackie EJ, Londos C. Perilipin, A major hormonally regulated adipocyte-specific phosphoprotein associated with the periphery of lipid storage droplets. *J Biol Chem* 1991; 266:11341-11346.
- Greenberg AS, Egan JJ, Wek SA, Moos MC Jr, Londos C, Kimmel AR. Isolation of cDNAs for perilipins a and b: sequence and expression of lipid droplet-associated proteins of adipocytes. *Proc Natl Acad Sci USA* 1993; 90:12035-12039.
- Lewis GF, Carpentier A, Adeli K, Giacca A. Disordered Fat Storage And Mobilization in the pathogenesis of insulin resistance and type 2 diabetes. *Endocr Rev* 2002; 23:201-229.
- Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, Hanson J, Khoury M, Lau J, O'Brien TR, Rothman N, Stroup D, Taioli E, Thomas D, Vainio H, Wacholder S, Weinberg C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002; 156:300-310.

- Londos C, Brasaemle DL, Gruia-Gray J, Servetnick DA, Schultz CJ, Levin DM, Kimmel AR. Perilipin: Unique proteins associated with intracellular neutral lipid droplets in adipocytes and steroidogenic cells. *Biochem Soc Trans* 1995; 23:611-615.
- Londos C, Gruia-Gray J, Brasaemle DL, Rondinone CM, Takeda T, Dwyer NK, Barber T, Kimmel AR, Blanchette-Mackie EJ. Perilipin: Possible roles in structure and metabolism of intracellular neutral lipids in adipocytes and steroidogenic cells. *Int J Obes Relat Metab Disord* 1996; 20 S3:S97-S101.
- Martinez-Botas J, Anderson JB, Tessier D, Lapillonne A, Chang BH, Quast MJ, Gorenstein D, Chen KH, Chan L. Absence of perilipin results in leanness and reverses obesity in *lepr(db/db)* mice. *Nat Genet* 2000; 26:474-479.
- Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P. Genome-wide search for type 2 diabetes in japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* 2002; 51:1247-1255.
- Nielsen S, Guo Z, Albu JB, Klein S, O'Brien PC, Jensen MD. Energy Expenditure, sex, and endogenous fuel availability in humans. *J Clin Invest* 2003; 111:981-988.
- Nishiu J, Tanaka T, Nakamura Y. Isolation and chromosomal mapping of the human homolog of perilipin (*plin*), A rat adipose tissue-specific gene, by differential display method. *Genomics* 1998; 48:254-257.
- Servetnick DA, Brasaemle DL, Gruia-Gray J, Kimmel AR, Wolff J, Londos C. Perilipins are associated with cholesteryl ester droplets in steroidogenic adrenal cortical and leydig cells. *J Biol Chem* 1995; 270:16970-16973.
- Souza SC, De Vargas LM, Yamamoto MT, Lien P, Franciosa MD, Moss LG, Greenberg AS. Overexpression of perilipin A and B blocks the ability of tumor necrosis factor alpha to increase lipolysis in 3T3-L1 adipocytes. *J Biol Chem* 1998; 273:24665-24669.
- Sztalryd C, Xu G, Dorward H, Tansey JT, Contreras JA, Kimmel AR, Londos C. Perilipin A is essential for the translocation of hormone-sensitive lipase during lipolytic activation. *J Cell Biol* 2002; 161:1093-1103.
- Tansey JT, Huml AM, Vogt R, Davis KE, Jones JM, Fraser KA, Brasaemle DL, Kimmel AR, Londos C. Functional studies on native and mutated forms of perilipins: A role in protein kinase A-mediated lipolysis of triacylglycerols in CHO cells. *J Biol Chem* 2003; 278:8401-8406.
- Tansey JT, Sztalryd C, Gruia-Gray J, Roush DL, Zee JV, Gavrilova O, Reitman ML, Deng CX, Li C, Kimmel AR, Londos C. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proc Natl Acad Sci USA* 2001; 98:6494-6499.

## 1.8.8 Bibliografía sobre soluciones bioinformáticas

### 1.8.8.1 Bibliografía sobre visualización de secuencias

Booch G., Rumbaugh J., Jacobson I. The Unified Modeling Language User Guide. Addison-Wesley, Reading, MA, 1999.

Fischer, C., Schweigert, S., Spreckelsen, C., Vogel, F. "Programs, databases, and expert systems for hu-man geneticists-a survey". Hum. Genet. (1997), 97:129-137.

GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.

GenBank-ftp. [ftp://ftp.ncbi.nih.gov/genomes/H\\_sapiens/](ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/). Accedido el 4 de junio de 2004.

<http://iubio.bio.indiana.edu/soft/molbio/mac/plasmid-maker.html>. Accedido el 4 de junio de 2004.

<http://iubio.bio.indiana.edu/soft/molbio/mac/plasmid-maker.html>. Accedido el 4 de junio de 2004.

<http://www.accessexcellence.org/AE/AEC/CC/restriction.html>. Accedido el 4 de junio de 2004.

<http://www.arches.uga.edu/~lace52/rplasmid.html>. Accedido el 4 de junio de 2004.

Rumbaugh J., Jacobson I., Booch G. El Lenguaje Unificado de Modelado. Manual de Referencia. Addison-Wesley, Madrid, 2000.

### 1.8.8.2 Bibliografía sobre búsqueda de polimorfismos

Christian Blaschke, Alfonso Valencia. The Frame-Based Module of the SUISEKI Information Extraction System. IEEE Intelligent Systems. March 2002: 14-20.

F. Olken. Bioinformatics Databases 1. 18th International Conference on Data Engineering (ICDE'02). February 2002: 649.

Lada A. Adamic, Dennis Wilkinson, Bernardo A. Huberman, Eytan Adar. A Literature Based Method for Identifying Gene-Disease Connections. IEEE Computer Society Bioinformatics Conference (CSB'02). August 2002: 109.

NCBI. The NCBI Handbook (electronic book). (2003). <http://www.ncbi.nlm.nih.gov/>. Accedido el 4 de junio de 2004.

### 1.8.8.3 Bibliografía sobre genes relacionados con las ECV

- Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler Thromb Vasc Biol.* 1999; 19:1734-43.
- Agerholm-Larsen B, Tybjaerg-Hansen A, Schnohr P, Steffensen R, Nordestgaard BG. Common cholesteryl ester transfer protein mutations, decreased HDL cholesterol, and possible decreased risk of ischemic heart disease: The Copenhagen City Heart Study. *Circulation.* 2000; 102:2197-203.
- Andersen RV, Wittrup HH, Tybjaerg-Hansen A, Steffensen R, Schnohr P, Nordestgaard BG. Hepatic lipase mutations, elevated high-density lipoprotein cholesterol, and increased risk of ischemic heart disease: the Copenhagen City Heart Study. *J Am Coll Cardiol.* 2003; 41:1972-82.
- Blankenberg S, Rupprecht HJ, Bickel C, Jiang XC, Poirier O, Lackner KJ, Meyer J, Cambien F, Tiret L; AtheroGene Investigators. Common genetic variation of the cholesteryl ester transfer protein gene strongly predicts future cardiovascular death in patients with coronary artery disease. *J Am Coll Cardiol.* 2003; 41:1983-9.
- Clee SM, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, Jukema JW, van Wijland M, van Dam M, Hudson TJ, Brooks-Wilson A, Genest J Jr, Kastelein JJ, Hayden MR. Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation.* 2001; 103:1198-205.
- Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Folch J, Ordovas JM. Associations of LPL and APOC3 gene polymorphisms on plasma lipids in a Mediterranean population: interaction with tobacco smoking and the APOE locus. *J Lipid Res.* 2002; 43:416-27.
- Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet.* 1998; 20:284-7.
- Deeb SS, Zambon A, Carr MC, Ayyobi AF, Brunzell JD. Hepatic lipase and dyslipidemia: interactions among genetic variants, obesity, gender, and diet. *J Lipid Res.* 2003; 44:1279-86.
- Flavell DM, Jamshidi Y, Hawe E, Pineda Torra I, Taskinen MR, Frick MH, Nieminen MS, Kesaniemi YA, Pasternack A, Staels B, Miller G, Humphries SE, Talmud PJ, Syvanne M. Peroxisome proliferator-activated receptor alpha gene variants influence progression of coronary atherosclerosis and risk of coronary artery disease. *Circulation.* 2002; 105:1440-5.

- Gagne SE, Larson MG, Pimstone SN, Schaefer EJ, Kastelein JJ, Wilson PW, Ordovas JM, Hayden MR. A common truncation variant of lipoprotein lipase (Ser447X) confers protection against coronary heart disease: the Framingham Offspring Study. *Clin Genet.* 1999; 55:450-4.
- Herrington DM, Howard TD, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Brosnihan KB, Meyers DA, Bleecker ER. Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease.[comment]. [Clinical Trial. Journal Article] *New England Journal of Medicine*; 2002; 346:967-74, 2002
- Hines L. M., Stampfer M. J., Ma J., Gaziano J. M., Ridker P. M., Hankinson S. E., Sacks F., Rimm E. B., Hunter D. J. N. Genetic Variation in Alcohol Dehydrogenase and the Beneficial Effect of Moderate Alcohol Consumption on Myocardial Infarction *Engl J Med* 2001; 344:549-555.
- Kastelein JJ, Ordovas JM, Wittekoek ME, Pimstone SN, Wilson WF, Gagne SE, Larson MG, Schaefer EJ, Boer JM, Gerdes C, Hayden MR. Two common mutations (D9N, N291S) in lipoprotein lipase: a cumulative analysis of their influence on plasma lipids and lipoproteins in men and women. *Clin Genet.* 1999; 56:297-305.
- Martin S, Nicaud V, Humphries SE, Talmud PJ; EARS group. Contribution of APOA5 gene variants to plasma triglyceride determination and to the response to both fat and glucose tolerance challenges. *Biochim Biophys Acta.* 2003;1637:217-25.
- Mata P, Ordovas JM, Lopez-Miranda J, Lichtenstein AH, Clevidence B, Judd JT, Schaefer EJ. ApoA-IV phenotype affects diet-induced plasma LDL cholesterol lowering. *Arterioscler Thromb.* 1994; 14:884-91.
- McCarthy JJ, Lehner T, Reeves C, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ; Genequest investigators. Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. *J Med Genet.* 2003; 40:453-8.
- Olivieri O, Stranieri C, Bassi A, Zaia B, Girelli D, Pizzolo F, Trabetti E, Cheng S, Grow MA, Pignatti PF, Corrocher R. ApoC-III gene polymorphisms and risk of coronary artery disease. *J Lipid Res.* 2002;43:1450-7.
- Ordovas JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PW, Schaefer EJ, Tucker K. Polyunsaturated fatty acids modulate the effects of the APOA1 G-A polymorphism on HDL Cholesterol concentrations in a sex-specific manner: the Framingham Study. *Am J Clin Nutr.* 2002; 75:38-46.
- Ordovas JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PW, Schaefer EJ, Tucker KL. Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study. *Circulation.* 2002; 106:2315-21.



- Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. *Arterioscler Thromb Vasc Biol.* 2000; 20:1323-9.
- Ordovas JM, Mooser V. The APOE locus and the pharmacogenetics of lipid response. *Curr Opin Lipidol.* 2002; 13:113-7. Review.
- Osgood D, Corella D, Demissie S, Cupples LA, Wilson PW, Meigs JB, Schaefer EJ, Coltell O, Ordovas JM. Genetic variation at the scavenger receptor class B type I gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: the Framingham study. *J Clin Endocrinol Metab.* 2003; 88:2869-79.
- Tai ES, Demissie S, Cupples LA, Corella D, Wilson PW, Schaefer EJ, Ordovas JM. Association between the PPARA L162V polymorphism and plasma lipid levels: the Framingham Offspring Study. *Arterioscler Thromb Vasc Biol.* 2002; 22:805-10.
- Talmud PJ, Hawe E, Martin S, Olivier M, Miller GJ, Rubin EM, Pennacchio LA, Humphries SE. Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. *Hum Mol Genet.* 2002; 11:3039-46.
- Zambon A, Deeb SS, Pauletto P, Crepaldi G, Brunzell JD. Hepatic lipase: a marker for cardiovascular disease risk and response to therapy. *Curr Opin Lipidol.* 2003;14:179-89.

### **1.8.9 Bibliografía de auditoría en laboratorios de investigación en Biomedicina**

- Benal R., Coltell O. Auditoría de los Sistemas de Información (reimpresión). Servicio de Publicaciones de la Universidad Politécnica de Valencia, Valencia, 1999.
- Cass S., Riezenmann M.J. "Improving Security, Preserving Privacy". *IEEE Spectrum*, Jan.; 2002: 44-49.
- Hass L.M., Kodali P., Rice J.E. "Integrating Life Sciences Data – With a Little Garlic". *Proc. IEEE Int. Symp. on Bio-Informatics and Biomedical Engineering.* IEEE; 2000: 5-12.
- ISACA home page. Information Systems Audit and Control Association. <http://www.isaca.org>. Accedido el 4 de junio de 2004.
- Piattini M., Del Peso E. (eds.) Auditoría Informática. Un enfoque práctico. Ra-Ma, Madrid, 1998.
- Rindfleisch T.C. "Privacy, Information Technology, and Health Care". *Communications of the ACM*, 40-8; 1997: 93-100.
- Rondel R. K., Varley S. A., Webb C. (eds.) *Clinical Data Management.* John Wiley, New York, 1993.

- 
- Sackman H. Biomedical Information Technology. Global Social Responsibilities for the Democratic Age. Academic Press, San Diego, CA (USA), 1997.
- Science's News staff. "A History of the Human Genome Project". Science Magazine; 291 (5507); 2001: 1195-1261.
- Shrotliffe E.H., Blois M.S. "The Computer Meets Medicine and Biology: Emergence of a Discipline". Computer Applications in Health Care and Biomedicine; <http://smi-web.stanford.edu/textbook/ChapterOne.htm> (accedido el 22/01/2002).
- Van Bommel J.H., Musen M.A.(eds.) Hadbook of Medical Informatics. Springer-Verlag, Heidelberg, 1997.



# 2

## HIPÓTESIS Y OBJETIVOS

### 2.1 Hipótesis

La hipótesis conceptual general es la siguiente:

La Bioinformática constituye un soporte disciplinar imprescindible en la investigación genómica cardiovascular mediante su contribución a la gestión, organización, y aseguramiento de varios procesos biomédicos que resultarían inviables de otro modo. Así, gracias a este soporte será posible obtener resultados prácticos de las interacciones gen\*gen y gen\*ambiente, inicialmente formulados a nivel teórico, sobre distintos fenotipos cardiovasculares.

Operativamente, esta hipótesis general puede expresarse a su vez a través de los objetivos que se plasman en la sección que viene a continuación.

### 2.2 Objetivos

El objetivo general de esta tesis es el de establecer un marco de integración de los enfoques metodológicos de la Ingeniería del Software y la Auditoría de Sistemas de Información con la Epidemiología Genómica Cardiovascular y la Bioinformática. En el curso de la investigación conducente a esta tesis, se pretende obtener la evidencia de que las soluciones bioinformáticas propuestas constituyen uno de los soportes fundamentales a la investigación genómica y cardiovascular, dentro del estudio de cohorte denominado *Framingham Heart Study*.

La bioinformática debe aportar su capacidad para facilitar y asegurar la gestión, almacenamiento, salvaguarda, recuperación y presentación de la información biomédica que se maneja en esos estudios. Para ello debe proporcionar los instrumentos necesarios que ayuden a conseguir uno de los objetivos epidemiológicos fundamentales: la prevención de las enfermedades cardiovasculares mediante la identificación, el análisis y el control de los factores de riesgo asociados a dichas enfermedades. Entonces, uno de los objetivos específicos y principales de esta tesis es el de desarrollar los instrumentos

bioinformáticos necesarios para poder calcular el riesgo de padecer enfermedades cardiovasculares teniendo en cuenta la estructura del genoma de los individuos, es decir, su fenotipo, y otros factores que la pueden alterar produciendo distintos efectos observables, su fenotipo.

Los objetivos generales presentados anteriormente pueden desglosarse en los siguientes subobjetivos más específicos:

1. Identificar, en la población del estudio, los factores genéticos y ambientales que determinen un subconjunto de nuevos factores de riesgo y factores de protección frente a las enfermedades cardiovasculares, así como las interacciones significativas entre ellos.
2. Identificar y caracterizar las fuentes y sumideros de información y las necesidades de tratamiento que se puedan plantear en los procesos y protocolos científicos genómicos del ámbito de trabajo. Dichos procesos sirven para la obtención de muestras biológicas de los participantes en los estudios, y también para la adquisición de los datos sobre dichas muestras.
3. Obtener el grado de calidad de la información biomédica y de los tratamientos bioinformáticos existentes en los estudios genómicos cardiovasculares, por si es preciso diseñar y aplicar determinados procedimientos y controles que garanticen un nivel mínimo de calidad es los aspectos mencionados.
4. Comprobar si existe una fuerte necesidad de herramientas bioinformáticas para poder proceder a estudios de viabilidad sobre la adecuación de las aplicaciones existentes, o la exigencia de desarrollar unas nuevas y específicas respecto a lo siguiente: la búsqueda de información clínica y genómica, la búsqueda y detección de polimorfismos, el tratamiento de datos de instrumentos de laboratorio y la integración de datos en bases de datos biomédicas.
5. Obtener un diseño estadístico de los experimentos que permita realizar formalmente el análisis estadístico de los resultados, especialmente con relación a los haplotipos.
6. Obtener los instrumentos de análisis y auditoría, en un enfoque sistémico, cuya aplicación pueda garantizar la seguridad, la privacidad y la eficiencia de los procesos bioinformáticos desarrollados en el estudio.

En estos objetivos se puede ver la concurrencia de aproximaciones disciplinares diferentes cuya conjugación se va a mostrar en esta memoria.

Para demostrar la hipótesis general y alcanzar los objetivos propuestos se utilizan en esta tesis los resultados obtenidos en seis estudios, supeditados al marco general del *Framingham Heart Study*, sobre los siguientes genes: CETP, APOE,

APOA1, LIPC, SR-BI y PLIN. A continuación se describen brevemente los objetivos específicos de cada uno de estos estudios, cuyos resultados, que permitirán validar la hipótesis y los objetivos generales presentados anteriormente, se exponen con amplitud en el Capítulo 4.

Concretamente, esta integración tendrá su aplicación más práctica respondiendo a los objetivos específicos de cada uno de los trabajos presentados y que a continuación se detallan:

1. **Estudio del CETP.** El objetivo ha sido la determinación de la frecuencia, la expresión fenotípica y la modulación potencial del riesgo de ECV en la población general por parte del RFLP de TaqIB en el primer intrón del gen CETP (Ordovás et al., 2000). En particular, se ha estudiado la variabilidad interindividual en los niveles de lípidos, en los perfiles de subclases de lipoproteínas y en el riesgo cardiovascular asociado con este polimorfismo de CETP entre los participantes del *Framingham Offspring Study*.
2. **Estudio del APOE.** El objetivo ha sido el examen sobre si la variación del locus del gen APOE modula la asociación entre la toma de alcohol y las concentraciones de LDL-C entre los participantes del *Framingham Offspring Study* (Corella et al., 2001).
3. **Estudio del APOA1.** El objetivo ha sido la investigación sobre la posible interacción de la ingesta de grasa en dieta, con los efectos del polimorfismo – 75 bp G/A en el promotor del gen APOA1, sobre los niveles de HDL-C y apoA-I entre los participantes del *Framingham Offspring Study* (Ordovás et al., 2001).
4. **Estudio del LIPC.** El objetivo ha sido la investigación de cómo la ingesta de grasa en dieta, centrándose en los ácidos grasos específicos y los orígenes de la grasa, interactúa con el polimorfismo de LIPC para determinar las concentraciones y el tamaño de partícula de HDL-C entre los participantes del *Framingham Offspring Study* (Ordovás et al., 2002).
5. **Estudio del SR-BI.** El objetivo ha sido la composición de los siguientes subobjetivos (Osgood et al., 2003):
  - 5.1. Evaluar el efecto fenotípico en concentraciones de lípidos, lipoproteínas y apolipoproteínas, así como en los perfiles de las subclases de lipoproteínas, de las variantes genéticas comunes del gen SR-BI en la cohorte de los participantes del *Framingham Offspring Study*.
  - 5.2. Examinar la asociación de la variación genética en el locus del gen SR-BI con el curso de la diabetes de tipo II en la cohorte de los participantes del *Framingham Offspring Study*.

- 5.3. Estimar si la diabetes modifica la asociación entre la variación del gen SR-BI y los rasgos fenotípicos examinados.
6. *Estudio del PLIN*. El objetivo ha sido examinar la asociación entre los polimorfismos comunes del gen PLIN y los fenotipos de obesidad en una amplia muestra de sujetos caucasianos extraídos de la población general así como de pacientes con obesidad severa (Qi et al., 2004). Este enfoque basado en la utilización de la combinación de población general y pacientes de hospital permite cubrir un amplio rango de fenotipos de obesidad y prevenir las limitaciones de diseños experimentales por separado.

## 2.3 Referencias

- Corella D, Tucker K, Lahoz C, Coltell O, Cupples LA, Wilson PWF, Schaefer EJ, Ordovas JM. "Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study". *American Journal of Clinical Nutrition*; 2001; 73(4):736-45. ISSN "0002-9165 (SCI: 2000, 5,012 p., 3/52 "NUTRITION AND DIETETICS")
- Doreen Osgood; Dolores Corella; Serkalem Demissie; L Adrienne Cupples; Peter WF Wilson; James B Meigs; Ernst J Schaefer; Oscar Coltell; Jose M Ordovas. "Genetic variation at the scavenger receptor class B type I (SR-BI) gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: The Framingham Study". *J Clin Endocrinol Metab*. 2003 Jun;88(6):2869-79. ISSN 0021-972X (SCI 2002, 5.199 p.; 13/88 "ENDOCRINOLOGY & METABOLISM").
- Lu Qi; Dolores Corella, José V. Sorlí, Olga Portolés, Haiqing Shen, Oscar Coltell, Diego Godoy, Andrew S. Greenberg, Jose M. Ordovas. "Genetic variation at the Perilipin (PLIN) locus is associated with obesity-related phenotypes in White women". *Clinical Genetics* 2004 Ref. CGE-00136-2004. ISSN 0009-9163 Online ISSN: 1339-0004 (SCI 2002, 2.237 p.; 57/115 "GENETICS & HEREDITY").
- Ordovás JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Polyunsaturated fatty acids modulate the effect of the APOA1-75(G/A) polymorphism on HDL-C levels in a specific fashion: The Framingham Study". *American Journal of Clinical Nutrition*, 2002; 75(1):38-46. ISSN "0002-9165 (SCI: 2000, 5,012 p., 3/52 "NUTRITION AND DIETETICS")

Ordovás JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on HDL metabolism: Evidence of a strong dose-effect in this gene-nutrient interaction in the Framingham Study". *Circulation*, 2002; 18: 2315-2321. ISSN 0009-7322 (SCI: (2000, 10,893 p.), (2001, 10,517 p.); 1/63 "CARDIAC & CARDIOVASCULAR SYSTEMS"; 1/60 "HEMATOLOGY"; 1/45 "PERIPHERAL VASCULAR DISEASE").

Ordovas JM, Cupples A, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PWF, Schaefer EJ. "Association of CETP TaqIB polymorphism with variation in lipoprotein subclasses and coronary heart disease risk: The Framingham Study". *Arteriosclerosis Thrombosis and Vascular Biology*; 2000; 20: 1323-1329. ISSN 1079-5642 (SCI: 1999, 5,406 p., 6/45 "PERIPHERAL VASCULAR DISEASE", 5/60 "HEMATOLOGY").



La ignorancia afirma o niega rotundamente; la ciencia duda.

François Marie Antoine Voltaire

### **3.1 Introducción**

Los estudios presentados en esta tesis son fruto de un trabajo multidisciplinar que implica una labor en equipo para la obtención de las conclusiones generales. A lo largo de los siete años de investigación recopilando la información necesaria para llevar a cabo los trabajos conducentes a esta tesis, el doctorando ha tenido la oportunidad de participar en todas y cada una de las tareas que forman parte del protocolo de la investigación genómica cardiovascular.

Así, por decisión propia, el doctorando se ha familiarizado con las técnicas de laboratorio incluyendo, desde la extracción de ADN, hasta el análisis de polimorfismos por distintas técnicas. Por otra parte, el doctorando ha tenido que emplear sus conocimientos en Ciencias de la Computación y Estadística, para llevar a cabo distintas actividades científico-tecnológicas: el diseño, desarrollo y gestión de proyectos bioinformáticos, desarrollo de software, gestión de datos, control de calidad de procesos y datos, diseño de guías de auditoría y aplicación de las mismas, elaboración de recomendaciones y herramientas para la mejora de procesos en el laboratorio genómico, acceso a bases de datos genómicas, tratamiento y análisis estadístico de datos, y participación en la presentación de los resultados y redacción de los manuscritos o comunicaciones a congresos.

Por todo ello, en esta sección se describen los principios, las metodologías, las técnicas y los recursos, aplicados en los trabajos de investigación de esta tesis, que se relacionan con las distintas y multidisciplinarias actividades mencionadas anteriormente. En primer lugar se procede a la caracterización de los sujetos participantes en el estudio. Después se describen los principios y técnicas aplicados para realizar las determinaciones bioquímicas, moleculares y dietéticas. A continuación se describen los procesos empleados en la extracción y amplificación

de ADN, así como del análisis de polimorfismos. Seguramente se entra en la exposición de los métodos estadísticos esenciales para el análisis de los datos obtenidos en los estudios, y se describen los análisis particulares realizados para cada uno de los genes en los respectivos estudios. Finalmente, se aborda la exposición de los recursos metodológicos y enfoques de ingeniería que se aplican en el desarrollo de los proyectos bioinformáticos.

## **3.2 Recursos y metodología de Epidemiología Genómica Cardiovascular**

### **3.2.1 Sujetos de estudio**

Las personas estudiadas son los participantes en el *Framingham Offspring Study* (Kannel et al., 1979). Dicho estudio comenzó en 1971 con el reclutamiento de 5.124 participantes (Dawber et al., 1951). Estos individuos estaban relacionados con los primeros integrantes del clásico Framingham Heart Study.

Aunque se parte de una muestra teórica de 5.124 participantes, el número de individuos incluido en cada análisis genético es variable e inferior a esta cifra debido a que se utilizan los datos válidos completos en las variables analizadas.

Se recolectaron las muestras de sangre para la extracción del ADN entre los años 1987 y 1991 de los individuos que participaron en la visita del quinto ciclo del *Framingham Offspring Study*, realizado entre 1992 y 1995. Las características de inclusión de estos individuos eran las de no haber padecido ningún episodio de enfermedad cardiovascular (infarto de miocardio, angina de pecho e insuficiencia coronaria), no estar tomando medicación para la disminución del colesterol, no estar recibiendo terapia de sustitución por estrógenos, o tener niveles de triglicéridos plasmáticos por debajo de 4,52 mmol/L. Casi todos los participantes del estudio son de raza caucásica. Y los datos de consumo de tabaco, peso y talla se obtuvieron por los procedimientos descritos en (Dawber et al., 1951) (Cupples et al., 1992). Por supuesto, todos los procedimientos y actuaciones fueron previamente aprobadas por los comités de ética correspondientes.

### **3.2.2 Determinaciones**

De todos los participantes se recoge información sobre variables socio-demográficas (sexo, raza, edad, residencia, estructura familiar, ocupación, etc.),

antropométricas (peso, talla, medida de la cintura, medida de la cadera), del estilo de vida (consumo de tabaco, consumo de alcohol, ejercicio, consumo de alimentos), clínicas (consumo de fármacos, enfermedades presentes y antecedentes de enfermedad, tensión arterial, pulso, así como determinaciones bioquímicas en muestras de sangre y orina) y variables genéticas (análisis de mutaciones y polimorfismos en genes candidatos).

### 3.2.2.1 Determinaciones de glucemia, lípidos, lipoproteínas y apolipoproteínas plasmáticas y actividades enzimáticas

Con el propósito de determinar las concentraciones de glucemia, lípidos, lipoproteínas y apolipoproteínas plasmáticas, el procedimiento empleado ha sido el de recolectar muestras de sangre venosa habiendo estado doce horas en ayunas en tubos con un 0,1% de anticoagulante EDTA.

Se ha separado el plasma de las células sanguíneas por centrifugación e inmediatamente se ha utilizado para la medida de los lípidos. Según se describe en (Manninen et al., 1988), se han determinado los niveles de colesterol plasmático total (TC), HDL-C y LDL-C. El HDL-C se ha medido después de la precipitación de apoB conteniendo lipoproteínas con sulfato de dextrano y sulfato de magnesio (Manninen et al., 1988). Las concentraciones de LDL-C se han estimado con la ecuación de Friedewald et al (Kannel et al., 1979):

$$\text{LDL-C} = \text{Colesterol total} - (\text{HDL} + \text{TG} / 5) \quad (\text{E: 3.1})$$

Los niveles plasmáticos de apolipoproteína AI (apoA-I) y apolipoproteína B (apoB) se han medido por ensayo inmunoabsorbente ligado a enzima no competitiva (en inglés, *enzyme-linked immunosorbent assay*, que se abrevia como ELISA), mediante anticuerpos policlonales purificados por afinidad (Tenkanen et al., 1991) (Cupples et al., 1992). Los coeficientes de variación para las mediciones de colesterol total, HDL-C y TG, han sido menores del 5% (Warnick et al., 1982). Se ha medido la glucosa en ayunas en especímenes frescos con un kit de reagente de hexoquinasa (Manninen et al., 1988). Y los ensayos de glucosa se han realizado en paralelo habiendo obtenido un coeficiente de variación interensayo menor del 3%.

La actividad de CETP se ha determinado mediante la aplicación de un Kit Diagnóstico de CETP de *Roar Biomedical, Inc* (New York, NY, USA). Este kit incluye partículas donantes (partículas de fosfolípidos sintéticos y ésteres de colesterol) y receptoras (VLDL). El lípido neutral fluorescente está presente en un estado auto-inactivado cuando se contiene dentro del núcleo del donante.

La transferencia de CETP mediadora se ha determinado mediante el incremento en intensidad de fluorescencia cuando el lípido neutral fluorescente se elimina del donante auto-inactivado y se deposita en el receptor.

Con respecto a las cantidades, se han diluido 10  $\mu$ l de plasma al (1:10) en 90  $\mu$ l de buffer de muestra (10mM tris, 150 mM NaCl, 2nM EDTA) a un pH de 7,4 en una microplaca compatible fluorescente, de Dynex Laboratories, e incubados durante tres horas a 37  $^{\circ}$ C. El ensayo se ha leído en un espectrómetro de fluorescencia con una excitación de longitud de onda de 465 nm y una emisión de longitud de onda de 535 nm. Se ha empleado una curva estándar, según las indicaciones de los fabricantes, para derivar las relaciones entre la intensidad de fluorescencia y la transferencia de masa. Las muestras de plasma de los sujetos del grupo de control se incluyeron en cada una de las placas anteriores para evitar los sesgos entre placas.

### 3.2.2.2 Determinaciones por resonancia magnética nuclear

Las concentraciones y subclases de lipoproteínas plasmáticas se han obtenido mediante espectroscopia por resonancia magnética nuclear protónica (NMR) como se describe en (Schaefer et al., 1986) (Ordovas et al., 1987). Cada perfil muestra las concentraciones de seis subclases de VLDL, una de IDL, tres de LDL y cinco de HDL, y el peso medio ponderado de tamaño de partícula de VLDL, LDL y HDL (Figura 3.1).

Las cinco subclases de HDL se han agrupado en las tres categorías siguientes (Schaefer et al., 1986):

1. HDL grande (8,8 a 13,0 nm).
2. HDL intermedio (7,8 a 8,7 nm).
3. HDL pequeño (7,3 a 7,7 nm).

Las tres subclases de LDL se han agrupado en las tres categorías siguientes:

1. LDL grande (21,3 a 27,0 nm).
2. LDL intermedio (19,8 a 21,2 nm).
3. LDL pequeño (18,3 a 19,7 nm).

Las concentraciones de las subclases de HDL y LDL se han expresado en unidades de colesterol(mmol/L).

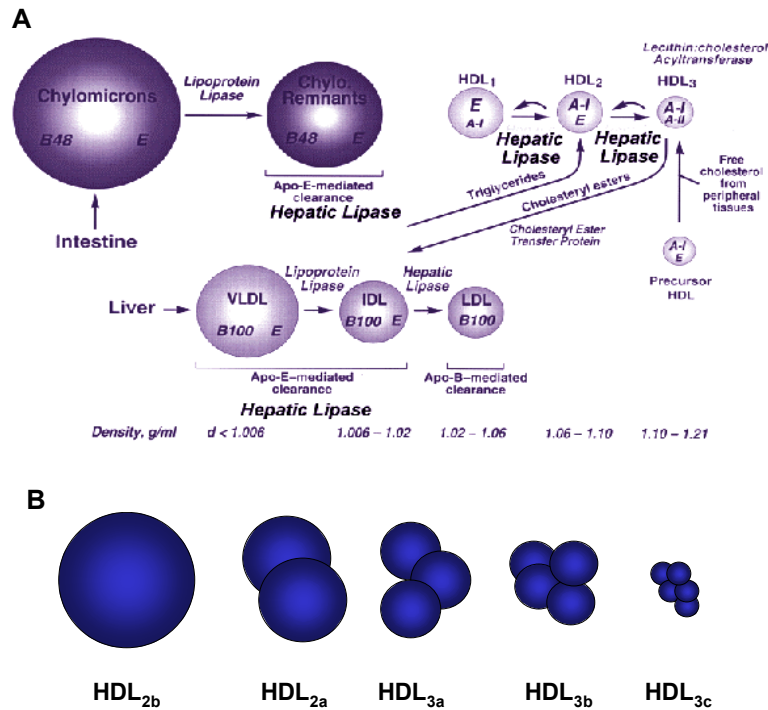


Figura 3.1. Tamaños de partículas de VLDL, HDL y LDL. (A) Ejemplo en la ruta metabólica de la Lipasa Hepática (Fuente: Ordovás et al., 2002, *Circulation*). (B) Distintos tamaños y morfologías de las partículas HDL

### 3.2.2.3 Estimación del consumo de alimentos y nutrientes

El consumo de alimentos de los participantes del *Framingham Offspring Study* se ha estimado mediante la aplicación del cuestionario de frecuencia de alimentos de Willett (McNamara et al., 1987). Este cuestionario incluye 136 ítems de alimentos, con algunas cuestiones sobre cerveza, vino y bebidas espirituales. A los sujetos se les ha pedido que marquen la frecuencia diaria, semanal o mensual en que consumen cada uno de los alimentos presentados. Posteriormente, en la Harvard University, las frecuencias de ingesta se han asociado con los datos de nutrientes para obtener las estimaciones de las ingestas diarias.

La información sobre la ingesta de grasa se ha obtenido en valores absolutos (g/día) y se ha modelado el efecto de la grasa como densidad de nutriente, es decir, la proporción de la energía proveniente de la grasa ingerida con respecto del aporte total de energía en forma de porcentaje de esta última. Se han calculado para cada

individuo la grasa total, los ácidos grasos saturados (SAFA), ácidos grasos monoinsaturados (MUFA) y ácidos grasos poliinsaturados (PUFA). Estas variables se han analizado como magnitudes continuas y categóricas. Como variables categóricas, se han clasificado los sujetos en dos grupos según el valor medio de la población. Además, se han definido tres categorías de ingesta de PUFA: baja (< 4% energía/día); intermedia (4 - 8% energía/día); y alta ( $\geq$ 8% energía/día). Estas categorías se han basado en la distribución de frecuencia de las variables de PUFA (correspondientes a los deciles más extremos), considerando el rango del consumo de PUFA en la población. En los hombres, el rango de ingesta de PUFA era de 2,03 a 16,54%. En las mujeres, el rango era de 1,19 a 13,66%.

Se ha calculado el consumo de alcohol en g/día de acuerdo con lo que cada individuo ha informado sobre las bebidas alcohólicas consumidas en el año anterior. Los sujetos se han dividido en dos categorías: los no bebedores, quienes no han informado de ningún consumo de bebidas alcohólicas; y los bebedores, quienes sí que informan del consumo de bebidas alcohólicas.

### 3.2.3 Análisis genómico

El análisis genómico se realiza estudiando la variación en la secuencia de ADN en lugares específicos de genes candidatos elegidos a priori por su posible relación con el metabolismo lipídico y el riesgo cardiovascular. A partir de una muestra de ADN genómico de cada individuo se determinan estas variaciones empleando bien métodos clásicos basados en la técnica de polimorfismos de longitud en los fragmentos de restricción (RFLP: *Restriction Fragment Length Polymorphisms*), o bien en las técnicas basadas en fluorescencia. En esta tesis los genes y polimorfismos seleccionados para su análisis han sido los siguientes:

1. CETP: polimorfismo TaqIB.
2. APOE: alelos E2, E3 y E4.
3. APOA1: polimorfismo -75G/A.
4. LIPC (LIPASA HEPÁTICA): polimorfismo -514C/T.
5. SRBI: polimorfismo en exón 1, intrón 5 y exón 8.

#### 3.2.3.1 Extracción de ADN

El DNA leucocítico se ha extraído de muestras de 5 a 10 ml de sangre total con el método descrito por Miller et al. (Ordovas et al., 1987). El método consiste en la obtención de ADN a partir de leucocitos de sangre periférica. Brevemente, en un

primer paso se lisan los hematíes con una solución hipotónica; a continuación se utiliza una solución detergente para romper las membranas de los leucocitos y liberar el ADN. Por último, se purifica el ADN con una mezcla fenol-cloroformo y se precipita con acetato amónico 10M y etanol 70%. El ADN obtenido se resuspende en tampón TE (Tris EDTA, pH 7,5) y se guarda a -4°C (Figura 3.2).

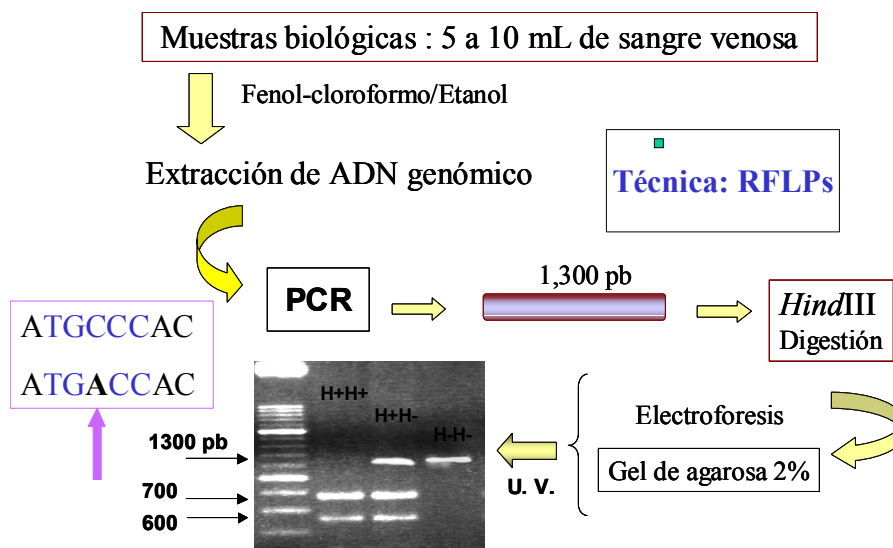


Figura 3.2. Obtención del ADN de los individuos y análisis de variaciones en epidemiología genómica

### 3.2.3.2 Amplificación por la reacción en cadena de la polimerasa (PCR)

La amplificación mediante PCR de los exones de los genes CETP, APOE, APOA1, LIPC, SR-BI y PLIN se realizó utilizando los oligonucleótidos correspondientes. Todas las amplificaciones se efectuaron en un volumen de 50 µl (para 300 y 600 ng de ADN de la muestra) y la mezcla de reacción que se utilizó fue: 1 µg de dNTP's, 25 µg de Mg<sub>2</sub>Cl<sub>2</sub>, 2 µg de cada primer y 2,5 U de Taq Polimerasa (5U/ml). Las condiciones de los ciclos de amplificación para el estudio de las diferentes variantes genéticas fueron básicamente las mismas; después de una desnaturalización inicial a 94°C durante 5 minutos, se realizaron los ciclos indicados para cada variante en las referencias correspondientes (que se detallan más adelante) (Figura 3.2). En general, se utilizó el siguiente protocolo: 30 segundos a 94°C, 30 segundos a la correspondiente temperatura de anillamiento y 40 segundos a 72°C. Para la extensión del amplificado se utilizó una temperatura de 72°C durante 5 minutos. La verificación de las amplificaciones se realizaba

cargando 5 ml de la reacción en geles horizontales de agarosa 2% en 1xTEB. El ADN se visualizaba mediante tinción con bromuro de etidio y exposición a luz UV, el tamaño del fragmento obtenido se obtenía por comparación con distintos marcadores de tamaño en función de los pares de bases del fragmento amplificado.

### 3.2.3.2 Análisis de polimorfismos

Para el análisis de los polimorfismos, en primer lugar se procedió a la amplificación de un fragmento de ADN conteniendo dicha variación. El método empleado fue la reacción en cadena de la polimerasa (PCR) con las especificaciones de tiempo, ciclos, temperatura, cebadores, reactivos y tampones que se detalla a continuación para cada polimorfismo. Tras la PCR, los amplificados de los genes CETP, APOE, APOA1 y LIPC, fueron sometidos a digestión con enzimas de restricción al emplear para estos análisis la técnica de los RFLP. Para el análisis de polimorfismos en SR-BI y PLIN, no se empleó la técnica de los RFLP sino el análisis de fluorescencia como posteriormente se detalla.

#### CETP: polimorfismo TaqIB

Se ha amplificado un fragmento de 535 pares de bases en el intrón 1 del gen CETP con la técnica de PCR en un termociclador de ADN (modelo PTC-100, de *MJ Research*, Inc, Watertown, MA, USA), utilizando primers de oligonucleótidos:

Forward: 5'-CACTAGCCCAGAGAGAGGAGTGCC-3'.

Reverse: 5'-CTGAGCCCAGCCGCACACTAAC-3'.

Cada amplificación se ha llevado a cabo con 100 ng de ADN genómico en un volumen de 50 µl conteniendo 40 pmol de cada nucleótido, 0,2 mM de dNTP, 1,5 mM de MgCl<sub>2</sub>, 10 mM de Tris a pH 8,4 y 0,25 U (unidades) de polimerasa Taq. Las plantillas de ADN se han desnaturalizado a 95°C durante 3 minutos y cada reacción de PCR se ha sometido a 30 ciclos con la temperatura de ciclo correspondiendo a la serie siguiente: 95°C durante 30 seg, 60°C durante 30 seg, y 72°C durante 45 seg, y finalmente una extensión a 72°C durante 5 min.

#### APOE: alelos E2, E3 y E4

Los genotipos de APOE se han tratado como describen Hixson and Vernier (Hixson et al., 1990). Se ha amplificado un fragmento de 244 pares de bases gen APOE incluyendo los dos sitios polimórficos con la técnica de PCR en un termociclador de ADN (modelo PTC-100, de *MJ Research*, Inc, Watertown, MA,



USA), utilizando los primers de oligonucleótidos F4 y F6 (Hixson et al., 1990). Cada una de las mezclas de reacción se ha calentado a 90°C durante 2 min y posteriormente se han aplicado 35 ciclos de amplificación en la serie: 94°C durante 40 seg, 62°C durante 30 seg y 72°C durante 1 min. Se han digerido los productos de la PCR con 5 unidades de HhaI y se han separado los fragmentos por electroforesis en un gel no desnaturizante de poliacrilamida. Después de la electroforesis, el gel se ha tratado con bromuro de etidio durante 30 minutos y se han examinado los fragmentos de ADN bajo iluminación UV.

#### **APOA1: polimorfismo -75G/A**

Se ha amplificado un fragmento de 432 pares de bases gen APOA1 región 5' con la técnica de PCR en un termociclador de ADN (modelo PTC-100, de *MJ Research*, Inc, Watertown, MA, USA), utilizando 250 ng de ADN genómico y 0,2 µM de cada primer de los oligonucleótidos siguientes en un volumen de 50 µL:

P1: 5'-AGGGACAGAGCTGATCCTTGA ACTCTTAAG-3'.

P2: 5'-TTAGGGGACACCTACCCGTCAGGAAGAGCA-3'.

Cada una de las mezclas de reacción se ha calentado a 95°C durante 5 min y posteriormente se han aplicado 30 ciclos de amplificación en la serie: 95°C durante 1 min, 58°C durante 1,5 min y 72°C durante 2 min. Se han digerido los productos de la PCR con 10 unidades de MspI (BRL y MD) y se han separado los fragmentos por electroforesis en un gel de agarosa al 3,5%. Después de la electroforesis, el gel se ha tratado con bromuro de etidio durante 20 minutos y se han examinado los fragmentos de ADN bajo iluminación UV.

#### **LIPC (LIPASA HEPÁTICA): polimorfismo -514C/T**

El ADN genómico se ha aislado a partir de leucocitos de sangre periférica mediante métodos estándar (Friedewald et al., 1972). Se ha realizado el genotipado de la lipasa hepática como se describe en Guerra et al. (1997). Se ha amplificado una secuencia del gen de la lipasa hepática de 285 pb mediante PCR en un termociclador de ADN (PTC-100, M.J. Research, Inc) utilizando los primers de los oligonucleótidos siguientes:

P1: 5'-TCTAGGATCACCTCTCAATGGGTCA-3'.

P2: 5'-GGTGGCTTCCACGTGG-CTGCCTAAG-3'.

Se han desnaturalizado las hebras de ADN a 95°C durante 3 minutos, y a continuación, cada PCR se ha sometido a 35 ciclos, cada uno de ellos compuesto de 1 minuto de desnaturalización a 95°C, 0,5 minutos de anillamiento a 63°C, y 0,5 minutos de extensión a 72°C. Los productos de PCR se han digerido con 10 U de NlaIII y los fragmentos se han separado por electroforesis en un gel de agarosa al 15%. Después de la electroforesis, el gel se ha tratado con bromuro de etidio durante 20 minutos y los fragmentos de ADN se han visualizado bajo iluminación ultravioleta. Los fragmentos resultantes son uno de 215 pb y otro de 70 pb para el alelo T, y uno de 285 pb para el alelo C no cortado.

### SR-BI: polimorfismo en exón 1, intrón 5 y exón 8

El genotipado del gen SR-BI se ha realizado mediante discriminación alélica utilizando el ensayo de la nucleasa 5' con sondas fluorogénicas en un instrumento *7700 Systems Detection System* (PE Applied Biosystems) como se describe en (Osgood et al., 2000). En resumen, se han usado dos sondas *TaqMan* (PE Applied Biosystems), una por cada alelo.

El genotipado mediante sondas *TaqMan* está basado en diseñar dos sondas alelo-específicas (A y B) cuyo extremo 3' coincide con la posición polimórfica. Dichas sondas poseen un fluoróforo *reporter* en su extremo 5' (p.e. Fam para sonda A y Tet para sonda B) y otro fluoróforo *quencher* en el extremo 3' (p.e. Tamra). El fluoróforo Tamra bloquea la emisión de fluorescencia por parte del Tet y Fam cuando están próximos físicamente. En el caso de que la sonda hibride perfectamente con la secuencia diana (p.e. sonda A en un individuo AA), la Taq DNA polimerasa degrada la sonda hibridada (debido a que la Taq DNA polimerasa posee actividad exonucleasa 5'-3') y se produce emisión de fluorescencia Fam debido a que dicho fluoróforo queda físicamente separado de Tamra. En un individuo heterocigoto, en cambio, se observaría fluorescencia para Fam y Tet. En el caso de que no exista hibridación (p.e. sonda A en individuo BB) no hay emisión de fluorescencia Fam, y sólo se observa fluorescencia Tet. En definitiva, dependiendo del tipo de fluorescencia observada puede identificarse el genotipo del individuo analizado (Fam = AA, Tet = BB, Fam+Tet = AB). Específicamente, para el caso de los polimorfismos del SR-BI, cada sonda consiste de un oligonucleótido con una 5' reporter dye y una 3' quencher dye. Las reporter dye usadas han sido fluorescina-6-carboxi (FAM) y VIC®; y en la quencher dye, tetrametilrodamina-6-carboxi (TAMRA). En (Osgood et al., 2000) se muestra una tabla con las secuencias de los primers y sondas, así como el programa del termociclador, aplicado para las variantes del exón 1, intrón 5 y exón 8. Estas secuencias están reproducidas en la Tabla 3.1. Se ha realizado la PCR en un volumen final de  $\mu\text{L}$  para cada SNP individual. La mezcla de reacción contenía 5  $\mu\text{L}$  de *TaqMan 2x Universal PCR Master Mix* (con la siguiente composición: polimerasa AmpliTaq

Gold, Amperasa uracilo-N-glicosilasa, dUTP, dGTP, dCTP, dATP, tintura de 6-carboxi-x-rodamina, Tris-HCl, KCl y MgCl<sub>2</sub>), 200 nmol/L de probando etiquetado FAM, 50 nmol/L de sonda etiquetada VIC, 900 nmol/L de primer reverso, 900 nmol/L de primer directo, y de 2 a 20 ng de ADN genómico.

**Tabla 3.1. Secuencias de los primers y sondas del gen SR-BI**  
(Fuente: Osgood et al., 2000)

| SNP                      | Primers y sondas   |
|--------------------------|--|
| <i>Exón 1 (G&gt;A)</i>   | Directo: 5'-GTCCCCGTCTCCTGCCA-3'<br>Inverso: 5'-CCCAGCACAGCGCACAGTA-3'<br>Sonda alelo G: 5'-FAM-AGACATGGGCTGCTCCGCCA-TAMRA-3' <sup>a</sup><br>Sonda alelo A: 5'-VIC-CAGACATGAGCTGCTCCGCCA-TAMRA-3' |
| <i>Intrón 5 (C&gt;T)</i> | Directo: 5'-CAAGTGGAACGGGCTGAGCAAGGT-3'<br>Inverso: 5'-TCTGGTCCCTGCCACTCCCGA-3'<br>Sonda alelo C: 5'-FAM-AGCCATGGCCGGGCCACCC-TAMRA-3'<br>Sonda alelo T: 5'-VIC-AGCCATGGCCAGGCCACCC-TAMRA-3'        |
| <i>Exón 8 (C&gt;T)</i>   | Directo: 5'-CCCCCTTGTCTCTCTCCCAT-3'<br>Inverso: 5'-AGGCCAGTCACCGCTTCTG-3'<br>Sonda alelo C: 5'-VIC-CCTCAACGCCGACCCGGTT-TAMRA-3'<br>Sonda alelo T: 5'-FAM-TTCCTCAACGCTGACCCGGTTC-TAMRA-3'           |

a : Las bases en negrita representan mutaciones puntuales.

### Perilipin (PLIN)

Los seis SNP del gen PLIN examinados se presentan en la Figura 3.3 y la Tabla 3.2. Se han denominado los polimorfismos según las recomendaciones más recientes (Antonarakis, 1998). La secuencia de referencia ha sido la identificada como GI21431190 en GenBank. El genotipado se ha realizado mediante la técnica de *primer-extension análisis*. Esta técnica consiste en diseñar un oligonucleótido cuyo extremo 3' hibride con el nucleótido 5' adyacente a la mutación que se desea detectar. Se realiza una reacción de extensión con 4 dideoxinucleótidos marcados con distintos fluoróforos y el resultado se analiza en un aparato de electroforesis capilar. Dependiendo del dideoxinucleótido que se haya incorporado, el pico de fluorescencia observado tendrá un color u otro y de este modo se puede discriminar el genotipo del individuo. El genotipado mediante *primer-extension analysis* se lleva a cabo mediante el kit *SNaPshot ddNTP Primer Extension kit* (Applied Biosystems).

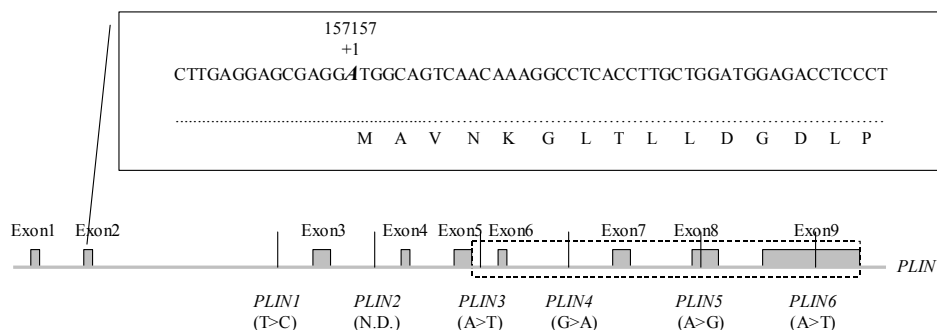


Figura 3.3. Esquema del gen *PLIN* y localización de polimorfismos

En primer lugar, se han amplificado los fragmentos de ADN mediante PCR múltiple. Los productos de la PCR han sido los fragmentos de 422 pb para el PLIN1, 391 pb para el PLIN2, 318 pb para el PLIN3, 350 pb para el PLIN4, 190 pb para el PLIN5, y 469 pb para el PLIN6. La amplificación con PCR se ha realizado en un volumen de reacción de 10  $\mu$ L conteniendo 0,2 mmol/L de cada uno de los dNTP, 0,2  $\mu$ mol/L de cada uno de los primers, 3,0 mmol/L de  $MgCl_2$  (clorhidrato de magnesio) y 0,8 U de polimerasa Qiagen Hotstar Taq. El proceso de la PCR ha seguido la serie: 95°C durante 10 min seguidos de 7 ciclos de 95°C durante 30 seg, 70°C durante 30 segs, y 72°C durante 1 min; a continuación, se han aplicado 41 ciclos de 95°C durante 30 seg, 65°C durante 30 seg, y 72°C durante 1 min.

Se ha incluido como cierre del protocolo una fase de extensión de 2 min a 72°C. Se han incubado los productos de la PCR durante 60 min a 37°C con 2,5 U de *Exonucleasa I* (New England Biolabs, Inc. Beverly, MA, USA) y *Fosfatasa Intestinal Calf* (New England Biolabs, Inc. Beverly, MA, USA) para remover los primers y dNTP no incorporados. A continuación se ha aplicado una incubación durante 15 min a 75°C para inactivar las enzimas. En un paso posterior, se ha llevado a cabo la *primer-extension análisis* mediante el sistema *ABI Prism SnaPshot multiplex system* (Applied Biosystems, Foster City, CA, USA). La reacción de extensión se ha desarrollado con un termociclador de PCR en una mezcla de reacción de 5  $\mu$ L que contenía 1,5  $\mu$ L del *Snapshot Ready Reaction Mastermix* (Applied Biosystems, Foster City, CA, USA), 1,0  $\mu$ L de agua, y 1,5  $\mu$ L de productos múltiples de PCR y 1,0  $\mu$ L de la mezcla de sondas (1,5  $\mu$ mol/L de PLIN1, PLIN2, PLIN3 y PLIN4, y 2,0  $\mu$ mol/L de PLIN5 y PLIN6).



Las condiciones de la reacción han sido las siguientes: 35 ciclos de 96°C durante 30 seg, 50°C durante 30 seg y 60°C durante 30 seg. Se han incubado los productos de reacción durante 60 min a 37°C con 3 U de *Fosfatasa Intestinal Calf* para remover los dNTP no incorporados, seguido de una incubación durante 15 min a 75°C para inactivar la enzima. El genotipado se ha realizado con los productos finales en un *ABI Prism 3100 genetic analyzer* (Applied Biosystems, Foster City, CA, USA) usando el software *Genotyper* versión 3.7 (Applied Biosystems, Foster City, CA, USA). Para conseguir la precisión de los datos de genotipado se han seguido prácticas de laboratorio normalizadas y de calidad. De este modo, se han aplicado controles internos y repetición de experimentos: para cualquier muestra que ha dado una señal débil se ha repetido su tratamiento. Además, el 20% de la muestras se han repetido aleatoriamente con el propósito de verificar la reproducibilidad.

### 3.3 Métodos estadísticos

Los datos obtenidos en cada uno de los estudios de los genes seleccionados se analizaron de acuerdo con las hipótesis correspondientes elaboradas para cada uno de ellos. Estos análisis están realizados en muestras de población. Por lo tanto, es necesario aplicar métodos estadísticos en su tratamiento. Estos métodos estadísticos corresponden a las dos grandes áreas de la Estadística: Estadística descriptiva y Estadística analítica o inferencial. La Estadística descriptiva se ha aplicado en primer lugar para conocer, describir y resumir la muestra. Posteriormente, se ha utilizado la Estadística inferencial para llevar a cabo el contraste de hipótesis.

Independientemente de las singularidades de cada gen analizado y de las variables de control incluidas en cada uno de los análisis, entre los estadísticos descriptivos se calcularon, para las variables cuantitativas, las medidas de tendencia central (media y mediana), medidas de dispersión (desviación típica, varianza, rango y coeficiente de variación) y medidas de forma (asimetría y apuntamiento). Y para las variables cualitativas se calcularon las frecuencias absolutas y relativas. La normalidad de las variables se comprobó mediante el test de Kolmogorov-Smirnov y gráficas de distribución de frecuencias. Cuando la variable original no mostró una distribución normal, se aplicaron transformaciones matemáticas (logaritmo, raíz cuadrada, potenciación al cuadrado, etc.) para conseguir la normalidad.

Para estimación de diferencia de medias de dos grupos independientes, se utilizó el test de la T de Student previa determinación de la homogeneidad de las varianzas mediante el test de Levene. Cuando el número de casos fue bajo, se

aplicó el test no paramétrico de Wilcoxon para la comparación de las medias de dos grupos independientes. En la comparación de medias de más de dos grupos, se utilizó el análisis de la varianza paramétrico, salvo cuando el número de casos en cada grupo hacía imprescindible recurrir al test no paramétrico. La comparación de las distribuciones de las frecuencias se realizó con el test del Chi-cuadrado de Pearson. Cuando las frecuencias esperadas fueron inferiores a cinco, se utilizó el test de Fisher o el de Montecarlo en tablas  $n \times n$ , siendo  $n > 2$ .

El grado de asociación entre dos variables continuas se estimó mediante técnicas de regresión lineal simple y múltiple incluyendo el control de potenciales variables de confusión. La interacción entre dos variables se estimó mediante el cálculo de la significación estadística del término de interacción en un modelo jerárquico. Mediante regresión logística simple o múltiple se estimaron los riesgos (*Odds Ratio*) asociados a cada una de las variables estudiadas. En todos los modelos ajustados se verificó siempre que se cumplieran los requisitos previos aplicando los test de control correspondientes. Además, este modelado estadístico estuvo siempre presidido por el principio de parsimonia.

### 3.3.1 Regresión lineal múltiple

La regresión múltiple se utiliza habitualmente para la predicción de una variable dependiente, por medio de la capacidad predictiva de otras variables independientes a través de una ecuación que se denomina *ecuación de regresión lineal múltiple*. El primer paso consiste en elegir qué variables serán explicativas de la variable dependiente considerada. Para poder utilizar este modelo es necesario que las variables consideradas cumplan una serie de requisitos, como los siguientes (Kleinbaum et al, 1988):

- La interdependencia entre las variables debe responder al modelo lineal. Para comprobar que se satisface esta condición se construyen previamente los correspondientes diagramas de dispersión. Si en algún caso se comprueba que los datos siguen un diagrama curvilíneo, se opta por practicar la transformación logarítmica de los mismos en aras de mejorar su linealidad.
- Los efectos de las variables deben poder sumarse entre sí para poder así predecir la variable dependiente.
- Todas las variables que componen la ecuación de regresión deben estar medidas como mínimo a nivel de intervalo.
- Ausencia de correlación entre variables independientes, ya que cuando alguna o todas las variables independientes se encuentran altamente

interrelacionadas (multicolinealidad) se resta fiabilidad a los resultados obtenidos.

- El método usado para resolver la ecuación de regresión es el método de mínimos cuadrados.

El modelo de regresión múltiple puede expresarse matemáticamente como:

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_k X_{ki} + e_{ii} \quad (\text{E: 3.2})$$

La notación  $X_{ki}$  indica el valor de la k-ésima variable independiente para el caso  $i$ . Los términos  $B_k$  son parámetros desconocidos y los términos  $e_{ii}$  son variables aleatorias, con media cero y varianza constante. Para construir intervalos de confianza para los coeficientes de regresión y probar hipótesis acerca de ellos, se supone que los  $e_{ii}$  tienen distribución normal de la variable dependiente para cada combinación lineal de los valores de las variables independientes.

Cuando el modelo consta de una variable dependiente y de dos variables independientes, puede ajustarse un plano en el espacio tridimensional a los puntos de los datos. Cuando el modelo contiene más de dos variables independientes se describe geoméricamente como un hiperplano. Los coeficientes representados como  $B$  se denominan "coeficientes de regresión parcial" ya que el coeficiente para una variable particular es ajustado por las otras variables independientes en el modelo (Rothman, 1986).

En una ecuación de regresión múltiple muchas veces se quiere conocer la importancia relativa de cada variable independiente. Para responder a esta pregunta es erróneo comparar la magnitud de los coeficientes de regresión parcial de las variables seleccionadas, ya que dicha magnitud depende de las unidades en que hayan sido medidas. Sólo si las variables independientes han sido expresadas en las mismas unidades, sus coeficientes de regresión parcial son directamente comparables. Un modo de hacer los coeficientes de regresión comparables es calcular los coeficientes  $\beta$  (BETA). Dichos coeficientes se obtienen directamente de los de regresión parcial sometiéndolos a estandarización, de modo que:

$$\beta = B_k (S_k + S_y) \quad (\text{E: 3.3})$$

donde  $S_k$  es la desviación estándar de la k-ésima variable independiente. Aunque este procedimiento es el más empleado para comparar magnitudes de variables, en un sentido estricto, estos coeficientes no sirven para todas las ocasiones, ya que también se ven afectados por las correlaciones de las variables independientes (Wayne, 1990).



Otra forma de determinar la importancia relativa de las variables independientes es considerar el incremento en el valor  $R^2$  cuando se introduce una nueva variable en una ecuación que ya contenía otras variables independientes. Este incremento  $\Delta$  responde a la siguiente fórmula siguiente:

$$R^2_{\Delta} = R^2 - R^2_{(i)} \quad (\text{E: 3.4})$$

donde  $R^2_{(i)}$  es el cuadrado del coeficiente de relación múltiple cuando todas las variables independientes excepto la  $i$ -ésima están presentes en la ecuación. Un elevado incremento en  $R^2$  indica que la variable en estudio proporciona una gran información acerca de la variable dependiente, que no es suministrada por las demás variables independientes incluidas en la ecuación (Kleinbaum et al, 1988).

### 3.3.2 Regresión logística

Mediante la regresión logística se calculan los coeficientes de un modelo probabilístico, constituido por un conjunto de variables independientes que mejor pronostica el valor de una variable dependiente dicotómica. Las variables independientes pueden ser continuas o categóricas. En las covariables categóricas con más de dos categorías se ha utiliza la primera categoría como referencia y el método de contraste simple (cada categoría de las variables de predicción, se compara con la categoría de referencia).

En la regresión logística, la estimación de los coeficientes alfa y beta del modelo se realiza mediante la *función de máxima verosimilitud* (en inglés *likelihood*), es decir, la estimación que hace que los datos observados sean los más verosímiles (probables). Para obtener el alfa y el beta que maximizan la función logística, se calcula la primera derivada de la *log-verosimilitud* y se iguala a cero. La ecuación no se puede resolver fácilmente por lo que hay que recurrir a métodos iterativos de resolución de ecuaciones como el de Newton-Rapshon. Evaluando el log-verosimilitud en los valores estimados de alfa y de beta, se obtiene un valor que se utiliza para contrastar el ajuste del modelo. Para obtener el error estándar de las estimaciones se calcula la inversa de la matriz del negativo de las segundas derivadas evaluadas en los alfas y beta estimados, obteniendo la matriz de varianzas-covarianzas de las estimaciones que permite calcular los intervalos de confianza (**IC**). Una vez estimado B, se puede calcular la **OR** (*odds ratio*), que indica la probabilidad de que los sujetos con el problema de salud estudiado estén expuestos al factor analizado, comparada con la probabilidad de que lo estén las personas sin el problema de salud. Esto se muestra a continuación:

$$\text{OR} = e^B \quad \text{IC al 95\% (B)} = B \pm 1,96 * \text{error estándar} \quad (\text{E: 3.5})$$

Clásicamente, el cálculo de las **OR** se ha reservado para los estudios de casos y controles, siendo la razón de prevalencia la medida de asociación típica de los estudios transversales. Sin embargo, en la actualidad está ampliamente aceptado el cálculo de las **OR** en los estudios de prevalencia, indicando que son **OR** de prevalencia.

Una vez obtenida la estimación de los coeficientes de la ecuación, el siguiente paso es contrastar la bondad de ajuste del modelo, es decir, si existe una asociación estadísticamente significativa. Aunque existen varias alternativas, se ha utilizado el test de Wald. El test de Wald contrasta la hipótesis nula de si el coeficiente de la ecuación es igual a cero. Está basado en el cálculo del coeficiente estandarizado tras dividir beta por su error estándar. El valor obtenido, en el caso de muestras grandes, se puede comparar con la raíz cuadrada de una Chi-cuadrado con un grado de libertad.

Otra forma de evaluar si aporta información la variable incluida en el modelo en la predicción del estado de salud, consiste en construir un test basado en la verosimilitud. Para ello se calcula la lejanía o discrepancia que hay entre el modelo ajustado con la variable de exposición, y el modelo saturado, donde cada individuo es una variable de exposición. Para calcular la lejanía (*deviance*) se realiza el siguiente test:

$$D = -2 \ln \left( \frac{\text{verosimilitud modelo ajustado}}{\text{verosimilitud modelo saturado}} \right) \quad (\text{E: 3.6})$$

Se calcula el negativo de dos veces el logaritmo del cociente de verosimilitud para garantizar la distribución Chi-cuadrado del test de contraste **D**. Si **D** es significativo, indica que el modelo ajustado se aleja del saturado y hace falta más información para explicar la variación. Si **D** no es significativo indica que no son necesarias más variables en el modelo. Para disminuir la lejanía se utilizan modelos multivariantes.

### 3.3.3 Determinación de las frecuencias alélicas del Equilibrio de Hardy-Weinberg

Un muestreo de individuos de una población proporciona números absolutos de genotipos que pueden convertirse en frecuencias relativas al dividirlos entre el número total de observaciones. Por otra parte, a partir de las frecuencias genotípicas observadas, se pueden determinar las frecuencias génicas o alélicas, teniendo en cuenta que cada genotipo autosómico tiene dos alelos.

Por ejemplo, sea una población con los genotipos A/A, A/B y B/B donde las frecuencias alélicas observadas son:

$$p = \frac{2 \text{ A/A} + \text{ A/B}}{2n} \quad q = \frac{2 \text{ B/B} + \text{ A/B}}{2n} \quad (\text{E: 3.7})$$

Siendo  $p$  y  $q$  las frecuencias alélicas para los alelos A y B respectivamente y  $n$  el número total de individuos.

La frecuencia de las tres combinaciones alélicas (A/A, A/B y B/B) es la siguiente:

$$p^2 + 2pq + q^2 = 1 \quad (\text{E: 3.8})$$

Una consecuencia importante de las relaciones entre genotipo y frecuencias alélicas es que las proporciones de los genotipos no cambian de generación en generación. El hecho de que los genotipos se distribuyan en proporción a las frecuencias de alelos individuales en una población y permanezcan constantes de generación en generación constituye el principio básico de la piedra angular de la genética poblacional, la ley de Hardy-Weinberg.

Se dice que una población se halla en *equilibrio de Hardy-Weinberg* cuando muestra las características básicas de la ley de Hardy-Weinberg. La prueba estadística que se emplea para la comparación de las frecuencias observada y esperada de los genotipos diferentes es el test de Ji cuadrado ( $\chi^2$ ). Para el ejemplo anterior, con dos alelos se tiene lo siguiente:

|            |        |        |        |
|------------|--------|--------|--------|
| Observadas | A/A    | A/B    | B/B    |
| Esperadas  | $p^2n$ | $2pqn$ | $q^2n$ |

El test  $\chi^2$  da lo siguiente:

$$\chi^2 = \frac{\sum (\text{Obs} - \text{Esp})^2}{\text{Esp}} \quad (\text{E: 3.9})$$

Según los grados de libertad y el error  $\alpha$  establecido se fija un valor crítico de  $\chi^2$ . Cuando el valor de  $\chi^2$  obtenido es mayor que el crítico, hay diferencias estadísticamente significativas entre las frecuencias observadas y las esperadas, y la población está en desequilibrio.

La ley de Hardy-Weinberg supone las siguientes condiciones:

1. La población es lo suficientemente grande como para que los errores de muestreo y efectos aleatorios sean despreciables.
2. El apareamiento de la población se da al azar.
3. No hay selección a favor o en contra de un fenotipo particular.
4. No hay otros factores como mutación, migración y deriva genética al azar.

Si en una determinada población no resultan verdaderas una o más de estas hipótesis, los resultados de la aplicación del test de equilibrio de Hardy-Weinberg pueden no ser válidos.

La ley de Hardy-Weinberg es uno de los conceptos fundamentales en genética de poblaciones y tiene tres propiedades importantes:

1. Las frecuencias alélicas predicen las frecuencias genotípicas.
2. En equilibrio, las frecuencias alélicas y genotípicas no cambian de generación en generación.
3. El equilibrio se alcanza con sólo una generación de apareamiento al azar.

Para el cálculo del equilibrio de Hardy-Weinberg se ha utilizado en primer lugar el programa *LINKUTYL* (Rockefeller linkage, 2004) y en los estudios posteriores el programa Arlequín (Arlequín, 2004).

### 3.3.4 Determinación del Desequilibrio de Ligamiento

Cuando se estudian diferentes variantes genéticas, un factor a tener en cuenta es el grado de asociación que existe entre ellos, denominado *desequilibrio de ligamiento*. Cuando dos variantes genéticas se hallan en desequilibrio de ligamiento, la asociación que existe entre ellos es mayor que la debida al azar, es decir existe una asociación preferente entre determinadas variantes genéticas. Una de las limitaciones de estos estudios es que los individuos analizados no deben estar emparentados, ya que se produciría un falseamiento del grado de asociación entre los marcadores polimórficos.

El grado de asociación entre dos pares de loci polimórficos se puede cuantificar mediante diferentes coeficientes. A continuación se estudian los dos más importantes.

**Valor  $\Delta$** 

El valor  $\Delta$  es uno de los valores que más se utiliza en la bibliografía para medir el grado de asociación entre dos loci polimórficos y se define matemáticamente como,

$$\Delta = \frac{P_{11} - P_{10} P_{01}}{(P_{10} q_{10} P_{01} q_{01})} \quad (\text{E: 3.10})$$

siendo  $P_{11}, P_{12}, P_{21},$  y  $P_{21}$  las frecuencias para los distintos genotipos (A/A, A/B, y B/B) y  $P_{10} = P_{11} + P_{12}$ ,  $q_{10} = 1 - P_{10}$ ,  $P_{01} = P_{11} + P_{21}$  y  $q_{01} = 1 - P_{01}$ .

Puesto que  $\Delta$  es un coeficiente de correlación, medida de asociación entre dos variables, la hipótesis del equilibrio de ligamiento ( $\Delta = 0$ ) es analizada usando un test  $\chi^2$ . Según los grados de libertad, en este caso uno, y el error  $\alpha$  establecido, se fija un valor crítico de  $\chi^2$ . Cuando el valor de  $\chi^2$  obtenido es mayor que el crítico, hay una asociación estadísticamente significativa entre los dos loci polimórficos y se afirma que existe desequilibrio de ligamiento significativo entre los dos polimorfismos analizados.

**Valor  $D$  y  $D'$** 

Para el cálculo de este parámetro de desequilibrio, se ha utilizado el programa *2BY2* de *LINKUTYL* (Rockefeller linkage, 2004). En este programa, se realiza un test exacto de Fischer con tablas 2x2. Además de los valores de  $p$ , para una cola y dos colas, el programa calcula el parámetro  $D$ , como valor absoluto y como porcentaje de su máximo ( $D' = D/D_{\max}$ ). Si el valor de  $D$  es negativo, el valor de  $D'$  se expresa como porcentaje de su mínimo.

**3.3.5 Análisis estadísticos para cada gen analizado**

En este apartado se presenta de manera detallada cada uno de los métodos estadísticos aplicados para cada gen analizado, siguiendo los postulados generales especificados en la subsección 3.4.

**3.3.5.1 Análisis estadístico para el CETP**

En el caso del análisis estadístico para el gen CETP se han empleado tests de  $\chi^2$  para las medidas categóricas y tests de t de Student de dos muestras para las medidas continuas. Se ha estimado la frecuencia alélica de los alelos B2 y APOE

con el método de conteo cromosómico y se aplicó un test de  $\chi^2$  para las comparaciones entre hombres y mujeres.

Para evaluar la relación entre los genotipos de CETP y los niveles lipídicos se han aplicado técnicas de análisis de covarianza que tenían en cuenta las relaciones familiares entre los miembros del estudio (la mayoría hermanos y primos). Se han tomado dos aproximaciones para realizar dichos análisis. En primer lugar, se ha empleado un enfoque de medidas repetidas con la asunción de una estructura de correlación intercambiable entre todos los miembros de una familia; para ello se ha contado con el soporte del programa PROC MIXED de SAS. Puesto que esta primera aproximación no representa con precisión la estructura de correlación verdadera en estas estructuras genealógicas, se ha aplicado también la medida del genotipo (Boerwinkle et al., 1988) mediante SOLAR, un programa de análisis de componentes de varianza para rasgos cuantitativos (*quantitative traits*) medidos en estructuras genealógicas de tamaño arbitrario (Almasy et al., 1998). El último enfoque aplicado considera los distintos tipos de relaciones en una estructura genealógica cuando se realiza un análisis de varianza sobre los genotipos definidos.

En estos análisis se han usado diferentes modelos para ajustar por variables de confusión potenciales (*potential confounders*). En primer lugar, se han obtenido resultados esencialmente en crudo que consideraban solamente la estructura familiar. En segundo lugar, se ha ajustado por edad, índice de masa corporal (IMC), consumo de tabaco y alcohol, betabloqueantes, y estado menopáusico y terapia de sustitución hormonal en las mujeres. En el análisis final, se han añadido al modelo los genotipos de APOE poniendo en un grupo los E2/E2 y E2/E3, en el segundo grupo los E3/E4 y E4/E4, y el E3/E3 en el grupo de referencia. Se han excluido los sujetos con genotipos E2/E4 que además eran bastante escasos.

Por otra parte se ha realizado un análisis de sensibilidad para estimar la validez y precisión de los coeficientes de regresión para las variables genotípicas de CETP cuando se han incluido en el modelo términos independientes. Como se han obtenido resultados similares para ambos sexos, se han analizado los datos conjuntamente para los dos sexos para aumentar la potencia estadística. Se han calculado los coeficientes de regresión y los intervalos de confianza al 95% para los genotipos B1B2 y B2B2 en comparación con el genotipo B1B1 mediante el ajuste de varios modelos de regresión lineal con variables mudas para términos categóricos y de interacción. Estos modelos son los siguientes:

- **Modelo 1:** genotipo CETP (B1B1, B1B2 y B2B2).
- **Modelo 2:** modelo 1 + sexo.
- **Modelo 3:** modelo 2 + IMC.

- **Modelo 4:** modelo 3+ consumo de tabaco (fumadores y no fumadores).
- **Modelo 5:** modelo 4 + consumo de alcohol (consumo y no consumo).
- **Modelo 6:** modelo 5 + genotipos APOE (E2, E3 y E4).

En todos los casos se ha tomado la primera categoría como referencia. Y los diagnósticos de regresión se han empleado para confirmar las asunciones y valorar la precisión de los cálculos.

Finalmente, mediante un test de  $\chi^2$  se ha estimado la ODDS de la prevalencia de ECV en el ciclo 5 del *Framingham Heart Study* para los sujetos con genotipos B1B2 o B2B2 que son familiares de los que tienen el genotipo B1B1. Las ECV comprenden el infarto de miocardio, la angina de pecho y la insuficiencia coronaria. Se ha empleado la regresión logística para ajustar estos odds ratios estimados sobre las covariantes. También se han empleado ecuaciones de estimación generalizada con un enlace logit para considerar la correlación entre las observaciones, obteniendo esencialmente los mismos resultados. Por lo tanto, en los resultados descritos se asume que las observaciones son independientes.

### 3.3.5.2 Análisis estadístico para el APOE

En el caso del análisis estadístico para el gen APOE se ha empleado el paquete estadístico SAS. Para mejorar la normalidad en las pruebas estadísticas con variables continuas, se han transformado los valores de niveles de triglicéridos a sus logaritmos decimales y los valores de ingesta de alcohol a sus raíces cuadradas. Se han aplicado tests de t de Student para grupos independientes con el objetivo de calcular las diferencias medias en variables continuas entre géneros.

Además, se ha utilizado el análisis de varianza de una vía (ANOVA) para las comparaciones múltiples de medias. También se ha usado ANOVA para calcular los valores de significación estadística ( $p$ ) en las tendencias lineales entre categorías mediante la partición de las sumas de los cuadrados intergrupos en componentes de tendencia.

Se han aplicado tests de Bonferroni para corregir las comparaciones múltiples. Y los tests de Chi-cuadrado han servido para comparar las diferencias entre porcentajes. Se calcularon los coeficientes de correlación de Pearson para describir las asociaciones originales entre variables. También se estimaron los coeficientes de correlación parciales con el ajuste por una o más variables de confusión. Se han realizado análisis de covarianza mediante el procedimiento del modelo lineal general (GLM) para evaluar la relación entre las concentraciones de LDL-C y las

variables independientes categóricas, controlando el efecto de las variables independientes continuas, y comprobando el efecto de los términos de interacción.

Se han utilizado técnicas de análisis de covarianza considerando las relaciones familiares entre los miembros del estudio (la mayoría hermanos y primos). Se ha empleado un enfoque de medidas repetidas, mediante el programa PROC MIXED, con la asunción de una estructura de correlación intercambiable entre todos los miembros de una familia. Finalmente, se ha utilizado análisis de regresión lineal múltiple con variables mudas para estimar la extensión, dirección y fortaleza de las relaciones entre varias variables independientes, con o sin términos de interacción, y concentraciones de LDL-C.

Se han creado los términos de interacción entre las variables mudas del consumo de alcohol y los genotipos de APOE como productos de dos factores. También se han empleado diagnósticos de regresión (análisis de residuos, influencia de puntos extremos y colinealidad) para confirmar las asunciones y evaluar la precisión de los cálculos.

### 3.3.5.3 Análisis estadístico para el APOA1

En el caso del análisis estadístico para el gen APOA1 se ha comprobado la normalidad de todas las variables continuas. Las variables TG y apoB se han transformado a logaritmos, y la ingesta de alcohol se ha transformado a raíz cuadrada para mejorar la normalidad en las pruebas estadísticas. Se han estimado las frecuencias alélicas con el método de cuenta génica. Para examinar si las frecuencias genotípicas estaban en equilibrio de Hardy-Weinberg se han aplicado tests de Chi-cuadrado. Se ha usado el test de t de Student para comparación de medias entre dos grupos independientes. Se ha realizado un análisis de varianza de una vía (ANOVA) para la comparación múltiple de medias, y se han calculado los valores de P para las tendencias lineales entre categorías mediante la partición de las sumas de cuadrados intergrupos en componentes de tendencia.

Debido a las marcadas diferencias por sexo en las variables de interés, se han realizado por separado los análisis estadísticos en hombre y mujeres. Se han calculado los coeficientes de correlación de Pearson para describir las asociaciones crudas entre variables continuas. Se han llevado a cabo análisis de covarianza mediante el modelo lineal general (GLM) para evaluar las relaciones entre los niveles de HDL-C o ApoA-I y las variables categóricas o continuas, y probar el efecto de los términos de interacción.

Finalmente, para estimar la extensión, sentido y fortaleza de las relaciones entre varias variables independientes, con o sin términos de interacción, y los



niveles de HDL-C o ApoA-I, se han aplicado análisis de regresión lineal múltiple con variables mudas (para términos categóricos y de interacción). Se han empleado diagnósticos de regresión, tales como análisis de residuos y tests de colinealidad, para comprobar las asunciones y evaluar la precisión de los cálculos. Todos los valores de P informados se han calculado con dos colas, y se han calculado los intervalos de confianza al 95% para los coeficientes estimados. Los cálculos estadísticos se han realizado con el paquete estadístico SPSS (*Statistical Package of Social Sciences*) para MS Windows versión 8.0.

#### **3.3.5.4 Análisis estadístico para el LIPC**

En el caso del análisis estadístico para el gen LIPC, se han aplicado el test de t de Student y el de Chi-cuadrado para comprobar las diferencias entre medias y porcentajes. Se ha usado un análisis de covarianza para evaluar las relaciones entre los genotipos de LIPC y las variables dependientes. Se han incluido en los modelos los términos de interacción para comprobar la hipótesis nula de que no hay interacción entre los genotipos de LIPC y la grasa en dieta. En primer lugar, se han obtenido los resultados crudos que tenían en cuenta sólo las relaciones familiares entre los miembros del estudio (la mayoría hermanos y primos) mediante el programa PROC MIXED de SAS (v.8) (Otvos et al., 1992).

En segundo lugar, se han obtenido los resultados ajustados después de haber tenido en cuenta las relaciones familiares y los factores de confusión potenciales (sexo, edad, IMC, tabaco, alcohol, estrógenos en mujeres, betabloqueantes y energía). Se han analizado conjuntamente los hombres y las mujeres debido a que no se ha detectado ninguna heterogeneidad del efecto alélico por sexo. Para describir el efecto de interacción entre el genotipo de LIPC y la grasa en la determinación del metabolismo de HDL-C, se han calculado los valores predichos para cada variable dependiente en el correspondiente modelo de regresión. Se ha calculado la raíz cuadrada del coeficiente de correlación como medida de la bondad del ajuste de los modelos. Además, se han añadido a las figuras las medias en las categorías de grasas como una segunda medida. Finalmente, todos los valores de P eran de dos colas.

#### **3.3.5.5 Análisis estadístico para el SR-BI**

En el caso del análisis estadístico para el gen SR-BI, se han estimado las frecuencias alélicas por el método del conteo de genes y se han calculado los intervalos de confianza al (CI) 95%. Se han usado tests de Chi-cuadrado para comprobar las diferencias en porcentajes. Este tipo de test se ha aplicado también para cada polimorfismo para estimar si hay diferencias estadísticamente

significativas entre las frecuencias genotípicas observadas y las frecuencias esperadas, asumiendo que existe equilibrio de Hardy-Weinberg.

Cuando el valor de P del test era mayor que 0,05, no se podía rechazar la hipótesis nula de que no hay diferencias entre los valores observados y los esperados. Entones, la conclusión era que no había desviación estadísticamente significativa del equilibrio de Hardy-Weinberg. Se han construido los haplotipos usando dos (exón 8, intrón 5) o tres marcadores en el gen SRB-I (exón 8, intrón 5 y exón 1) mediante el programa *haplo.score* (Schaid et al., 2002), que estima las frecuencias haplotípicas por medio del algoritmo de la maximización de expectación (EM) y realiza tests estadísticos de asociación usando estadísticos marcadores para rasgos cuantitativos y cualitativos con opciones de ajuste por covariantes. Puesto que en *haplo.score* se asume que las observaciones son independientes, se ha restringido los análisis de datos a 1.607 sujetos no relacionados (donde se ha seleccionado aleatoriamente un individuo de cada estructura familiar). Se ha calculado también la D' de Lewontin (Lewontin, 1964), mediante las frecuencias haplotípicas estimadas por *haplo.score*, para cada par de marcadores en la evaluación del desequilibrio de ligamiento.

Se ha comprobado la normalidad de todas las variables continuas y se han aplicado transformaciones logarítmicas a los triglicéridos y el tamaño de partículas. Se ha usado las técnicas de análisis de covarianza (ANCOVA), con el programa *PROC GENMOD* de SAS, para evaluar las relaciones entre la variación del gen SR-BI (genotipos para el exón 1, el intrón 5 y el exón 8) y las concentraciones y tamaños de partícula de lípidos y lipoproteínas, teniendo en cuenta las relaciones familiares entre los miembros del estudio (la mayoría hermanos y primos) como se ha indicado previamente (Ordovas et al., 2000). En primer lugar, para cada variante genética, se han evaluado los efectos crudos considerando solamente la estructura familiar, y después se ha ajustado por factores de confusión adicionales (edad, IMC, tabaco, ingesta de alcohol, uso de betabloqueantes, genotipo de apoE y situación menopáusica y terapia estrógena en mujeres. Se han considerado tres categorías para el control de apoE: E2 (sujetos e2/e2 + e2/e3); E3 (sujetos e3/e3) y E4 (sujetos e3/e4 + e4/e4). La prevalencia de cada categoría en esta población era la siguiente: 7,5% para E2; 80,5% para E3 y 12% para E4.

Para probar la hipótesis nula de que no hay interacción entre los genotipos de SR-BI y la diabetes de tipo 2 en la determinación de los lípidos plasmáticos y los tamaños de partículas, se han aplicado técnicas ANCOVA con términos de interacción jerárquica, después de haber ajustado por factores de confusión potenciales. Y para examinar el efecto de las interacciones, se han realizado análisis estratificados adicionales por el estado de la diabetes. En estos análisis, se han estimado medias con ajuste multivariado en diabéticos y no diabéticos por las

variantes del gen SR-BI. Considerando la relación entre lípidos y variantes del gen SR-BI especificada por el sexo que se ha descrito previamente, se han ajustado diferentes modelos para hombres y mujeres. Además, después de haber examinado el sentido del efecto, y con el propósito de mejorar la potencia estadística, se han analizado conjuntamente los datos de hombres y mujeres incluyendo una variable muda para el sexo. Finalmente, para probar la asociación entre haplotipos y rasgos cuantitativos, se ha aplicado el software *haplo.score* y el procedimiento GLM en SAS para estimar los niveles medios de lípidos para cada haplotipo.

Puesto que los haplotipos para sujetos con más de un locus heterocigótico no pueden determinarse sin ambigüedad, a los sujetos se les han asignado entre dos y ocho haplotipos posibles. A los sujetos con cero o un locus heterocigótico se les han asignado dos haplotipos, cada uno con la probabilidad de uno (fase conocida). A los sujetos con dos loci heterocigóticos se les ha asignado cuatro haplotipos, y a los heterocigóticos para los tres marcadores se les ha asignado ocho haplotipos. Para los dos últimos grupos se han inferido las probabilidades de haplotipos condicionalmente a sus genotipos y a las frecuencias haplotípicas de la población estimadas. Se han incluido las distintas probabilidades haplotípicas en el programa PROC GLM mediante regresión de mínimos cuadrados ponderados. Todos los valores de P eran de doble cola. Se han ajustado los valores de P para pruebas múltiples sólo cuando se han realizado más de dos comparaciones de medias ajustadas en el mismo test estadístico, tal como en el caso del exón 8 y en los análisis de haplotipos. En estos casos, se ha usado un ajuste de Scheffe para comparaciones múltiples. No se han considerado ajustes adicionales para el número total de tests aplicados en el caso del gen SR-BI.

### 3.3.5.6 Análisis estadístico para el PLIN

En el caso del análisis estadístico para el gen PLIN, se han aplicado tests de Chi-cuadrado (Pearson, test exacto de Fisher, o la aproximación de Monte Carlo) para probar las diferencias entre las frecuencias observadas y las esperadas, asumiendo el equilibrio de Hardy-Weinberg, para comprobar el desequilibrio de ligamiento, y para comprobar las diferencias entre porcentajes. Se han estimado los coeficientes de desequilibrio de ligamiento de pares mediante el programa LINKAGE. También se han calculado los coeficientes D y D' ( $D/D_{\max}$ ). Se han estimado los haplotipos mediante el programa EH (Rockefeller haplotype, 2004) que utiliza el algoritmo de maximización de expectación para obtener las estimaciones de umbral máximo de las frecuencias haplotípicas.

Se ha comprobado la distribución normal de todas las variables continuas. Se han transformado logarítmicamente los valores de triacilglicéridos (TG). Se han aplicado tests paramétricos para comparar medias. Además, cuando el número de

casos en cada subgrupo era muy pequeño, se han empleado tests no paramétricos (Mann-Whitney o Kruskal-Wallis).

Se ha utilizado análisis de regresión lineal multivariado con variables mudas para términos categóricos para probar la hipótesis nula de que no existe asociación entre las variantes genéticas y los fenotipos relacionados con la obesidad. Las covariantes principales eran sexo, edad, IMC, origen, diabetes, tabaco, consumo de alcohol, medicación, actividad física y educación. Se han estimado en los modelos los coeficientes de regresión y las medias ajustadas para cada predictor. Se ha comprobado la homogeneidad de los efectos alélicos según el sexo, o los factores genéticos o ambientales, mediante la introducción de los correspondientes términos de interacción en el modelo de regresión lineal de mayor parsimonia. Se han empleado procedimientos de diagnóstico de regresión estándar para asegurar la adecuación de estos modelos.

En el análisis de casos y controles, se ha definido dicotómicamente la obesidad como  $IMC \geq 30 \text{ Kg/m}^2$ . Para estimar el riesgo se ha ajustado modelos de regresión logística en la forma siguiente: se han comparado con el tipo natural la odds ratio (OR) y el intervalo de confianza (CI) al 95% de la obesidad asociada con la presencia de cada variante genética. También se han ajustado modelos de regresión logística múltiple para controlar por el efecto de las covariantes y los modificadores de efecto. Los análisis de asociación se han llevado a cabo mediante el programa estadístico SPSS (*Statistical Package of Social Sciences*) para MS Windows versión 10.0.

### 3.4 Enfoque de Ingeniería informática y recursos metodológicos

La bioinformática debe plantearse que, en general, la Bioinformática no es un fin, sino un medio. Así, la Bioinformática debe ser entendida con una naturaleza dual: ciencia / ingeniería. La Bioinformática como Ciencia debe buscar modelos que expliquen computacionalmente los procesos biológicos relacionados principalmente con la Genómica y la Proteómica. Por otra parte, la Bioinformática como Ingeniería debe buscar soluciones prácticas mediante procesos suficientemente probados y basados en conocimientos científicos.

El contenido de esta sección se corresponde con esta segunda naturaleza. La Bioinformática necesita de recursos básicos para poder llevar a cabo sus actividades. Dichos recursos pueden ser de diversa naturaleza, pero se pueden agrupar en dos grandes categorías:

- Generales o inespecíficos: Los que utilizan otras disciplinas científicas y tecnológicas. Por ejemplo, software y herramientas estándares (lenguajes de programación, compiladores y utilidades para programación, entornos de desarrollo de sistemas basados en la ingeniería de componentes, herramientas de modelado de ciclo completo, herramientas de ofimática, etc.); equipos informáticos de propósito general; y redes y otros recursos distribuidos.
- Particulares o específicos: Los que son exclusivos de la disciplina e incluso se han diseñado y creado en el seno de la misma. Por ejemplo, software y herramientas bioinformáticas; sistemas informáticos para la gestión y el control de instrumentos (Genómica y Proteómica); redes para la computación distribuida (GRIS); redes para el intercambio y almacenamiento de información genómica y proteómica (GenBank, Entrez, etc.); sistemas tecnológicos de análisis (micromatrices de ADN, ARN, ADNc, SNP, tejidos, etc.); y técnicas de análisis (estadísticas, lingüísticas computacionales, Inteligencia Artificial y Minería de Datos).

Todo ello ha obligado a la aplicación de enfoques más formales y globales desde la Ingeniería del Software, como el desarrollo de herramientas que integran varias funcionalidades, la incorporación de interfaces gráficas y ayudas para el usuario, la infraestructura para el mantenimiento y gestión de configuraciones y el acceso a versiones libres y comerciales de software.

Además de los recursos mencionados, en Bioinformática se debe tener un conjunto de métodos y técnicas que permitan resolver los problemas con precisión, eficiencia y sentido científico. Hay que considerar que la Bioinformática desarrolla esencialmente sistemas informáticos y es el soporte de los estudios científicos. Por tanto, se pueden plantear dos grandes categorías de técnicas y métodos: los que se aplican como enfoque de ingeniería informática; y los que se aplican como enfoque de ciencia colaborativa (Biología molecular, Genética, etc.)

Fundamentalmente, la Bioinformática produce software, sea en forma de herramientas aisladas para propósitos específicos, o en forma de sistemas integrados. Por tanto, se tienen que aplicar los principios, normas, metodologías y técnicas provenientes de la Ingeniería del Software. Y además, las soluciones producidas deben obedecer a los siguientes criterios (Sommerville, 2004):

- Calidad: ajustar los procesos y los productos a las normas UNE, ISO, IEEE.
- Robustez: los productos deben ofrecer cierto grado de tolerancia a fallos.
- Fiabilidad: los procesos y resultados deben ser replicables bajo las mismas condiciones.

- Facilidad de uso: la dificultad del producto no debe estar en el uso.
- Portabilidad: se deben diseñar los productos para que puedan ofrecer la posibilidad de ejecutarse en distintas plataformas informáticas.

Todo ello siguiendo los principios de la Ingeniería del Software: Principio del Ciclo del Software, Principio del Modelo de Desarrollo, Principio de la Integración en Áreas y Procesos de Negocio (investigación científica), Principio de la Dualidad Proceso/Producto, Principio del Desarrollo por medio de Proyectos, Principio de la Gestión de Configuraciones (control de las sucesivas versiones de los productos), y Principio de la Gestión del Cambio (control de los sucesivos cambios desde los requisitos iniciales) (Sommerville, 2004).

Por lo tanto, en la Bioinformática, como enfoque de Ingeniería del Software, se debe proceder, desde un punto de vista formal y práctico, a la elección de técnicas según el Ciclo del Software y el Modelo de Desarrollo aplicado; de técnicas para la Integración; y de técnicas para, por ejemplo, la estimación de recursos (COMOMO II), planificación de tareas (PERT, GANTT), dirección, control y supervisión de proyectos; y de técnicas para la Gestión de Configuraciones y del Cambio.

En esta tesis se ha adoptado sistemáticamente este enfoque de ingeniería y por ello en esta sección describen los recursos conceptuales y materiales necesarios para poder obtener soluciones bioinformáticas eficaces y robustas. En principio se abordan las metodologías y elementos de representación. A continuación se tratan las bases de datos y las técnicas de almacenamiento y salvaguarda de los datos. Después se describe el instrumental de laboratorio empleado en los procesos científicos a los que la Bioinformática da soporte. A continuación se describen las herramientas informáticas empleadas, tanto en el desarrollo como en el análisis. Y finalmente, se describe el marco de auditoría para el control y aseguramiento de la calidad de los sistemas bioinformáticas.

### 3.4.1 Metodologías y elementos de representación

En esta tesis se intenta aplicar un enfoque formal de Ingeniería del Software sobre la integración de los procesos bioinformáticos desarrollados en esta tesis. El primer aspecto a abordar es el desarrollo de los principios fundamentales y normas de la Ingeniería del Software sobre el proceso para obtener el producto en forma de soluciones bioinformáticas. El segundo punto es el de la elección de notaciones de representación de los modelos de los sistemas a desarrollar y, en segundo lugar, la o las metodologías que sirven para obtener dichos modelos usando la notación.

En este sentido, la elección no ha revestido de ninguna dificultad. Puesto que se busca una metodología orientada al proceso, que sea ampliamente utilizada, y que esté amparada por una o varias normativas técnicas, la decisión ha sido aplicar el Proceso Unificado para el Desarrollo de Software (Jacobson et al., 2000), que se suele abreviar como RUP (del término inglés *Rational Unified Process*). Con respecto a la notación se ha elegido en lenguaje UML (Booch et al., 1999). En esta subsección se describen brevemente, en este orden, UML y RUP.

#### **3.4.1.1 Principios fundamentales de la Ingeniería del Software y arquitectura formal**

La Ingeniería del Software, como disciplina de ingeniería, presenta un esquema formal que parte de los principios fundamentales y finaliza en el producto (Figura 3.4) (Coltell, 2003). En este marco, los principios fundamentales son asertos de ingeniería que prescriben restricciones sobre soluciones de problemas de software o sobre el proceso de desarrollo de software, se evalúan rigurosamente en la práctica, y se juzgan sobre la base de la utilidad, la relevancia y la significación.

En Bourque et al. (2002) se plantea un cuerpo de principios fundamentales de la Ingeniería del Software que es el resultado de un proceso de revisión, iniciado en 1996 bajo el auspicio del *IEEE Software Engineering Standards Executive Committee*, y continuado con dos workshops, dos estudios Delphi y la publicación posterior en una web ([http:// www.lrgl.uqam.ca/fpse/](http://www.lrgl.uqam.ca/fpse/)) para su revisión global. La versión final se compone de 15 principios que son los siguientes:

- A) Aplicar y utilizar mediciones cuantitativas en la toma de decisiones.
- B) Construir por y para reutilizar.
- C) Controlar la complejidad con múltiples perspectivas y niveles de abstracción.
- D) Definir artefactos de software rigurosamente.
- E) Establecer un proceso de software que provea flexibilidad.
- F) Implementar un enfoque disciplinado y mejorarlo continuamente.
- G) Invertir en el conocimiento del problema.
- H) Gestionar la calidad a través del ciclo de vida tan formalmente como sea posible.

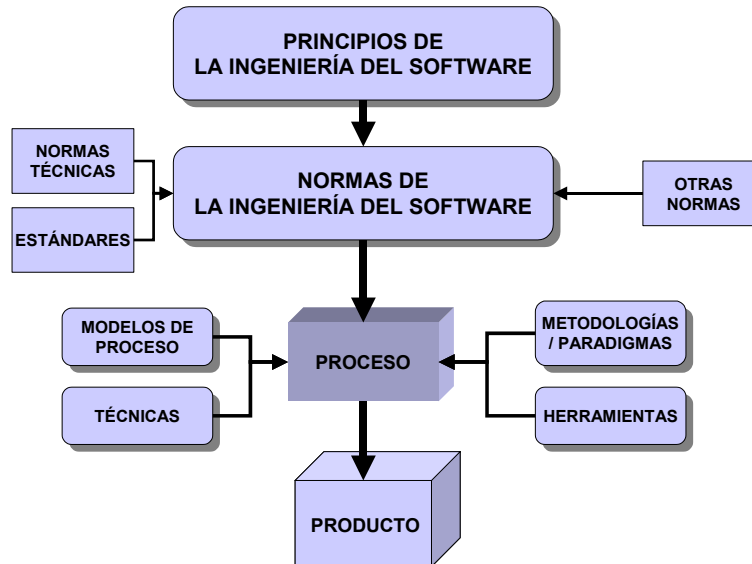


Figura 3.4. Estructura formal de la Ingeniería del Software

- I) Minimizar la interacción de componentes de software.
- J) Producir software en una forma escalonada.
- K) Fijar objetivos de calidad para cada producto a librar.
- L) Puesto que el cambio es inherente al software, procurar planificarlo y gestionarlo.
- M) Puesto que las negociaciones y compensaciones son inherentes a la Ingeniería del Software, hacerlas aflorar explícitamente y documentarlas.
- N) Mejorar el diseño mediante el estudio de soluciones anteriores a problemas similares.
- O) La incertidumbre es inevitable en la Ingeniería del Software. Por lo tanto, identificarla y gestionarla de forma que se pueda paliar.

Está claro que en esta tesis no se han aplicado todos estos principios puesto que no se tocan todos los aspectos de la disciplina. Se han tomado los más importantes y significativos que están relacionados directamente con la producción de sistemas bioinformáticos, tomando éstos como una variedad o subconjunto de sistemas software. Por lo tanto, la estructura formal de la Ingeniería del Software Bioinformático, como adaptación de la estructura general se muestra en la Figura 3.5.



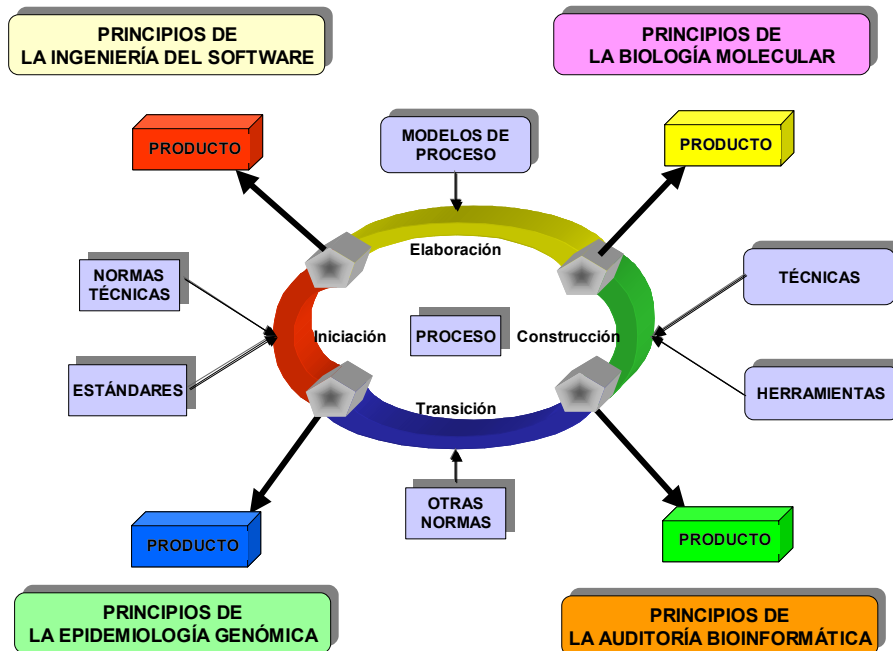


Figura 3.5. Estructura formal de la Ingeniería del Software Bioinformático

En la Figura 3.5 se puede ver que también se consideran los principios de las otras disciplinas asociadas y que definen el marco conceptual en el que se adscribe el proceso bioinformático general. Los artefactos estructurales que aparecen (normas técnicas, estándares, otras normas, modelos de proceso, técnicas y herramientas) corresponden a los artefactos técnicos en el desarrollo disciplinario de cada uno de los tipos de principios.

#### 3.4.1.2 La notación UML para el Modelo Orientado a Objetos

La notación UML (*Unified Modeling Language*) fue diseñada principalmente para el modelado de sistemas de software orientado a objetos por James Rumbaugh y Grady Booch, a los que se unió Ivar Jacobson. En 1997, tras haberse convertido casi en un estándar de facto, la versión UML 1.0 se sometió a la evaluación de la organización OMG (*Object Management Group*) (OMG, 2004) cuyas sucesivas mejoras pasaron a ser las correspondientes versiones hasta 1999, en que se publicó la versión 1.3 como el estándar OMG. A partir de entonces, UML se sometió a la evaluación de una *Task Force* de ISO (*International Standard Organization*) (ISO, 2004) con la intención de publicar una versión UML ISO. Actualmente, aunque continua el proceso de revisión en ISO, se está procediendo paralelamente a la

elaboración de UML versión 2.0 (UML, 2004), algunas de cuyas partes ya se están aplicando. Sin embargo, no se puede afirmar taxativamente que UML 2 es ISO.

UML no es una metodología de diseño, a diferencia de OMT (de James Rumbaugh y otros colaboradores), Booch Method (de Grady Booch), u OOSE (de Ivar Jacobson). Por lo tanto, el modelo conceptual de UML debe entenderse solamente en los límites estrictos de la representación de sistemas orientados a objetos, pero no del proceso de representación. El modelo conceptual del lenguaje lo componen tres elementos: bloques básicos de construcción, reglas de combinación de los bloques y mecanismos comunes. Los bloques básicos de construcción son los siguientes:

- *Elementos*: son las abstracciones de elementos del problema.
- *Relaciones*: son las ligaduras entre los elementos.
- *Diagramas*: son las agrupaciones de colecciones de elementos y relaciones.

Hay cuatro tipos de elementos que son los siguientes:

- *Elementos estructurales*: constituyen los nombres del modelo o las partes estáticas del mismo. Hay siete tipos distintos.
- *Elementos de comportamiento*: constituyen los verbos del modelo o las partes dinámicas del mismo. Hay dos tipos distintos.
- *Elementos de agrupación*: son cajas de descomposición del modelo que ayudan a estructurarlo en partes organizativas.
- *Elementos de anotación*: son comentarios o las partes explicativas del modelo.

Hay cuatro tipos de relaciones que son las siguientes:

- *Dependencia*: relación semántica entre elementos.
- *Asociación*: relación estructural entre elementos, principalmente entre clases y entre objetos.
- *Generalización*: relación de especialización y/o generalización entre elementos, principalmente entre clases y entre objetos.
- *Realización*: relación semántica entre clasificadores.

La *arquitectura* de un sistema en UML (arquitectura software) es un artefacto que permite decidir sobre la organización de un sistema software, mediante la selección de elementos estructurales y sus interfaces; sobre el comportamiento de

dicho sistema, mediante la especificación de operaciones en las colaboraciones entre los elementos estructurales; y el estilo arquitectónico, mediante la composición de elementos estáticos y dinámicos. La arquitectura software también se relaciona con otros aspectos también importantes: la utilización, la funcionalidad, el rendimiento, la reutilización, las restricciones económicas y tecnológicas, etc.

Sin embargo, cualquier sistema software es tan complejo que debe ser estudiado desde diferentes perspectivas en distintos momentos del proyecto y bajo intereses distintos. Los artefactos en UML que encarnan estas perspectivas e intereses son las vistas. Las vistas son las proyecciones de la organización y la estructura del sistema que se centran en un aspecto particular del sistema. Así, en UML se presentan cinco vistas según sendos criterios: vista de casos de uso, vista de diseño, vista de procesos, vista de implementación y vista de despliegue.

Como las vistas no son suficientes para describir la arquitectura por su generalidad se utilizan los diagramas de UML, que son la representación gráfica de un conjunto de elementos como grafo conexo donde los nodos son los elementos y los arcos son las relaciones. Visualizan el sistema desde distintas perspectivas, que son las proyecciones de los modelos según diferentes criterios. Hay nueve diagramas en UML agrupados en cinco vistas.

En la Tabla 3.3 se muestra la adscripción de cada uno de los diagramas presentados a las cinco vistas que componen la arquitectura del sistema. En el Anexo 12.2 se describe con mayor amplitud esta notación. Una de las herramientas utilizadas para la aplicación de la notación UML ha sido Rational Rose Enterprise de Rational/IBM (Figura 3.6).

Tabla 3.3. Vistas y diagramas de UML

|                                      | Vista de Casos de Uso |      | Vista de Diseño |      | Vista de Procesos |      | Vista de Implementación |      | Vista de Despliegue |      |
|--------------------------------------|-----------------------|------|-----------------|------|-------------------|------|-------------------------|------|---------------------|------|
|                                      | Est.                  | Din. | Est.            | Din. | Est.              | Din. | Est.                    | Din. | Est.                | Din. |
| Diagrama de Casos de Uso             | X                     |      |                 |      |                   |      |                         |      |                     |      |
| Diagrama de Interacción-Secuencia    |                       | X    |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Interacción-Colaboración |                       | X    |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Clases                   |                       |      | X               |      | X                 |      |                         |      |                     |      |
| Diagrama de Objetos                  |                       |      | X               |      | X                 |      |                         |      |                     |      |
| Diagrama de Estados                  |                       | X    |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Actividades              |                       |      |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Componentes              |                       |      |                 |      |                   |      | X                       |      |                     |      |
| Diagrama de Despliegue               |                       |      |                 |      |                   |      |                         |      | X                   |      |

Est.: Vista Estática Din.: Vista Dinámica

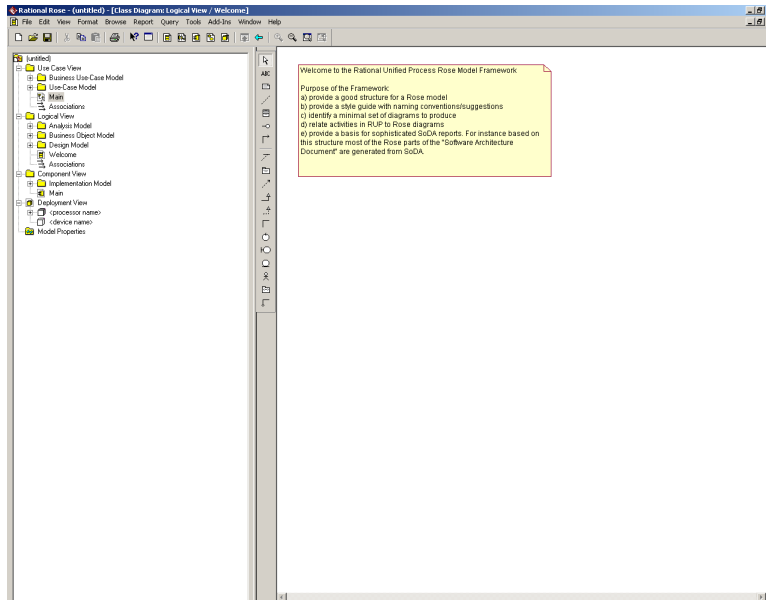


Figura 3.6. Interfaz principal de Rational Rose Enterprise

### 3.4.1.3 Proceso Unificado de Rational

El Proceso Unificado de Rational es un proceso iterativo. Un enfoque iterativo propone una comprensión incremental del problema a través de refinamientos sucesivos y un crecimiento incremental de una solución efectiva a través de varias versiones. Como parte del enfoque iterativo se encuentra la flexibilidad para acomodarse a nuevos requisitos o a cambios tácticos en los objetivos del negocio. También permite que el proyecto identifique y resuelva los riesgos más bien pronto que tarde. Las actividades del Proceso Unificado de Rational destacan en la creación y el mantenimiento de modelos que proporcionan representaciones semánticas del sistema software que se está desarrollando. Además, estos modelos se basan en los conceptos de objeto y clase y las relaciones entre ellos, y utilizan UML como la notación común.

El desarrollo bajo el Proceso Unificado está *centrado en la arquitectura*. El proceso se centra en establecer al principio una arquitectura software que guía el desarrollo del sistema. Con ello se facilita el desarrollo en paralelo, se minimiza la repetición de trabajos y se incrementa la probabilidad de reutilización de componentes y el mantenimiento posterior del sistema.

Las actividades de desarrollo bajo el Proceso Unificado están *dirigidas por los casos de uso*. El Proceso Unificado pone un gran énfasis en la construcción de sistemas basada en una amplia comprensión de cómo se utilizará el sistema que se entregue. Las nociones de los casos de uso y los escenarios se utilizan para guiar el flujo de procesos desde la captura de los requisitos hasta las pruebas, y para proporcionar caminos que se pueden reproducir durante el desarrollo del sistema.

El Proceso Unificado es un *proceso configurable*. Aunque un único proceso no es adecuado para todas las organizaciones de desarrollo de software, el Proceso Unificado es adaptable y puede configurarse para cubrir las necesidades de proyectos que van desde pequeños equipos de desarrollo de software hasta grandes empresas de desarrollo. También se basa en una arquitectura de proceso simple y clara, que proporciona un marco común a toda una familia de procesos y que, además, puede variarse para acomodarse a distintas situaciones.

El Proceso Unificado soporta las *técnicas orientadas a objetos*. Cada modelo es orientado a objetos. Los modelos del Proceso Unificado se basan en los conceptos de objeto y clase y las relaciones entre ellos, y utilizan UML como la notación común.

El Proceso Unificado impulsa un *control de calidad* y una *gestión del riesgo* objetivos y continuos. La evaluación de la calidad va contenida en el proceso, en

todas las actividades, e implicando a todos los participantes, mediante medidas y criterios objetivos. La gestión del riesgo va contenida en el proceso, de manera que los riesgos para el éxito del proyecto se identifican y se acometen al principio del proceso de desarrollo, cuando todavía hay tiempo de reaccionar.

El Proceso Unificado tiene una estructura matricial donde se relacionan esfuerzos y tiempos (Figura 3.7):. Los tiempos están definidos por las fases y las iteraciones. Los esfuerzos están definidos por los flujos de trabajo del proceso y de soporte.

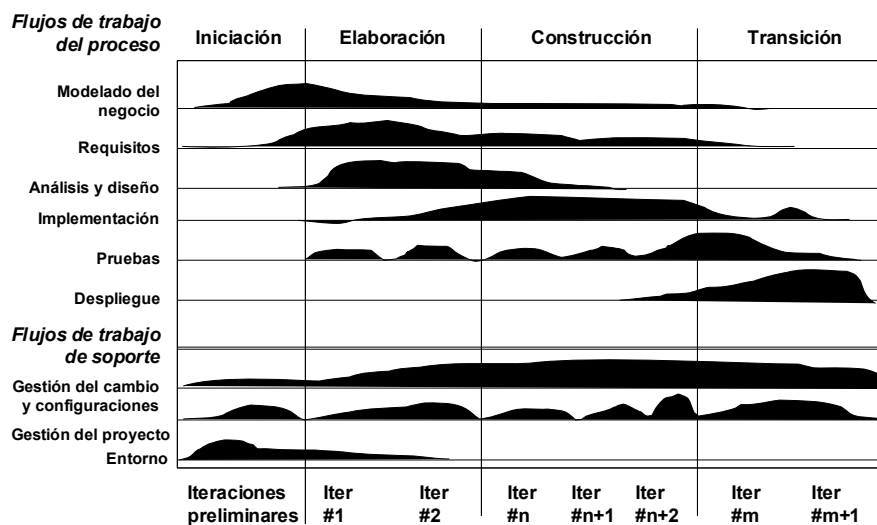


Figura 3.7. El ciclo de vida del desarrollo del software (Fuente: Rational Corporation)

El paso a través de las cuatro fases principales constituye un *ciclo de vida del desarrollo*, y produce una generación de software. La primera pasada a través de las cuatro fases se denomina ciclo de desarrollo inicial. A menos que acabe la vida del producto, un producto existente evolucionará a la siguiente generación repitiendo la misma secuencia de inicio, elaboración, construcción y transición.

Los *modelos* son el tipo de artefacto más importante en el Proceso Unificado de Rational. Hay nueve modelos que en conjunto cubren todas las decisiones importantes implicadas en la visualización, especificación, construcción y documentación de un sistema con gran cantidad de software. Son los siguientes:

1. **Modelo del negocio:** establece una abstracción de la organización.
2. **Modelo del dominio:** establece el contexto del sistema.

3. **Modelo de casos de uso:** establece los requisitos funcionales del sistema.
4. **Modelo de análisis** (opcional): establece un diseño de las ideas.
5. **Modelo de diseño:** establece el vocabulario del problema y su solución.
6. **Modelo del proceso** (opcional): establece los mecanismos de concurrencia y sincronización del sistema.
7. **Modelo de despliegue:** establece la topología hardware sobre la cual se ejecutará el sistema.
8. **Modelo de implementación:** establece las partes que se utilizarán para ensamblar y hacer disponible el sistema físico.
9. **Modelo de pruebas:** establece las formas de validar y verificar el sistema.

En cada uno de los flujos de trabajo del ciclo de vida del desarrollo del software se trabaja con los modelos descritos, pero no con todos al mismo tiempo, sino siguiendo una secuencia lógica determinada por el flujo de trabajo y la naturaleza del modelo. En la Tabla 3.4 se muestra qué modelos se manejan en cada uno de los flujos de trabajo del proceso de desarrollo.

**Tabla 3.4. Modelos y flujos de trabajo del Proceso Unificado**

|                          | Modelado del Negocio | Requisitos | Análisis | Diseño | Implementación | Prueba | Despliegue |
|--------------------------|----------------------|------------|----------|--------|----------------|--------|------------|
| Modelo del Negocio       | X                    |            |          |        |                |        |            |
| Modelo del Dominio       | X                    | X          |          |        |                |        |            |
| Modelo de Casos de Uso   |                      | X          |          |        |                |        |            |
| Modelo de Análisis       |                      |            | X        |        |                |        |            |
| Modelo de Diseño         |                      |            |          | X      |                |        |            |
| Modelo de Procesos       |                      |            |          | X      |                |        |            |
| Modelo de Despliegue     |                      |            |          | X      |                |        | X          |
| Modelo de Implementación |                      |            |          |        | X              |        | X          |
| Modelo de Prueba         |                      |            |          |        |                | X      | X          |

En el Anexo 12.3 se describe con mayor amplitud esta metodología.

### 3.4.2 El proceso de búsqueda y detección de polimorfismos

El proyecto para la elaboración de un Modelo Genómico Cardiovascular, que recoja y documente los polimorfismos de los genes relacionados directamente con manifestaciones genotípicas y fenotípicas de las patologías cardiovasculares, los factores de protección y la eficacia a la intervención terapéutica, debe abordarse, desde el punto de vista metodológico, teniendo en cuenta los siguientes aspectos:

1. Para la identificación de genes se deben aplicar las búsquedas bibliográficas en PubMed y otras bibliotecas digitales disponibles, siempre contando con el conocimiento experto previo y los resultados experimentales propios, para centrar las búsquedas en determinados genes (Lada et al., 2003).
2. Para la identificación de polimorfismos se deben aplicar las mismas estrategias que en el punto anterior.
3. Debe realizarse un proceso de identificación preliminar de genes y sus polimorfismos significativos, desde el punto de vista cardiovascular, en la literatura publicada.
4. Para la documentación de los polimorfismos y los SNP asociados se deben realizar búsquedas en determinadas bases de datos de NCBI GenBank.
5. Para almacenar la información obtenida sobre los genes, polimorfismos, SNP, proteínas y evidencia publicada se debe diseñar y utilizar una base de datos (Olken, 2002). Teniendo en cuenta la disponibilidad de herramientas estándar accesibles al resto de participantes en el proyecto, se ha determinado la aplicación de MS Access. En particular, se ha seleccionado la versión 2000 por su disponibilidad inmediata. Sin embargo, cualquier sistema de gestión de base de datos relacional puede servir.
6. Para extraer las secuencias seleccionadas de polimorfismos se debe utilizar la utilidad *NCBI Sequence Viewer* dando la posición central y los límites de la subcadena en términos de posiciones de nucleótidos.
7. Para almacenar las secuencias de cada polimorfismo, aparte de hacerlo en la propia base de datos mencionada en el punto anterior, cada secuencia se almacenará en un fichero de texto distinto en formato FASTA.

Las bases de datos de NCBI GenBank que deben explorarse se relacionan a continuación:

1. **NCBI OMIM**: para obtener información de los polimorfismos: identificación de polimorfismos para cada gen en forma de (a) posición relativa en número de nucleótidos desde el Start Codon del gen, o número de contig, más el cambio de nucleótido (ejemplo: -477C/T, +320G/C); (b) posición relativa en número de



aminoácidos en la proteína más en cambio de aminoácido (ejemplo: I523V, C600Y).

2. **NCBI SNP**: para obtener información de los SNP asociados a los polimorfismos: según el identificador del polimorfismo del punto (1), se trata de encontrar el SNP correspondiente.
3. **NCBI Nucleotide**: para obtener información de genes y las subcadenas: información de los genes (Start Codon, Gene ID, posiciones de los contig, etc.) y extracción de las subcadenas a partir de la contig position del SNP. Se ha determinado arbitrariamente que las subcadenas tengan una longitud de 501 nucleótidos: 250 a la izquierda y 250 a la derecha del SNP.
4. **NCBI Protein**: para ver las variaciones de aminoácidos que se asocian a disfunciones fisiológicas.

Los ficheros para almacenar las secuencias de cada polimorfismo deben denominarse según la regla siguiente:

```
"GID" & < GenBank ID > & "_" & <dbSNP id> & "_" & <SNP
contig position> & ".snp"
```

donde:

1. <GenBank ID>: es el GenBank ID de cada gen. Ejemplo: ABCA1 tiene el 29795229, ABCG5 tiene el 29791392, APOA1 tiene el 29805597, etc.
2. &: es el operador de concatenación para formar una cadena con cada parte.
3. "\_": es el literal guión subrayado.
4. <dbSNP id>: es el identificador de cada SNP dado en la SNP database que empieza con el literal "rs" y sigue con un conjunto de cifras. Ejemplo: rs1883024, rs2742458, etc.
5. <SNP contig position>: es la posición del SNP expresado con relación a la posición de contig del gen. Ejemplo: 9388519, 9395124, etc.

Entonces, los nombres de los ficheros pueden ser por ejemplo:

```
GID29795229_rs1883024_9388519.snip
```

```
GID29791392_rs6720173_22856334.snip
```

Se efectuó una investigación preliminar sobre los posibles genes implicados en las enfermedades cardiovasculares, y en los sistemas biológicos como manifestaciones fenotípicas y factores de riesgo (Blaschke et al., 2002). Esta

investigación se basó en la búsqueda en los artículos publicados en las revistas más afines (*New England Journal of Medicine, Circulation, Stroke, etc.*) y en la selección de los genes reportados según un conjunto mínimo de criterios aplicados. Además, se intentó identificar los polimorfismos significativos de estos genes que pudieran ser de interés en el proyecto. En la lista preliminar los genes candidatos se pueden agrupar según aspectos bioquímicos o patológicos que son los siguientes:

- Asociados a lípidos: se encontraron más de 100 genes.
- Hipertrigliceridemia: aspecto patológico y bioquímico caracterizado por tener concentraciones altas de triglicéridos.
- Hiperlipidemia: aspecto patológico y bioquímico caracterizado por tener concentraciones anormales (altas o bajas respectivamente) de colesterol, HDL-C, LDL-C, etc.
- Diabetes: aspecto patológico caracterizado por manifestar resistencia a la insulina o no producirla de forma natural en cantidad suficiente.
- Inflamación: aspecto patológico caracterizado por presentar diversos tipos de inflamaciones o edemas.
- Etc.

Parte del resultado está presentado en la Tabla 3.5.

**Tabla 3.5. Genes y Polimorfismos Identificados. Proceso preliminar (parcial)**

| Gen   | Cromosoma | SNP                | Diana                         | Evidencia  |
|-------|-----------|--------------------|-------------------------------|--|
| APOA1 | 11q21     | -75G/A             | Respuesta a terapia dietética | Determina la respuesta de los niveles de C-HDL a cambios en el consumo de ácidos grasos poliinsaturados.                 |
| APOA4 | 11q21     | G360H<br>Ser347Tre | Respuesta a terapia dietética | Determina la respuesta de C-LDL a cambios en el consumo de colesterol. También afecta el estado de lipemia postprandial. |
| APOC3 | 11q21     | SstI               | Riesgo                        | Niveles de triglicéridos, potencialmente asociado con un aumento del riesgo cardiovascular                               |
| APOE  | 19        | E2/E3/E4           | Riesgo y respuesta            | Riesgo Cardiovascular. Respuesta a terapias (Dieta, estatinas).  |
| APOA5 | 11q21     | 56C/G<br>-1131T/C  | Riesgo                        | Hipertrigliceridemia   |
| LIPC  | 15q21-q23 | C-514T             | Riesgo y respuesta            | Niveles de C-HDL, Riesgo cardiovascular y respuesta a terapia  |

Las fuentes de información realmente disponibles para la ejecución del proyecto fueron las siguientes:

- Resultados experimentales propios del Nutrigenomics Lab.

- Evidencia publicada: NCBI PubMed.
- Conocimiento de expertos. En este caso se trata de los siguientes expertos:
  - Dr. J.M. Ordovás: Nutrigenómica y Lípidos.
  - Dra. D. Corella: Epidemiología Genómica.
  - Dr. L. Parnell: Bioinformática y Biología Molecular.

Los recursos realmente disponibles para la ejecución del proyecto fueron los siguientes:

- BD referencias publicadas: NCBI PubMed, que se encuentra en <http://www.ncbi.nlm.nih.gov/Entrez/>.
- BD enfermedades con origen genético. NCBI OMIM (On Mendelian).
- BD secundarias de enfermedades: GeneCards, que se encuentra en <http://bioinfo.weizmann.ac.il/cards/>.
- BD secuencias de ADN. NCBI Nucleotide, NCBI LocusLink.
- BD SNP: NCBI SNP.
- BD proteínas: NCBI Protein.
- BD integradas de ECV: CPAN. <http://www.cpan.org/>.

Dado que la lista inicial de genes era relativamente grande (mas de 100) y, la de polimorfismos asociados podía llegar a ser el quíntuplo, se determinó trabajar inicialmente con un grupo reducido de ellos, a modo de proyecto piloto, para después ampliar la lista a conveniencia y en función de la obtención y publicación de nuevas evidencias.

### 3.4.3 El marco de auditoría

La *Information Systems Audit and Control Association* (ISACA), es una asociación profesional de ámbito mundial, con más de 40.000 asociados individuales e institucionales, para la regulación de la práctica de la Auditoría de Sistemas de Información. La sede principal esta en EE. UU. (ISACA, 2004). En algunos países la ISACA tiene reconocimiento oficial como la institución reguladora del ejercicio de sus auditores informáticos, con unas competencias similares a los organismos que controlan la actividad de auditoría de cuentas. En otros países, como España, el reconocimiento es tácito ya que no existen normas legales específicas.

Como resultado de un proceso relativamente largo, donde los miembros de la ISACA han ido aportando sus ideas, sugerencias y experiencia, se publicó en 1997 la primera versión de un marco metodológico formal denominado COBIT (*Control Objectives for Information and related Technology* - Objetivos de Control para la Información y Tecnologías Afines) (ISACAF-B, 2000). COBIT, por tanto, está ampliamente aceptado por la comunidad internacional de auditores de sistemas de información como una norma estándar.

#### 3.4.3.1 Objetivos de COBIT

La Misión de COBIT es la siguiente: “Investigar, desarrollar, publicar y promover un conjunto de objetivos de controlen tecnología de información con autoridad, actualizados, de carácter internacional y aceptados generalmente para el uso cotidiano de gerentes de empresas y auditores.” (Bernal et al., 1999).

COBIT está diseñado como un estándar aplicable y aceptable en general para la buena práctica de la auditoría de las tecnologías de la Información en todo el mundo. El producto COBIT utiliza los Objetivos de Control de ISACA, mejorados con estándares específicos de tipo técnico, profesional, normativo e industrial existentes y emergentes. Los objetivos de control se han desarrollado para su aplicación en el amplio espectro de sistemas de información en la empresa. Estos objetivos de control tienen en cuenta lo siguiente (ISACAF-B, 2000):

- Adecuación a los estándares y normativas legislativos y de hecho existentes que se aplican en el marco global, así como en los objetivos de control individuales.
- Revisión crítica de las diferentes actividades y tareas bajo los dominios de control y posibilitando la especificación de indicadores de prestaciones importantes (normas, reglas, etc.)
- Establecimiento de unas directrices y fundamentos para proporcionar investigación consistente sobre los temas de auditoría y control de Tecnologías de Información (TI).

COBIT se ha diseñado como sistema metodológico que consiste en un conjunto de objetivos de control de TI de alto nivel y una estructura global para su clasificación y funcionamiento(Figura 3.8).

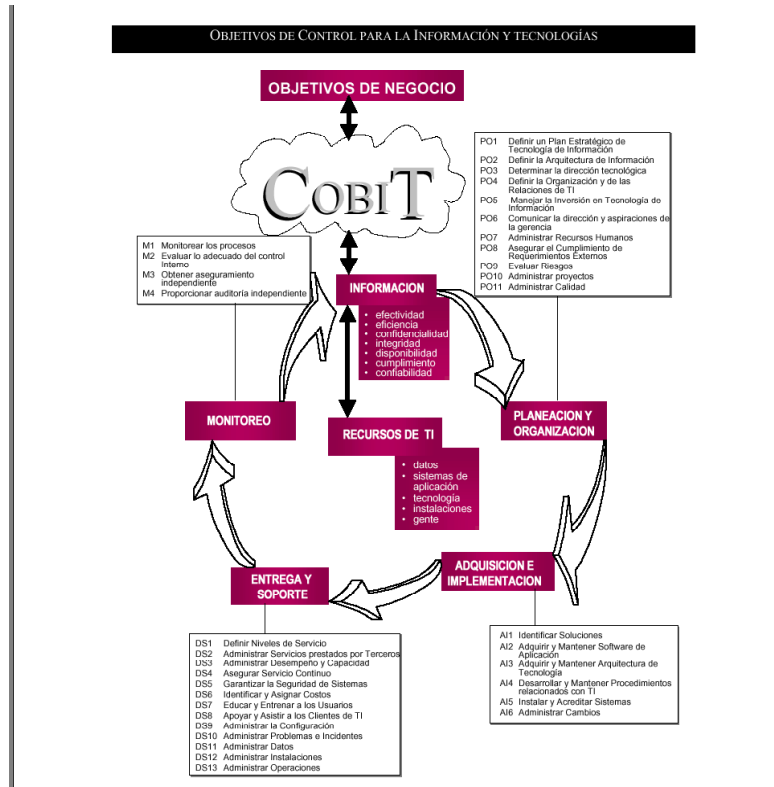


Figura 3.8. Recursos de TI, Objetivos de Negocio y Dominios de COBIT “Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute.”

### 3.4.3.2 Estructura y objetivos de control

La teoría subyacente para la clasificación elegida, en línea con las experiencias de Re-Ingeniería, es que hay, en esencia, tres niveles de esfuerzos en TI cuando se considera la gestión de los recursos de TI (ISACAF-B, 2000):

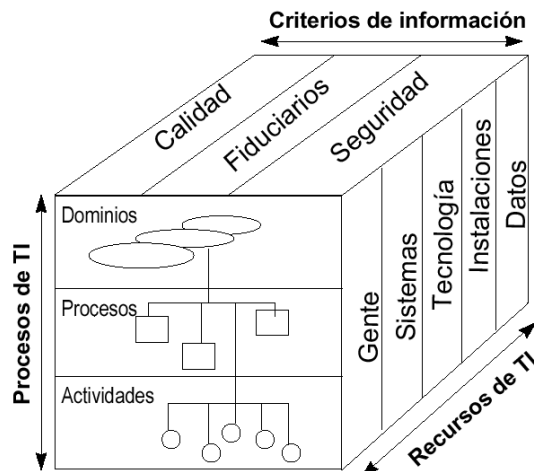
1. Actividades: Las actividades, junto con las tareas están en el nivel inferior. Las actividades tienen el concepto de ciclo de vida mientras que las tareas se consideran discretas en el tiempo.
2. Procesos: Se definen en un nivel superior como series de actividades unidas con puntos de control naturales.

3. Dominios: Correspondientes al nivel superior, son agrupaciones de procesos. COBIT distingue cuatro dominios en línea con el ciclo de gestión o el ciclo de vida aplicables a los procesos de TI (Tabla 3.6).

**Tabla 3.6. Dominios de COBIT**

|  |   |
|--|---|
| <p><b>Planificación y Organización:</b><br/>Conduce la estrategia y las tácticas y corresponde a la identificación de la forma en que la información tecnológica puede contribuir mejor a alcanzar los objetivos de gestión.</p> | <p><b>Distribución y Soporte:</b><br/>Corresponde con la distribución normal de los servicios requeridos, que van desde las tradicionales operaciones sobre seguridad y continuidad hasta la formación.</p> |
| <p><b>Adquisición e Implementación:</b><br/>Para llevar a cabo la estrategia es necesario identificar, desarrollar y adquirir soluciones de TI apropiadas, así como implementarlas e integrarlas en los procesos de gestión.</p> | <p><b>Monitorización:</b><br/>Todos los procesos de TI deben evaluarse regularmente en el tiempo para comprobar su calidad.</p>   |

El marco conceptual se enfoca desde tres puntos de vista distintos: criterios de gestión para la información, recursos de TI y procesos de TI. Estos tres puntos de vista se ensamblan en un formato cúbico y permiten que se obtengan referencias cruzadas en dicho marco y se pueda acceder a él eficientemente (Figura 3.9).



*Figura 3.9. El cubo de COBIT. "Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute."*

Los objetivos de control de TI están organizados inicialmente por proceso / actividad, pero las ayudas para la navegación que se aportan, facilitan la entrada desde cualquier punto estratégico. También facilitan la adopción de enfoques combinados o globales, tal como la instalación / implementación de un proceso, responsabilidades de gestión global para un proceso, y el uso de los recursos de TI por un proceso (Figura 3.10).

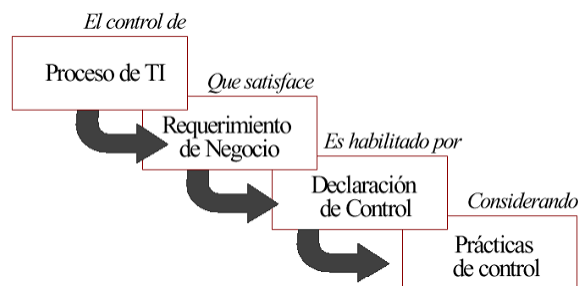


Figura 3.10. Objetivos de control de COBIT definidos genéricamente. "Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute."

La información que los procesos de gestión necesitan está proporcionada por el uso de los recursos de TI. Para asegurar que los requisitos de gestión para la información se aplican, se tiene que definir medidas de control adecuadas, se tiene que implementarlas y monitorizarlas sobre estos recursos. Está claro que no todas las medidas de control satisfarán los requisitos de gestión en el mismo grado, así que se hace una distinción en COBIT contemplando el cumplimiento (Figura 3.11):

- **Primario (P):** grado en que el objetivo de control satisface completamente el requisito de información correspondiente.
- **Secundario (S):** grado en que el objetivo de control satisface solamente en menor extensión o indirectamente el requisito de información correspondiente.

Se ha descrito el COBIT tal como está formulado en la actualidad para la ASI, que no es adecuado para realizar auditorías bioinformáticas. Por lo tanto, es necesario estudiar cómo se puede adaptar para poder obtener un marco metodológico específico para la auditoría bioinformática.

TABLA RESUMEN

| DOMINIO                   | PROCESO | Criterios de Información                          |            |                  |            |                |             |               |              | Recursos de TI         |            |               |       |  |  |  |  |
|---------------------------|---------|---|------------|------------------|------------|----------------|-------------|---------------|--------------|------------------------|------------|---------------|-------|--|--|--|--|
|                           |         | efectividad                                       | eficiencia | confidencialidad | integridad | disponibilidad | completitud | confiabilidad | portabilidad | sistemas de aplicación | tecnología | instalaciones | datos |  |  |  |  |
| Planeación y Organización | PO1     | Definir un plan estratégico de sistemas           | P          | S                |            |                |             |               |              |                        |            |               |       |  |  |  |  |
|                           | PO2     | Definir la arquitectura de información            | P          | S                | S          | S              |             |               |              |                        |            |               |       |  |  |  |  |
|                           | PO3     | Determinar la dirección tecnológica               | P          | S                |            |                |             |               |              |                        |            |               |       |  |  |  |  |
|                           | PO4     | Definir la organización y sus relaciones          | P          | S                |            |                |             |               |              |                        |            |               |       |  |  |  |  |
|                           | PO5     | Administrar las inversiones (en TI)               | P          | P                |            |                |             |               |              | S                      |            |               |       |  |  |  |  |
|                           | PO6     | Comunicar la dirección y objetivos de la gerencia | P          | P                |            |                |             | S             |              |                        |            |               |       |  |  |  |  |
|                           | PO7     | Administrar los recursos humanos                  | P          | P                |            |                |             |               |              |                        |            |               |       |  |  |  |  |
|                           | PO8     | Asegurar el apego a disposiciones externas        | P          | P                |            |                |             |               |              | P                      | S          |               |       |  |  |  |  |
|                           | PO9     | Evaluar riesgos                                   | S          | S                | P          | P              | P           | S             | S            |                        |            |               |       |  |  |  |  |
|                           | PO10    | Administrar proyectos                             | P          | P                |            |                |             |               |              |                        |            |               |       |  |  |  |  |
|                           | PO11    | Administrar calidad                               | P          | P                |            | P              |             |               | S            |                        |            |               |       |  |  |  |  |

Figura 3.11. Tabla resumen de COBIT. “Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute.”

Se plantea a continuación un modelo de desarrollo de auditoría para la Función Bioinformática basado en el COBIT (ISACAF-B, 2000) (ISACAF-E, 2000) que está ampliamente descrito en (Coltell et al., 2002). Dicho modelo consiste en la definición de las subfunciones principales de la Bioinformática: Investigación Teórica, Gestión de la Investigación, Aplicación de la Investigación, Desarrollo Tecnológico, Formación y Seguridad y Protección de la Información.

**3.4.3.3 Descripción y estructura organizativa del objeto de estudio**

El *Nutrition and Genomics Laboratory*, perteneciente al *JM-USDA Human Nutrition Research Center on Aging at Tufts University*, Boston, MA, EE. UU., es un laboratorio de investigación creado en 2000 por su responsable, el Dr. José M. Ordovás, que surge del más antiguo *Lipid Metabolism Laboratory*, encuadrado en el mismo centro. El *Human Nutrition Research Center* es un instituto de investigación cofinanciado por el Departamento de Agricultura de los EE. UU. (USDA), y la Universidad Tufts.

Las actividades de investigación realizadas por este laboratorio combinan la alta investigación en Epidemiología Genética, Nutrición Genómica (Nutrigenómica), Metabolómica y Transcriptómica. En la Epidemiología Genética se trabaja sobre las asociaciones genotipo-fenotipo y las interacciones gen-dieta. En la Nutrigenómica se trabaja sobre estudios de intervención dietaria. En la Metabolómica se trabaja sobre los biomarcadores no invasivos del envejecimiento saludable. Y finalmente, en la Transcriptómica se trabaja sobre los mecanismos de la respuesta dietaria y nuevos genes relacionados con los procesos de



envejecimiento. La estructura orgánica estable del laboratorio se muestra en la Tabla 3.7. Por cuestiones de confidencialidad, no se incluyen los nombres propios de los miembros del laboratorio excepto la de su responsable.

**Tabla 3.7. Estructura orgánica estable del Nutrition and Genomics Laboratory**

| Miembro       | Perfil                   | Asignación                              | Puesto                 |
|---------------|--------------------------|---|------------------------|
| J. M. Ordovás | Director                 | Despacho de investigación               | Staff                  |
| Miembro N° 02 | Genetista                | Despacho de investigación y laboratorio | Staff                  |
| Miembro N° 03 | Técnico de laboratorio   | Laboratorio                             | Staff                  |
| Miembro N° 04 | Técnico de laboratorio   | Laboratorio                             | Staff                  |
| Miembro N° 05 | Técnico de investigación | Despacho de investigación y laboratorio | Estudiante predoctoral |
| Miembro N° 06 | Técnico de investigación | Despacho de investigación y laboratorio | Estudiante predoctoral |
| Miembro N° 07 | Técnico de investigación | Despacho de investigación y laboratorio | Estudiante             |
| Miembro N° 08 | Graduate Student         | Despacho de investigación y laboratorio | Estudiante             |
| Miembro N° 09 | Graduate Student         | Despacho de investigación y laboratorio | Estudiante             |

Además, se produce durante el año una rotación de investigadores procedentes del extranjero, en calidad de científicos visitantes o de becarios posdoctorales. Esta afluencia de visitas ha dado lugar a que el laboratorio mantenga una gran cantidad de colaboraciones con los investigadores que han estado allí y con sus respectivos centros. En la Figura 3.12 se muestra la arquitectura funcional y de responsabilidades del laboratorio, donde se destacan los instrumentos principales.

Como se puede observar, la dirección y supervisión de los técnicos de laboratorio, los estudiantes y los científicos visitantes, está compartido entre el director del laboratorio y el genetista. Por otra parte, se muestra una Función Bioinformática que tiene a su cargo el control de la instrumentación principal y que también depende del director y del genetista. Esta función no está cubierta por ningún personal en la actualidad. El laboratorio como unidad orgánica dispone de dos laboratorios de investigación biomédica y molecular que contienen esta instrumentación principal y otros instrumentos y elementos auxiliares habituales en este tipo de laboratorios. En la actualidad tampoco no existe personal auxiliar asignado para desempeñar funciones de administración general y económica de las operaciones diarias del laboratorio.

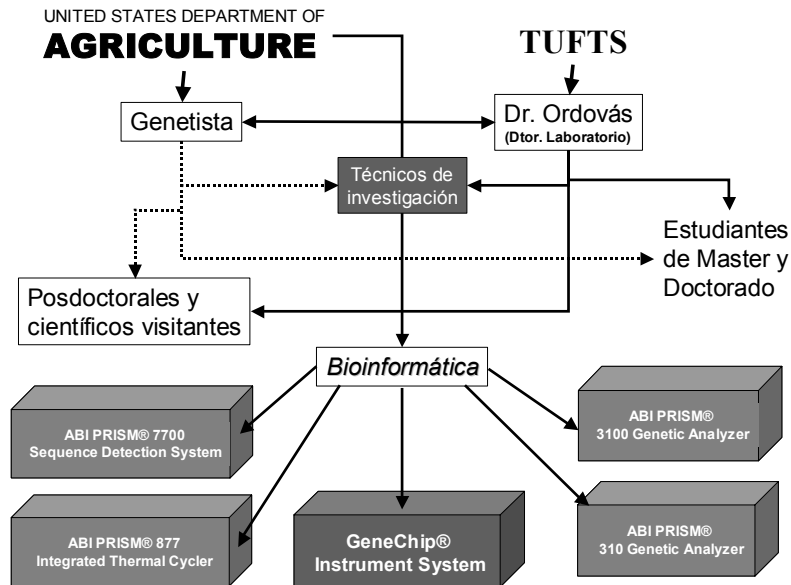


Figura 3.12. Arquitectura funcional y de responsabilidades del laboratorio

La estructura de sistemas informáticos del laboratorio se muestra en la Tabla 3.8. Cabe destacar que casi todos los integrantes del mismo tienen un equipo informático asignado para llevar a cabo trabajos generales, redacción de documentos científicos, búsqueda de bibliografía por Internet e intranet, comunicación a través de correo electrónico y otras tareas ofimáticas auxiliares. La excepción son los dos técnicos de laboratorio, que actualmente comparten el mismo equipo, puesto que uno de ellos se ha incorporado recientemente, y un *graduate student* que solamente utiliza uno de los laboratorios para los experimentos que determinan los supervisores.

El laboratorio no dispone de personal técnico informático propio sino que está asistido por personal del Departamento de Computación del centro. Dicho departamento dispone de dos técnicos para dar servicio a un edificio de catorce plantas, incluyendo el parque de ordenadores y la red local del edificio. No se ha facilitado datos sobre el parque de ordenadores del centro.

Por lo tanto, una vez perfilada la propuesta metodológica para la realización de auditorías sobre las funciones bioinformáticas, se planteó el siguiente paso: la aplicación práctica de la misma a un caso concreto y al alcance del doctorando, que es el laboratorio descrito.

**Tabla 3.8. Estructura de sistemas informáticos del Nutrition and Genomics Laboratory**

| Equipo                 | Sistema Operativo | Lugar de asignación                                    | Responsable              |
|------------------------|-------------------|--|--------------------------|
| PC, DELL, P-III        | Windows 2000 Pro. | Despacho de investigación director                     | J. M. Ordovás            |
| PC, Gateway, P-II      | Windows 2000 Pro. | Despacho de investigación director                     | J. M. Ordovás            |
| Laptop, Sony, P-III    | Windows XP HE     | Despacho de investigación y exterior                   | J. M. Ordovás            |
| Laptop, Toshiba, P-III | Windows 2000 Pro. | Despacho de investigación staff y exterior             | Genetista                |
| PC, Apple, Imac        | Apple MacOS 9.0   | Laboratorio  | Técnico de laboratorio   |
| PC, DELL, P-III        | Windows 2000 Pro. | Despacho de investigación staff                        | Técnico de investigación |
| PC, DELL, P-III        | Windows 2000 Pro. | Despacho de investigación estudiantes                  | Técnico de investigación |
| PC, DELL, P-III        | Windows 2000 Pro. | Despacho de investigación estudiantes                  | Técnico de investigación |
| PC, Gateway, P-II      | Windows 98 SE     | Despacho de investigación estudiantes                  | Graduate Student         |
| PC, DELL, P-III        | Windows NT 4.0.   | Laboratorio. ABI PRISM® 7700 Sequence Detection System | Técnico de laboratorio   |
| PC, DELL, P-II         | Windows NT 4.0.   | Laboratorio. ABI PRISM® 3100 Genetic Analyzer          | Técnico de laboratorio   |
| PC, DELL, P-II         | Windows 98        | Laboratorio. ABI PRISM® 310 Genetic Analyzer           | Técnico de laboratorio   |
| PC, Apple              | Apple MacOS 8.0   | Laboratorio. ABI PRISM® 877 Integrated Thermal Cyclers | Técnico de laboratorio   |
| HP 4500 Laserjet       | Postscript        | Salita común   | Técnico de laboratorio   |

### 3.5 Bibliografía

#### 3.5.1 Bibliografía de técnicas de determinación

Kleinbaum, D.G., Kupper, L.L. and Muller, K.E. (1988) Applied regression analysis and other multivariable methods. 2<sup>a</sup> ed. Boston: PWS Kente Publishing company.

- Rothman, K. (1986). *Modern Epidemiology*, eds. Little Brown and Company. Boston.
- Tilzer, L.L., Thomas, S.M. and Moreno, R.F. (1989). Use of silica gel polymer for DNA extraction with organic solvents. *Anal Biochem* 183:13-15.
- Wayne, S.D. (1990). *Bioestadística*. 3ª ed. Mexico D.F. Limusa.

### 3.5.2 Bibliografía sobre técnicas de análisis de ADN y análisis de datos

- Arlequin home page. A software for population genetics data analysis.  
<http://amthro.unige.ch/arlequin>. Accedido el 4 de junio de 2004.
- Rockefeller University. Haplotype analysis programs home page.  
<http://linkage.rockefeller.edu/ott>. Accedido el 4 de junio de 2004.
- Rockefeller University. Linkage programs home page.  
<http://linkage.rockefeller.edu/ott/linkutil.htm>. Accedido el 4 de junio de 2004.
- SAS statistics home page.  
<http://www.sas.com/technologies/analytics/statistics/stat/index.html>. Accedido el 4 de junio de 2004.
- SPSS home page. <http://www.spss.com>. Accedido el 4 de junio de 2004.

### 3.5.3 Bibliografía sobre técnicas de desarrollo de software

- ANSI/IEEE Std. 830-1993. «Recommended Practice for Software Requirements», IEEE, 1993.
- Boehm B. *Software Engineering Economics*, Prentice-Hall, 1981.
- Booch G., Rumbaugh J., Jacobson I. *El Lenguaje Unificado de Modelado*, Addison-Wesley, Madrid, 1999.
- Bruegge B., Dutoit A.H. *Ingeniería de Software Orientado a Objetos*. Prentice Hall – Pearson educación, México, 2002.
- Budd T. *Introducción a la programación orientada a objetos*, Addison-Wesley Iberoamericana, Wilmington, DW (USA), 1994.
- Bourque, Pierre; Dupuis, Robert; Abran, Alain; Moore, James W.; Tripp, Leonard; y Wolf, Sybille. Fundamental principles of software engineering – a journey. *The Journal of Systems and Software* 2002; 62, 59-70.
- Castagna G. *Object-Oriented Programming. A Unified Foundation*. Birkhäuser, Boston, MA (USA), 1996.

- Cleland D.I. *Project Mangement. Strategic Desing and Implementation*, TAB Professional and Reference Books, Blue Ridge Summit, PA (USA), 1990.
- Coltell O. "Ingeniería del Software. Introducción. Ciclos de vida, herramientas CASE. Generación automática de código. Reusabilidad de componentes". García F. J., del Valle F. J. (Eds.) Temario para Oposiciones a Analista de Sistemas de Información. Volumen I. Colegio Oficial de Ingenieros en Informática de la Comunidad Valenciana, Valencia, 2003: 657-707. ISBN 84-607-9851-8, 84-607-9852-6 (Vol. I).
- Harmon P., Hall C. *Intelligent Software Systems Development. An IS Managers Guide*. John Wiley, New York-USA, 1993.
- Jacobson I., Booch G., Rumbaugh J. *El Proceso Unificado de Desarrollo de Software*. Addison-Wesley, Madrid, 2000.
- Pressman R.S. *Ingeniería del Software. Un enfoque práctico* (5ª ed.) Mc Graw-Hill; New York, 2001.
- Rumbaugh J., Jacobson I., Booch G. *El Lenguaje Unificado de Modelado. Manual de Referencia*. Addison-Wesley, Madrid, 2000.
- Sommerville I. *Ingeniería de software*. 6ª edición. PrenticeHall – Pearson educación, México, 2002.
- OMG home page. <http://www.omg.org>. Accedido el 4 de junio de 2004.
- UML home page. <http://www.uml.org>. Accedido el 4 de junio de 2004.
- ISO home page. <http://www.iso.com>. Accedido el 4 de junio de 2004.

### 3.5.4 Bibliografía CETP

- Tenkanen H, Koskinen P, Kontula K, Aalto\_Setälä K, Mänttari M, Manninen V, et al. Polymorphisms of the gene encoding cholesterol ester transfer protein and serum lipoprotein levels in subjects with and without coronary heart disease. *Hum Genet* 1991; 87:574-578.
- Manninen V, Elo MO, Frick MH, Haapa K, Heinonen OP, Heinsalmi P, et al. Lipid alterations and decline in the incidence of coronary heart disease in the Helsinki Heart Study. *JAMA* 1988; 260:641-651.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol* 1979; 110:281-290.
- Dawber TR, Meadors GF, Moore R. Epidemiological approaches to heart disease: The Framingham Study. *Am J Public Health* 1951; 41:279-286.
- Cupples LA, Gagnon DR, Kannel WB. Long- and short-term risk of sudden coronary death. *Circulation* 1992; 85:111-118.

- Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg<sup>2+</sup> precipitation procedure for quantitation of high-density-lipoprotein cholesterol. *Clin Chem* 1982;28:1379-88.
- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972; 18:499-502.
- McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and apolipoprotein fractions. *Clin Chim Acta* 1987; 166:1-9.
- Schaefer EJ, Ordovas JM. Metabolism of the apolipoproteins A-I, A-II, and A-IV. In: Segrest J, Albers J, editors. *Methods in Enzymology, Plasma Lipoproteins, Part B: Characterization, Cell Biology and Metabolism*. Academic Press, 1986:420-442.
- Ordovas JM, Peterson JP, Santaniello P, Cohn J, Wilson PWF, Schaefer EJ. Enzyme linked immunosorbent assay for human apolipoprotein B. *J Lipid Res* 1987; 28:1216
- Boerwinkle E, Utermann G. Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am J Hum Genet* 1988; 42:104-112.
- Almasy L, Blangero J. Multipoint quantitative trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; 62:1198-1211.

### 3.5.5 Bibliografía APOE

- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972;18:499-502.
- Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res* 1990;31:545-8.

### 3.5.6 Bibliografía LIPC

- Otvos JD, Jeyarajah EJ, Bennett DW, et al. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. *Clin Chem* 1992;38:1632-1638.

### 3.5.7 Bibliografía SR-BI

- Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. *Arterioscler Thromb Vasc Biol* 2000; 20:1323-1329.
- Osgood-McWeeney D, Galluzzi JR, Ordovas JM. Allelic discrimination for single nucleotide polymorphisms in the human scavenger receptor class B type 1 gene locus using fluorescent probes. *Clin Chem* 2000; 46:118-119.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score Tests for Association between Traits and Haplotypes when Linkage Phase Is Ambiguous. *Am. J. Hum. Genet.* 2002; 70:425-434.
- Lewontin RC. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 1964; 49:49-67.

### 3.5.8 Bibliografía PLIN

- Antonarakis SE. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat* 1998; 11:1-3.

### 3.5.9 Bibliografía de auditoría de sistemas de información en laboratorios de investigación en Biomedicina

- Benal R., Coltell O. Auditoría de los Sistemas de Información (reimpresión). Servicio de Publicaciones de la Universidad Politécnica de Valencia, Valencia, 1999.
- Cass S., Riezenmann M.J. "Improving Security, Preserving Privacy". *IEEE Spectrum*, Jan.; 2002: 44-49.
- Coltell O., Chalmeta R. "Auditoría Bioinformática". Actas del V Congreso Nacional de Informática y Salud, INFORSALUD 2002. Madrid, 2002.
- Cornell M., Paton N.W., Wu S., Goble C.A., Miller C.J., Kirby P., Eilbeck K., Brass A., Hayes A., Oliver S.G. "GIMS-A Data Warehouse for Storage and Analysis of Genome Sequence and Functional Data". EBI, the European Bioinformatics Institute (EMBL Outstation, Hinxton, UK), <http://www.ebi.ac.uk/> (accedido 6/02/2002).
- Ermolaeva O., Rastogi M., Pret K.D., Schuler G.D., Bittner M.L., Chen Y., Simon R., Meltzer P., Trent J.M., Boguski M.S. "Data management and analysis for gene expression arrays". *Nature genetics*, 20; 1998: 19-23.

- Hass L.M., Kodali P., Rice J.E. "Integrating Life Sciences Data – With a Little Garlic". Proc. IEEE Int. Symp. on Bio-Informatics and Biomedical Engineering. IEEE; 2000: 5-12.
- ISACA home page. Information Systems Audit and Control Association. <http://www.isaca.org>. Accedido el 4 de junio de 2004.
- ISACAF-B. COBIT. Framework. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- ISACAF-D. COBIT. Control Objectives. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- ISACAF-E. COBIT. Audit Guidelines. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- ISACAF-F. COBIT. Implementation Tool Set. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- Piattini M., Del Peso E. (eds.) Auditoría Informática. Un enfoque práctico. Ra-Ma, Madrid, 1998.
- Rindfleisch T.C. "Privacy, Information Technology, and Health Care". Communications of the ACM, 40-8; 1997: 93-100.
- Rondel R. K., Varley S. A., Webb C. (eds.) Clinical Data Management. John Wiley, New York, 1993.
- Sackman H. Biomedical Information Technology. Global Social Responsibilities for the Democratic Age. Academic Press, San Diego, CA (USA), 1997.
- Science's News staff. "A History of the Human Genome Project". Science Magazine; 291 (5507); 2001: 1195-1261.
- Shrothliffe E.H., Blois M.S. "The Computer Meets Medicine and Biology: Emergence of a Discipline". Computer Applications in Health Care and Biomedicine; <http://smi-web.stanford.edu/textbook/ChapterOne.htm> (accedido el 22/01/2002).
- Van Bommel J.H., Musen M.A.(eds.) Handbook of Medical Informatics. Springer-Verlag, Heidelberg, 1997.





Las ciencias tienen las raíces amargas, pero muy dulces los frutos.

Aristóteles

#### 4.1 Introducción

Los resultados presentados en esta tesis son fruto de un trabajo multidisciplinar que implica una labor en equipo para la obtención de las conclusiones generales. A lo largo de los siete años de investigación recopilando la información necesaria para llevar a cabo los trabajos conducentes a esta tesis, el doctorando ha tenido la oportunidad de participar en todas y cada una de las tareas que forman parte del protocolo de la investigación genómica cardiovascular. Así, por decisión propia, el doctorando se ha familiarizado con las técnicas de laboratorio incluyendo, desde la extracción de ADN, hasta el análisis de polimorfismos por distintas técnicas.

Si bien, para la optimización del proceso ha sido necesaria su especialización en aquellas tareas que demandaban conocimientos más profundos en Ciencias de la Computación y Estadística, incluyendo diseño, desarrollo y gestión de proyectos bioinformáticos, desarrollo de software, gestión de datos, control de calidad de procesos y datos, diseño de guías de auditoría y aplicación de las mismas, elaboración de recomendaciones y herramientas para la mejora de procesos en el laboratorio genómico, acceso a bases de datos genómicas, tratamiento y análisis estadístico de datos, y participación en la presentación de los resultados y redacción de los manuscritos o comunicaciones a congresos.

Seguidamente se detallan los resultados obtenidos en los trabajos de investigación realizados en el marco de esta tesis, así como las propuestas y herramientas elaboradas en función de las demandas bioinformáticas surgidas en la investigación genómica cardiovascular.

## 4.2 Resultados de los proyectos bioinformáticos

En esta sección se describen los resultados obtenidos en los trabajos de desarrollo de soluciones bioinformáticas a los problemas planteados en la Sección 1.6 del Capítulo 1, cumpliendo los objetivos generales y particulares que se han establecido en el Capítulo 2. En particular, los proyectos desarrollados se relacionan con los problemas de comunicación de datos entre instrumentos de laboratorio, visualización y manejo de secuencias de ADN, visualización y compresión de secuencias de ADN y búsqueda y detección de polimorfismos.

### 4.2.1 Soporte de comunicación instrumental (PLATEX)

Estos resultados se han publicado en la revista *Clinica e Investigación en Arteriosclerosis* en el año 2004 con el título “PLATEX: Una herramienta bioinformática para la conversión de datos en el estudio genético de la arteriosclerosis” (Coltell et al., 2004). A continuación se presenta un resumen de los principales hallazgos.

En esta Subsección se describe uno de los resultados de los trabajos de investigación, que ha consistido en la resolución de la falta de comunicación en línea entre ficheros que contenían los códigos y las posiciones de placas de análisis de secuenciación de 96 ó 384 pocillos, y el software para la colección de datos del Analizador Genético ABI PRISM® modelo 3100 (Applied Biosystems, 2001). Este instrumento está siendo utilizado en el *Nutrigenomics Laboratory* del *Human Nutrition Research Center on Aging at Tufts University* (Boston).

#### 4.4.1. El Gestor de Conversión de Datos Automático

La solución aplicada ha sido un conjunto de macros adjuntas a un libro Microsoft® Excel™ que automatizan completamente el proceso descrito en la sección anterior. Este conjunto, denominado PLATEX, realiza entre otras algunas operaciones importantes. En primer lugar, hay macros que gestionan los ficheros de texto con las muestras de electroforesis. En segundo lugar, otras macros gestionan las etiquetas y los datos de las muestras, mediante hojas de cálculo temporales y específicas, para asignar cada código de muestra al pocillo y color Dye correspondiente (hasta cinco colores distintos) y guardar los datos actualizados en un formato especial. Este formato lo puede importar directamente el software de captación de datos.

El sistema PLATEX tiene un proceso fundamental encargado de la conversión de ficheros que contienen datos que identifican muestras en placas de 96 o 384

pocillos (que se denominan ficheros de rejilla de placa), en ficheros de texto delimitados por tabuladores. Estos últimos pueden ser importados como registros de placa por el software de captación de datos (*Data Collection System*) (Figura 4.1).

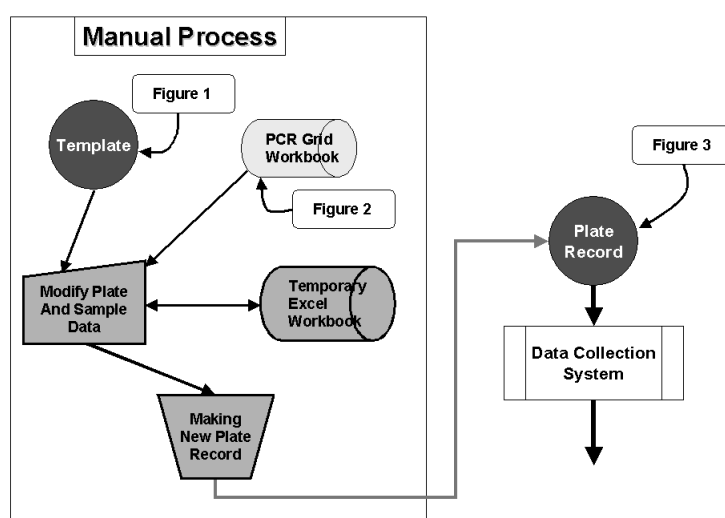


Figura 4.1. El proceso de conversión manual que produce los ficheros de datos (*plate record*) para la secuenciación y análisis de ADN

El proceso completo que se muestra en la Figura 4.1 puede ser cubierto por un conjunto de macros adjunto a un libro Excel<sup>TM</sup>, que es el Gestor del Proceso de Conversión de Datos. Este libro está compuesto por varias hojas de cálculo. Una de ellas sirve como área de trabajo y ejecuta las operaciones básicas sobre los ficheros (creación, apertura, clausura, etc.), y el contenido de las celdas (borrado, copia y pegado). Este proceso automático simplifica un poco la estructura de ficheros en comparación con el proceso manual, tal y como se muestra en la Figura 4.2.

El Gestor tiene una interfaz simple para la ejecución individual de las macros, para realizar alguna tarea atómica del proceso, o la ejecución conjunta para hacer todas las tareas como si fueran transacciones. La construcción de macros individuales para ejecutar tareas atómicas ayuda a la depuración del código VBasic<sup>TM</sup> de cada uno de los módulos. Además, permite crear una arquitectura de componentes donde cada componente se puede modificar o sustituir sin provocar efectos colaterales sobre el resto de componentes (Figura 4.3).

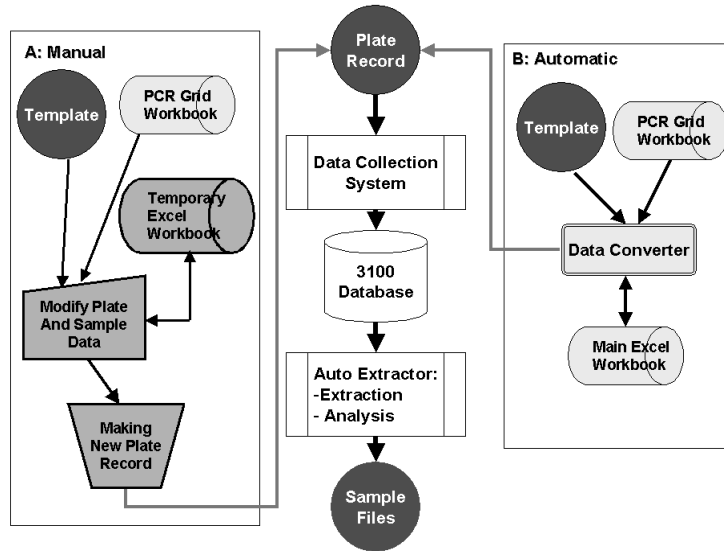


Figura 4.2. El proceso de conversión que obtiene los ficheros de datos para la secuenciación y análisis de ADN. Parte A: proceso manual. Parte B: proceso automático

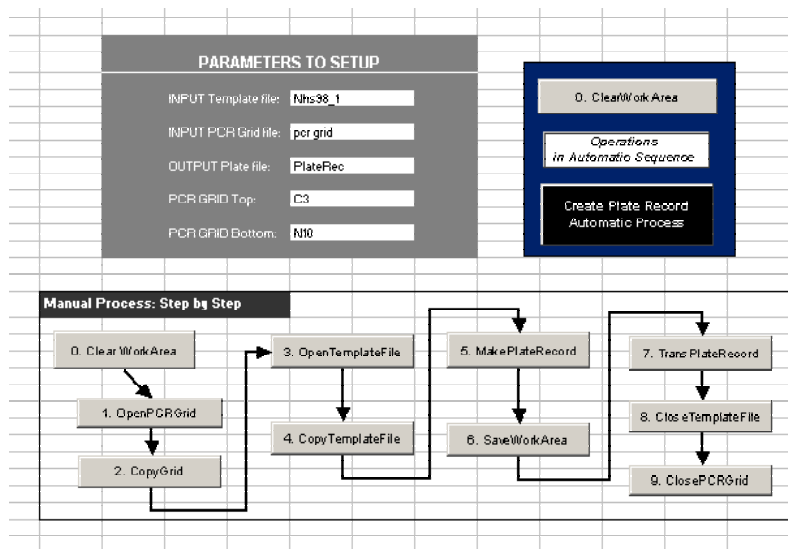


Figura 4.3. Interfaz de la hoja de cálculo que actúa de área de trabajo con un ejemplo de la tarea fundamental en que se asigna etiquetas de muestra a los conjuntos de muestras. Algunos parámetros se pueden modificar para permitir el trabajo con distintos ficheros y placas

#### 4.4.2. Descripción de las macros

Las macros se agrupan en las siguientes categorías según la naturaleza de las operaciones que llevan a cabo: metamacros, macros de ficheros, macros de datos y macros de procesos. Las metamacros son macros que solamente ejecutan otras macros. Las macros de ficheros llevan a cabo operaciones básicas sobre ficheros (creación, apertura, clausura, etc.). Las macros de datos se encargan de las operaciones básicas con los datos (borrado, copia y pegado). Y las macros de procesos ejecutan las tareas principales. En la Tabla 4.1 se puede ver el nombre y la descripción de cada macro agrupada en su categoría correspondiente (se ha mantenido el idioma inglés puesto que esta Tabla se muestra así en la documentación original de la solución).

Los nombres de los ficheros de plantilla, rejilla PCR y registro de placa, y las referencias izquierda superior y derecha inferior de la estructura de datos de la placa del fichero de rejilla PCR, son parámetros que los puede modificar el usuario para manejar otros conjuntos de datos. Para ello se muestran en las celdas correspondientes del Área de Trabajo (Figura 4.4). El poder fijar las referencias de celda de la rejilla de PCR previene los errores debidos a la distribución irregular de las estructuras de datos de placa, ya que es el propio usuario quien detecta la zona de trabajo correcta (Figura 4.4).

|    | A | B     | C      | D     | E     | F     | G     | H     | I     | J     | K     | L     | M     | N |
|----|---|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|
| 2  |   | 1     | 2      | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |   |
| 3  | A | F0385 | F03733 | F0390 | F0395 | F0331 | F0391 | F0353 | T0007 | F0363 | F0371 | T0095 |       |   |
| 4  | B | F0394 | F0378  | F0391 | F0374 | F0380 | F0372 | F0391 | F0323 | F0373 | F0378 | F0381 | F0383 |   |
| 5  | C | F0323 | F0616  | F0395 | F0384 | F0383 | F0373 | F0385 |       | F0371 | F0363 | F0362 | F0327 |   |
| 6  | D | F0393 | F0395  | F0373 | F0383 | F0383 | F0372 | F0395 | F0324 | F0461 | F0363 | F0363 | F0332 |   |
| 7  | E | F0383 | F0370  | F0363 | F0337 | F0381 | F0363 | F0385 | F0397 | F0463 | F0340 | F0464 | F0453 |   |
| 8  | F | F0397 | F0395  | F0463 | F0387 |       | F0325 | F0463 | F0454 | F0391 | F0381 | F0387 | F0385 |   |
| 9  | G | F0385 | F0345  | F0335 | F0338 | F0390 | F0334 | F0333 | F0331 | F0335 | F0390 | F0391 | F0336 |   |
| 10 | H | F0411 | F0393  | F0380 | F0453 | F0472 | F0391 | F0392 | F0470 | F0321 | F0320 | F0399 |       |   |
| 11 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 12 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 13 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 14 |   | 1     | 2      | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |   |
| 15 | A | F0314 | F0382  | F0377 | F0380 | F0326 | F0434 |       | F0433 | F0393 | F0468 | F0423 | F0388 |   |
| 16 | B | F0374 | F0348  | F0339 | F0324 | F0438 | F0359 | F0321 | F0388 | F0393 | F0393 | F0366 | F0318 |   |
| 17 | C | F0370 | F0491  | F0341 | F0389 | F0440 | F0458 | F0435 | F0343 | F0370 | F0393 | F0392 | F0395 |   |
| 18 | D | F0320 |        | F0319 | F0383 | F0324 | F0394 | F0397 | F0393 | F0393 | F0426 | F0389 | F0390 |   |
| 19 | E | F0410 | F0395  | F0395 | F0388 | F0383 | F0380 | F0450 | F0378 | F0463 | F0332 | F0491 | F0310 |   |
| 20 | F | F0391 | F0438  | F0388 | F0373 | F0390 | F0381 | F0377 | F0388 | F0460 | F0364 | F0382 | F0397 |   |
| 21 | G | F0393 | F0377  | F0376 | F0471 | F0493 | F0463 | F0469 | F0436 | F0398 | F0473 | F0467 | F0382 |   |
| 22 | H | F0379 | F0391  | F0447 | F0448 | F0428 | F0438 | F0386 | F0456 | F0394 | F0396 | F0468 |       |   |
| 23 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 24 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 25 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 26 |   |       |        |       |       |       |       |       |       |       |       |       |       |   |
| 27 |   | 1     | 2      | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |   |
| 28 | A | F0361 | F0370  | F0385 | F0436 | F0437 | F0407 | F0388 | F0387 | F0376 | F0360 | F0446 | F0457 |   |

Figura 4.4. Fichero de rejillas de PCR donde se puede ver la distribución irregular de las rejillas

Tabla 4.1. Descripción de las macros que componen PLATEX

| Macro components in PLATEX |                     |  |
|----------------------------|---------------------|--|
| Category                   | Macro               | Description  |
| <i>Metamacros</i>          | ExecConverter()     |  |
| <i>File macros</i>         | OpenTemplateFile()  | Opens the template text file and formats it as a new spreadsheet. The name of the template file is a parameter that can be changed by user on the corresponding cells of the Work Area spreadsheet.  |
|                            | OpenPCRGrid()       | Opens the plate grid Excel Book. The name of the PCR Grid file is a parameter that can be changed by user on the corresponding cells of the Work Area spreadsheet.   |
|                            | SaveWorkarea()      | Saves contents of the Work Area spreadsheet to prevent data lost when plate record files are created.  |
|                            | TransPlateRecord()  | Saves updated plate records as tab-delimited files with “.plt” extension as required by the data collection software. The name of plate record file is a parameter that can be changed by user on the corresponding cell of the Work Area spreadsheet.   |
|                            | SaveEmptyWorkArea() | Saves the Work Area after cleared.   |
|                            | CloseTemplateFile() | Closes the template text file.   |
| <i>Data macros</i>         | CopyGrid()          | Copies plate grid labels from PCR GRID spreadsheet to Work Area spreadsheet. The top-leftmost and bottom-rightmost range references of plate data structure in the PCR Grid file are parameters that can be changed by user on the corresponding cells of the Work Area spreadsheet.   |
|                            | CopyTemplateFile()  | Copies the sample columns from the sample spreadsheet to the Work Area spreadsheet.  |
|                            | ClearWorkArea()     | Clears the old work area contents.   |
| <i>Process macros</i>      | MakePlateRecord()   | Assigns each label to the corresponding Dye set represented by columns named “Sample Name” and “Color Info”. Typically, each DNA molecule is labeled with one Dye molecule, but up to five Dyes can be used to label the DNA sample. Each Dye molecule is represented by different color (black, blue, green, orange, and red). When all Dye molecules are used and then, each sample label is addressed to ranges of five cells corresponding to “Sample Name” and “Color Info” columns. Each task to process label assignments have the following atomic operations: (1) select label cell; (2) copy label; (3) select five cells in “Sample Name” column; (4) paste labels; (5) select five cells in “Color Info” column; and (6) paste labels. |

En la Tabla 4.2 se muestra el código fuente de la macro más importante del conjunto, denominada “MakePlateRecord”, que asigna cada etiqueta al correspondiente conjunto de pocillos (Dye).

Tabla 4.2. Código fuente de la macro MakePlateRecord

```

Sub MakePlateRecord()
'
' Macro saved on 19/09/2002 by Oscar Coltell
'
' Outer loop by columns: from "B"(2) though "M"(13)
' Inner loop by rows: from "2" though "9"
  Dim plateLeft, plateRigth, plateTop, plateBottom As Integer
  Dim rwIndex, colIndex, rwOffset As Integer
  Dim rwIndcopy1, colIndcopy1, rwIndcopy2, colIndcopy2 As Integer
  plateLeft = 2      '[Const]Leftmost column within Samples Table: "B2" or
                    ' column no. 2
  plateRigth = 13   '[Const]Rightmost column within Samples Table: "M2" or
                    ' column no. 13
  plateTop = 2      '[Const]Upper row within Samples Table: "B2:M2" or
                    ' row no. 2
  plateBottom = 9   '[Const]Lower row within Samples Table: "B9:M9" or
                    ' row no. 9
  rwIndex = 0       'row index within Samples Table: from 2 to 9
  colIndex = 0      'column index within Samples Table: from 2(B) to 13(M)
  rwOffset = 5      '[Const]offset within PCR Table for each sample set
  rwIndcopy1 = 4    'row index within PCR Table, column "Sample name" ("O")
  colIndcopy1 = 15 '[Const]column index within PCR Table, column
                    ' "Sample name" ("O")
  rwIndcopy2 = 4    'row index within PCR Table, column "Color Info" ("S")
  colIndcopy2 = 19 '[Const]column index within PCR Table, column
                    ' "Color Info" ("S")
'
  For colIndex = plateLeft To plateRigth
    For rwIndex = plateTop To plateBottom
      Cells(rwIndex, colIndex).Select
      Selection.Copy
      Range(Cells(rwIndcopy2, colIndcopy2), _
            Cells(rwIndcopy2 + 4, colIndcopy2)).Select
      Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
        SkipBlanks:= False, Transpose:=False
      Cells(rwIndcopy1, colIndcopy1).Select
      Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
        SkipBlanks:= False, Transpose:=False
      Range(Cells(rwIndcopy1 + 1, colIndcopy1), _
            Cells(rwIndcopy1 + 4, colIndcopy1)).Select
      Application.CutCopyMode = False
      Selection.ClearContents
' If Sample cell is EMPTY, it puts values '99999' into PCR cells
      If Cells(rwIndex, colIndex).Value = "" Then
        Range(Cells(rwIndcopy2, colIndcopy2), _
              Cells(rwIndcopy2 + 4, colIndcopy2)).Value = "99999"
        Range(Cells(rwIndcopy1, colIndcopy1), _
              Cells(rwIndcopy1 + 4, colIndcopy1)).Value = "99999"
      End If
      rwIndcopy1 = rwIndcopy1 + rwOffset
      rwIndcopy2 = rwIndcopy2 + rwOffset
    Next rwIndex
  Next colIndex
End Sub

```



#### 4.4.4. Descripción de las operaciones

La actividad principal del proceso se compone de seis operaciones atómicas por cada etiqueta de fragmento de ADN. Entonces, para procesar una placa de 96 pocillos se necesitan un total de 576 operaciones, o 2304 operaciones para una placa de 384 pocillos. El tiempo medio empleado en cada operación manual es de 15 segundos. Esto hace que el tiempo medio en procesar manualmente todas las celdas en los ficheros de rejillas de PCR sea de 2,4 horas (17.280 segundos) para las placas de 96, y de 9,6 horas (69.120 segundos) para las placas de 384. Si se suma el tiempo necesario para abrir los ficheros, copiar y pegar el contenido de las celdas, y guardar los datos actualizados en nuevos ficheros de texto, el tiempo total medio varía entre 2,5 horas y 10 horas para cada uno de los tipos de placas mencionados.

Por otra parte, el tiempo medio medido para ejecutar automáticamente el proceso mostrado en la Figura 4.5.5, mediante la solución del conjunto de macros, es de 30 segundos para las placas de 96, y de 120 segundos para las placas de 384. Esto muestra que dicho proceso es alrededor de 288 veces más rápido.

#### 4.2.2 Visualización y manejo de secuencias (SEQPACKER)

Estos resultados se han remitido respectivamente, en el año 2004, al *Fourth IEEE Symposium on Bioinformatics and Bioengineering 2004* (BIBE2004) con el título “SeqPacker: a nucleotide sequence manipulation utility” (Coltell et al., 2004 [BIBE04]); y al *2nd International Workshop on Biological Data Management 2004* (BIDM'04) con el título “A Biologist-Friendly User Interface to Manipulate Nucleotide Sequences in Genomic Epidemiology” (Coltell et al., 2004 [BDIM04]).

Se describe en este apartado una solución bioinformática que es una herramienta que integra las funciones de manipulación de secuencias de ADN y ARN, con las funciones de visualización gráfica potente. Esta herramienta es parte de un proyecto más ambicioso para construir un paquete (o suite) integrado de sencillas soluciones bioinformáticas de soporte a la investigación genómica, que permitan manipular, visualizar, comprimir, y otras operaciones adicionales, secuencias de ADN, ARN y aminoácidos. Este paquete se ha denominado inicialmente “SeqPacker”, porque se dedica a manipular secuencias de nucleótidos y a aplicar algoritmos de compresión para reducir los espacios de almacenamiento y tiempos de transmisión. Parte de este paquete se presenta en este apartado y otra parte desarrollada se presenta en el apartado siguiente.

#### 4.2.2.1 Características generales de la solución bioinformática

La herramienta que se describe aquí se ha llamado también “SeqPacker” por traslación del paquete global. Se puede entender SeqPacker como un sencillo pero potente editor de secuencias de nucleótidos, ADN y ARN por el momento, mediante la aplicación de tecnologías estándar (OMG, 2004) y estándares de representación de nucleótidos como FASTA (GenBank, 2004), que ha servido de ayuda en la investigación en Epidemiología Genómica Cardiovascular.

La versión de SeqPacker publicada hasta la fecha, la versión 2.0) tiene las siguientes características:

1. Representación de secuencias de nucleótidos en un formato agradable y fácilmente manejable, agrupando las bases en columnas de 5 o 10, con posibilidad de numerar o no las líneas, y representando cada nucleótido con su color estándar según el *NBI SNPSHOT system* (A: verde; C: negro; G: azul; T o U: rojo) (Figura 4.5) (GenBank, 2004). También se puede aplicar el formato FASTA (FASTA, 2004). Como facilidad adicional, el tamaño y el tipo de letra de los nucleótidos se pueden modificar. Para preservar la regularidad de las columnas, solamente se permiten los tipos de letra que no ajustan los espacios en blanco.

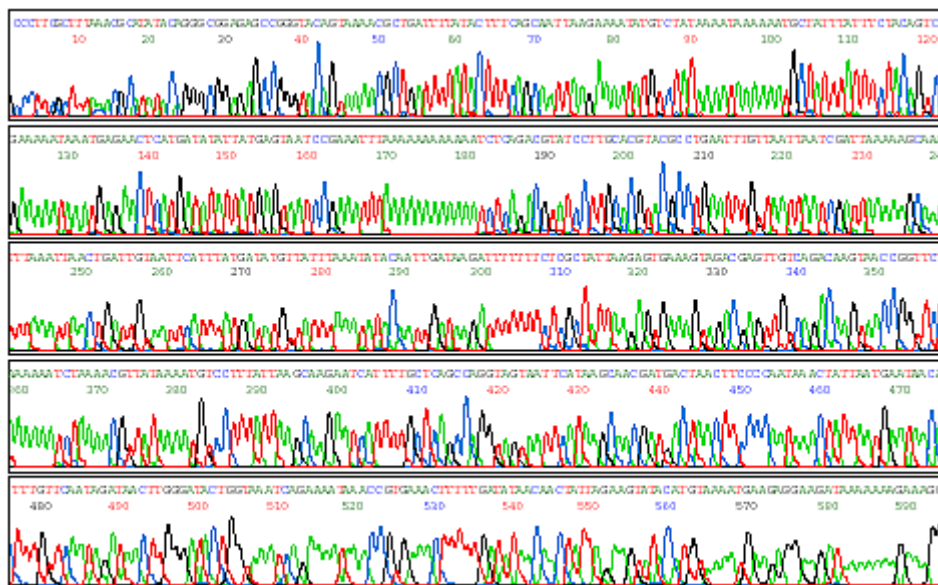


Figura 4.5. Representación habitual de la salida de un secuenciador ABI PRISM con el código de colores para las bases. Se denomina cromatograma

- Una interfaz gráfica con elementos que flexibilizan mucho el uso: división en paneles de la información a presentar, botones de acceso rápido asociados a las operaciones principales de cada panel, reconfiguración automática de cada panel cuando se modifica el tamaño de la pantalla general de la herramienta (Figura 4.6).

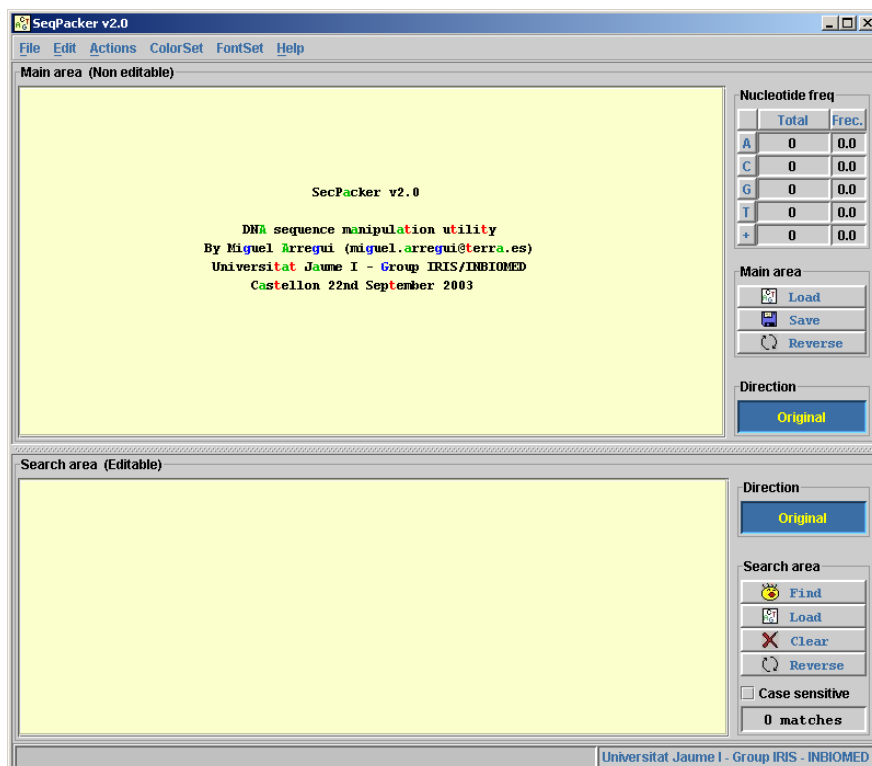


Figura 4.6. Interfaz general de SeqPacker

- Se pueden leer y escribir ficheros de secuencias en formato FASTA, en texto llano y en formato ABI (Applied Biosystems, 2004). Este último es un formato binario que utilizan los instrumentos de Applied Biosystems en los ficheros de salida de su software de control y gestión.
- Permite trasladar secuencias obtenidas desde Internet directamente a la ventana de trabajo mediante el clipboard.
- Contiene un motor de búsqueda para búsquedas exactas. La secuencia principal se sitúa en el panel superior, y la secuencia a buscar o patrón se sitúa en el panel inferior (Figura 4.7). El modo de búsqueda reconoce mayúsculas y

minúsculas pero inicialmente está desactivado (modo por defecto). Se puede activar desde el menú. La secuencia patrón se puede cargar desde un fichero o se puede escribir directamente en el panel inferior.

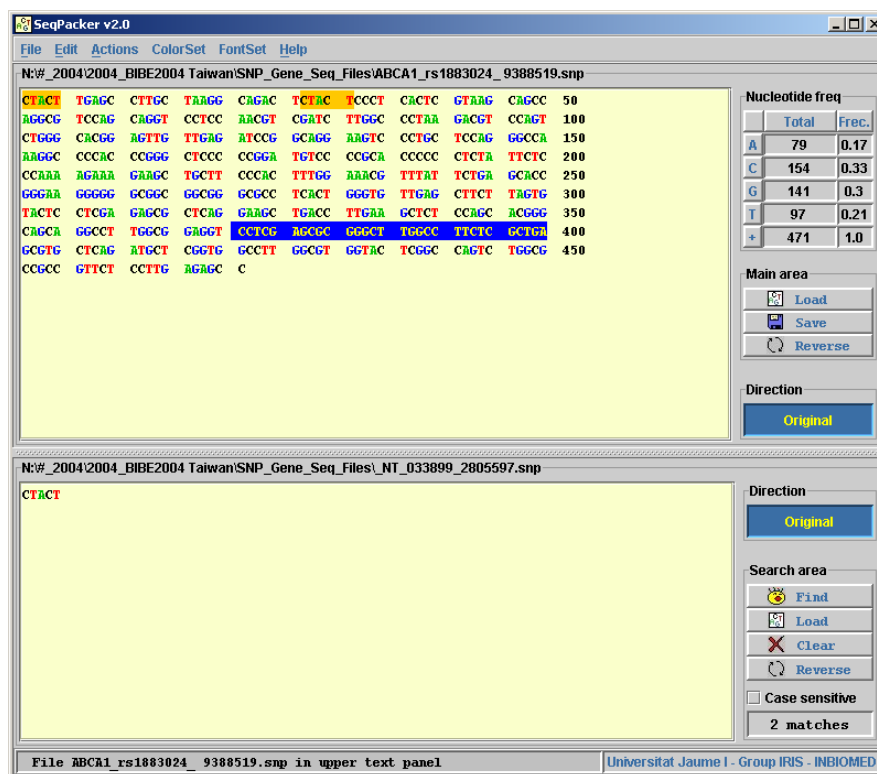


Figura 4.7. Un caso de búsqueda de una pequeña secuencia

6. Puede trabajar con cadenas directas (forward, “5’-“) o inversas (reverse, “3’-“). Cada panel muestra en un botón el sentido de la cadena que contiene (Figura 4.7). Esto permite realizar búsquedas combinando los sentidos de las secuencias: directo-directo, directo-inverso, inverso-directo e inverso-inverso.
7. Dado que la herramienta está codificada en Java, está asegurada su portabilidad a una amplia gama de plataformas. Utiliza el Java Runtime Environment (JRE) y puede ejecutarse en las siguientes plataformas: MS Windows: 98SE, 2000, XP; Linux: RedHat 8.0, SUSE 9.0; y MacOS X. Con esta tecnología, es muy fácil extender la aplicación a otras plataformas.

El propósito del desarrollo de SeqPacker es proporcionar un editor para manipulación de secuencias de nucleótidos con tecnología estandarizada para cada

plataforma informática, de forma pueda beneficiar a tantos usuarios como sea posible.

#### 4.2.2.2 Estructura interna de las funciones

Se diseñó SeqPacker inicialmente para gestionar ficheros en formato ABI producidos por el software Chromas (Chromas, 2004), que se utiliza para ver el cromatograma producido por un secuenciador de ADN de Applied Biosystems o de Amersham MegaBace (Figura 4.5). Pero pronto se amplió para trabajar con otros formatos de ficheros como FASTA y GenBank. Las funciones que ofrece esta aplicación se muestran en la Figura 4.8.

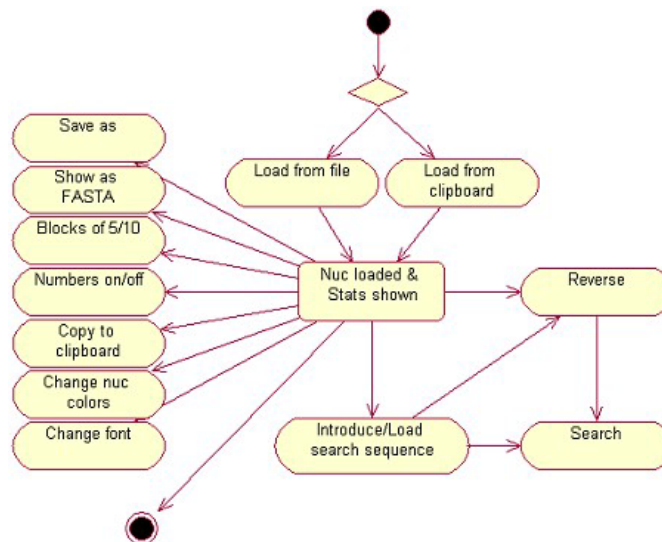


Figura 4.8. Flujo de trabajo de SeqPacker

El usuario tiene la opción de abrir un nuevo fichero o cargar una secuencia desde el clipboard. Esta última opción permite la copia de secuencias desde fuentes heterogéneas, tales como Internet (por ejemplo de una página HTML de GenBank), puesto que el algoritmo de inspección (parsing) es capaz de filtrar cualquier carácter distinto de los nucleótidos que forman la secuencia de interés (retornos de carro, espacios, etc.) Una vez se ha obtenido la secuencia, la aplicación permite realizar diversas transformaciones visuales sobre dicha secuencia (Figura 4.9).

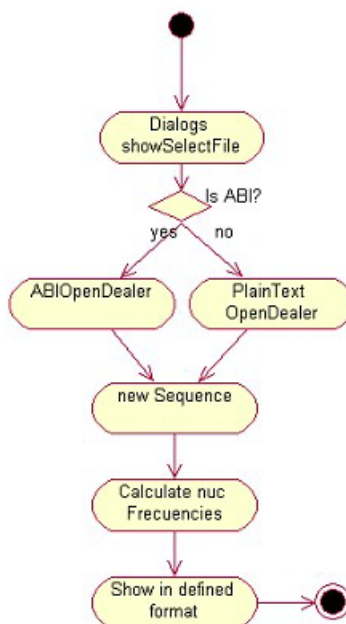


Figura 4.9. Flujo de trabajo para la visualización de secuencias

Por ejemplo, se puede poner la secuencia de entrada en formato FASTA, o en grupos de 5 o 10 bases. Además, también se pueden mostrar u ocultar los contadores de nucleótidos que se sitúan en el extremo derecho de cada línea, o también cambiar el color, tipo y tamaño de la fuente. SeqPacker también calcula muy rápidamente el complementario inverso de una secuencia nada más pulsar el botón del ratón. Otra función es la de hacer búsquedas exactas simples sobre la secuencia de trabajo, detectando si el usuario está intentando casar secuencias no compatibles, como ARN con ADN o viceversa. Una vez el usuario ha terminado de transformar la secuencia, la puede guardar para usos futuros, siempre en el mismo formato que se muestra en el panel, pero añadiendo una línea con las típicas cabeceras de FASTA (nombre del fichero y comentarios).

Las Figuras 4.10 y 4.11 muestran respectivamente el modelo de objetos de la interfaz gráfica y de la aplicación en notación UML. El segundo modelo representa las clases implicadas en el flujo de trabajo de lectura de los ficheros y aprovecha parte de las clases de los submodelos del formato ABI (Figura 4.5).

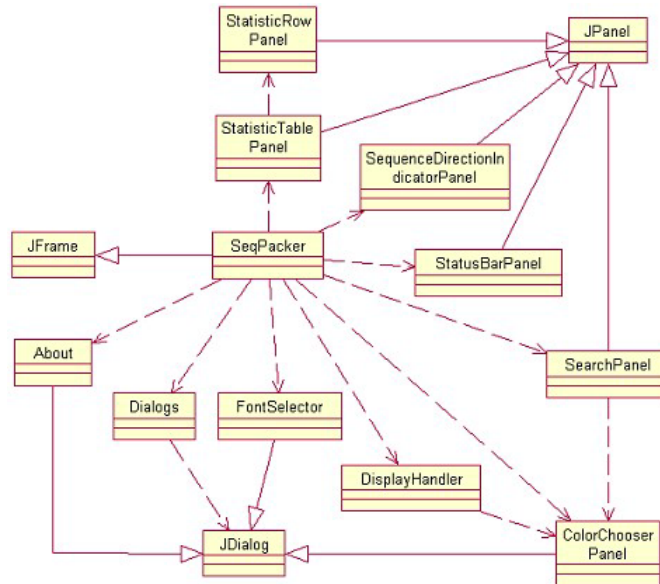


Figure 4.10. El modelo de objetos de la interfaz gráfica

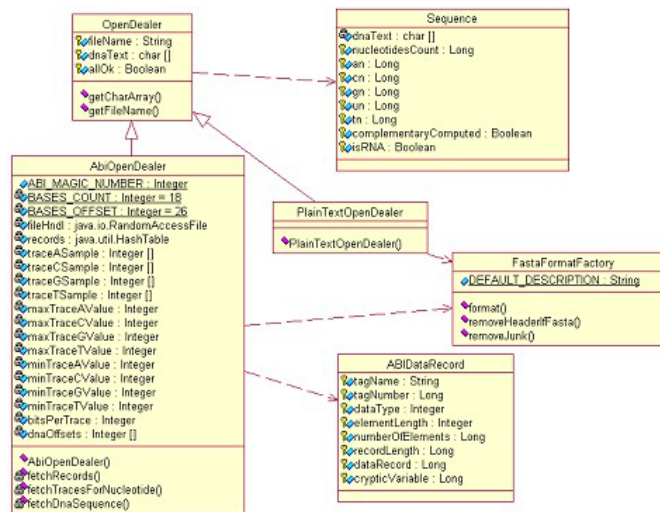


Figure 4.11. El modelo de objetos de aplicación

### 4.2.3 Búsqueda y detección de polimorfismos (PREDICARD)

Estos resultados se han presentado en el “Curso de Informática Biomédica”, celebrado el 13 de Octubre de 2003 en el Instituto de Salud Carlos III (ISCIII) y se ha elaborado un artículo con el título “La representación, almacenamiento e interrelación de entidades genómicas y proteómicas en las bases de datos públicas: un caso de estudio” (Coltell et al., 2003 [ISCIII]), que aparecerá en las actas del curso en el curso de este año de 2004.

Se describe en este apartado el proceso que se aplicó para la búsqueda e identificación de polimorfismos de genes asociados con la ocurrencia de ECV. Como la lista inicial de genes encontrados en una primera exploración era muy grande (mas de 100 genes), se calculó que la lista de polimorfismos asociados podría llegar a ser el quintuplo. Por lo tanto, se decidió que se debería trabajar inicialmente con un grupo reducido de dichos genes y polimorfismos asociados, a modo de proyecto piloto, para después ampliar la lista a conveniencia y en función de la obtención y publicación de nuevas evidencias. Este es el proceso que se describe en esta subsección como resultado concreto del proyecto realizado en el marco de esta tesis. Este proceso consta de 5 actividades (o fases) que son las siguientes:

1. Confección de la lista de Genes.
2. Obtención de la lista de genes.
3. Identificación de los polimorfismos.
4. Documentación de los polimorfismos.
5. Extracción de las subsecuencias de SNP de los polimorfismos.

En esta sección se describen con detalle cada una de estas actividades.

#### 4.2.3.1 Confección de la lista de Genes

Esta actividad se ha llevado a cabo con la función *EVIDENCE COLLECT* del proyecto. Se han recogido las evidencias experimentales a partir de los genes en estudio en el Nutrigenomics Lab. Por otra parte, se han buscado las evidencias en los genes reportados en la literatura científica mediante varias prospecciones con el NCBI PubMed.

Parte importante es la investigación preliminar sobre los posibles genes implicados en las enfermedades cardiovasculares, y en los sistemas biológicos como manifestaciones fenotípicas y factores de riesgo (Blaschke et al., 2002). Esta



investigación se basó en la búsqueda en los artículos publicados en las revistas más afines (*New England Journal of Medicine, Circulation, Stroke, etc.*) y en la selección de los genes reportados según un conjunto mínimo de criterios aplicados. Además, se intentó identificar los polimorfismos significativos de estos genes que pudieran ser de interés en el proyecto.

Para la selección de las referencias que pudieran ser válidas, se establecieron algunos criterios según el conocimiento de los expertos participantes y las evidencias obtenidas. Estos criterios fueron los siguientes:

1. Que hubiera más de una evidencia publicada sobre el mismo gen.
2. Que se garantizara la reproducibilidad de los resultados.
3. Que se manifestara la generalidad de la población estudiada.
4. Que se asegurara el tamaño mínimo de muestra utilizada en los estudios.
5. Que se tuviera conocimiento del equipo investigador y de su fiabilidad investigadora (no es oro todo lo que reluce, ni siquiera en investigación).
6. Que se aplicara en último lugar la opinión de los expertos para la toma de decisiones.

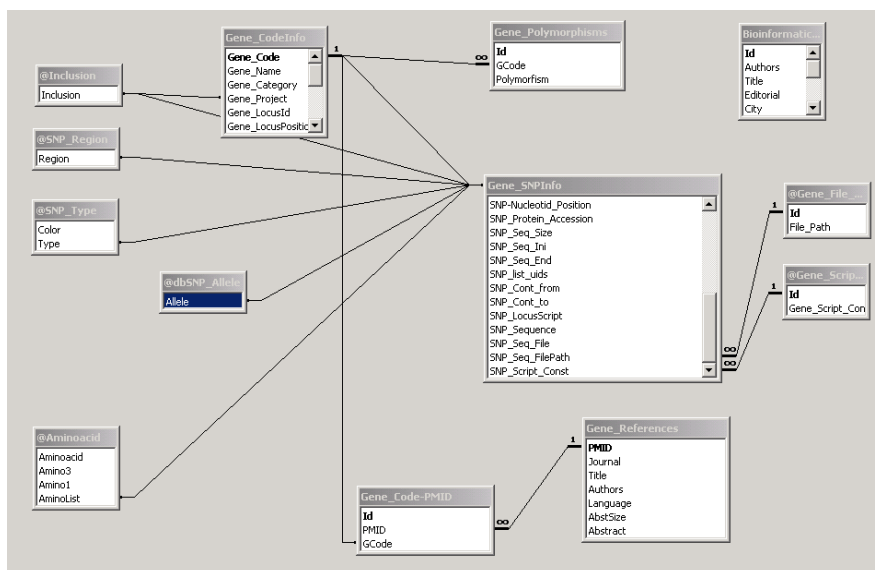


Figura 4.12. Esquema lógico de la base de datos

---

Los resultados parciales se guardaron en una base de datos creada para almacenar toda la información del proyecto. En la Figura 4.12 se muestra el esquema lógico de esta base de datos.

Los resultados de esta actividad se guardaron en las tablas Gene-CodeInfo, Gene-Code-PMID, Gene-References de la base de datos. La lista preliminar obtenida de los genes candidatos se presenta en la Tabla 4.3, donde se muestran los genes encontrados, la posición que ocupan en los cromosomas respectivos y las referencias utilizadas para el proceso. Estas referencias se han incluido en una subsección específica al final de este Capítulo y también en el Capítulo de Bibliografía general.

Tabla 4.3. Genes y Polimorfismos Identificados. Proceso preliminar

| Gen   | Cromo-soma | SNP | Diana | Evi-den-cia | Referencias  |
|-------|------------|-----|-------|-------------|--|
| APOA1 | 11q21      |     |       |             | Ordovas JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PW, Schaefer EJ, Tucker K. Polyunsaturated fatty acids modulate the effects of the APOA1 G-A polymorphism on HDL Cholesterol concentrations in a sex-specific manner: the Framingham Study. <i>Am J Clin Nutr.</i> 2002; 75:38-46.  |
| APOA4 | 11q21      |     |       |             | Mata P, Ordovas JM, Lopez-Miranda J, Lichtenstein AH, Clevidence B, Judd JT, Schaefer EJ. ApoA-IV phenotype affects diet-induced plasma LDL cholesterol lowering. <i>Arterioscler Thromb.</i> 1994; 14:884-91.   |
| APOC3 | 11q21      |     |       |             | Olivieri O, Stranieri C, Bassi A, Zaia B, Girelli D, Pizzolo F, Trabetti E, Cheng S, Grow MA, Pignatti PF, Corrocher R. ApoC-III gene polymorphisms and risk of coronary artery disease. <i>J Lipid Res.</i> 2002;43:1450-7.   |
| APOE  | 19         |     |       |             | Ordovas JM, Mooser V. The APOE locus and the pharmacogenetics of lipid response. <i>Curr Opin Lipidol.</i> 2002; 13:113-7. Review.   |
| APOA5 | 11q21      |     |       |             | Martin S, Nicaud V, Humphries SE, Talmud PJ; EARS group. Contribution of APOA5 gene variants to plasma triglyceride determination and to the response to both fat and glucose tolerance challenges. <i>Biochim Biophys Acta.</i> 2003;1637:217-25.<br>Talmud PJ, Hawe E, Martin S, Olivier M, Miller GJ, Rubin EM, Pennacchio LA, Humphries SE. Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. <i>Hum Mol Genet.</i> 2002; 11:3039-46.  |
| LIPC  | 15q21-q23  |     |       |             | Zambon A, Deeb SS, Pualetto P, Crepaldi G, Brunzell JD. Hepatic lipase: a marker for cardiovascular disease risk and response to therapy. <i>Curr Opin Lipidol.</i> 2003;14:179-89.<br>Deeb SS, Zambon A, Carr MC, Ayyobi AF, Brunzell JD. Hepatic lipase and dyslipidemia: interactions among genetic variants, obesity, gender, and diet. <i>J Lipid Res.</i> 2003; 44:1279-86.<br>Ordovas JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PW, Schaefer EJ, Tucker KL. Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study. <i>Circulation.</i> 2002; 106:2315-21.<br>Andersen RV, Wittrup HH, Tybjaerg-Hansen A, Steffensen R, Schnohr P, Nordestgaard BG. Hepatic lipase mutations, elevated high-density lipoprotein cholesterol, and increased risk of ischemic heart disease: the Copenhagen City Heart Study. <i>J Am Coll Cardiol.</i> 2003; 41:1972-82. |

| Gen  | Cromo-<br>soma | SNP | Diana | Evi-<br>den-<br>cia | Referencias  |
|------|----------------|-----|-------|---------------------|--|
| LPL  | 8p22           |     |       |                     | <p>Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Folch J, Ordovas JM. Associations of LPL and APOC3 gene polymorphisms on plasma lipids in a Mediterranean population: interaction with tobacco smoking and the APOE locus. <i>J Lipid Res.</i> 2002; 43:416-27.</p> <p>Kastelein JJ, Ordovas JM, Wittekoek ME, Pimstone SN, Wilson WF, Gagne SE, Larson MG, Schaefer EJ, Boer JM, Gerdes C, Hayden MR. Two common mutations (D9N, N291S) in lipoprotein lipase: a cumulative analysis of their influence on plasma lipids and lipoproteins in men and women. <i>Clin Genet.</i> 1999; 56:297-305.</p> <p>Gagne SE, Larson MG, Pimstone SN, Schaefer EJ, Kastelein JJ, Wilson PW, Ordovas JM, Hayden MR. A common truncation variant of lipoprotein lipase (Ser447X) confers protection against coronary heart disease: the Framingham Offspring Study. <i>Clin Genet.</i> 1999; 55:450-4.</p>  |
| CETP | 16q21          |     |       |                     | <p>Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. <i>Arterioscler Thromb Vasc Biol.</i> 2000; 20:1323-9.</p> <p>Agerholm-Larsen B, Tybjaerg-Hansen A, Schnohr P, Steffensen R, Nordestgaard BG. Common cholesteryl ester transfer protein mutations, decreased HDL cholesterol, and possible decreased risk of ischemic heart disease: The Copenhagen City Heart Study. <i>Circulation.</i> 2000; 102:2197-203.</p> <p>Blankenberg S, Rupprecht HJ, Bickel C, Jiang XC, Poirier O, Lackner KJ, Meyer J, Cambien F, Tiret L; AtheroGene Investigators. Common genetic variation of the cholesteryl ester transfer protein gene strongly predicts future cardiovascular death in patients with coronary artery disease. <i>J Am Coll Cardiol.</i> 2003; 41:1983-9.</p> |

| Gen    | Cromosoma   | SNP | Diana | Evidencia | Referencias  |
|--------|-------------|-----|-------|-----------|--|
| SCARB1 | 12q         |     |       |           | <p>McCarthy JJ, Lehner T, Reeves C, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ; Genequest investigators. Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. <i>J Med Genet.</i> 2003; 40:453-8.</p> <p>Osgood D, Corella D, Demissie S, Cupples LA, Wilson PW, Meigs JB, Schaefer EJ, Coltell O, Ordovas JM. Genetic variation at the scavenger receptor class B type I gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: the Framingham study. <i>J Clin Endocrinol Metab.</i> 2003; 88:2869-79.</p> <p>Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. <i>Arterioscler Thromb Vasc Biol.</i> 1999; 19:1734-43.</p> |
| PPARA  | 22q12-q13.1 |     |       |           | <p>Tai ES, Demissie S, Cupples LA, Corella D, Wilson PW, Schaefer EJ, Ordovas JM. Association between the PPARA L162V polymorphism and plasma lipid levels: the Framingham Offspring Study. <i>Arterioscler Thromb Vasc Biol.</i> 2002; 22:805-10.</p> <p>Flavell DM, Jamshidi Y, Hawe E, Pineda Torra I, Taskinen MR, Frick MH, Nieminen MS, Kesaniemi YA, Pasternack A, Staels B, Miller G, Humphries SE, Talmud PJ, Syvanne M. Peroxisome proliferator-activated receptor alpha gene variants influence progression of coronary atherosclerosis and risk of coronary artery disease. <i>Circulation.</i> 2002; 105:1440-5.</p>  |
| PPARG  | 3p25        |     |       |           | <p>Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. <i>Nat Genet.</i> 1998; 20:284-7.</p>   |
| ABCA1  | 9q22-q31    |     |       |           | <p>Clee SM, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, Jukema JW, van Wijland M, van Dam M, Hudson TJ, Brooks-Wilson A, Genest J Jr, Kastelein JJ, Hayden MR. Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. <i>Circulation.</i> 2001; 103:1198-205.</p>  |

| Gen  | Cromo-soma    | SNP | Diana | Evi-dencia | Referencias  |
|------|---------------|-----|-------|------------|--|
| ESR1 | 6q25.1        |     |       |            | Herrington DM, Howard TD, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Brosnihan KB, Meyers DA, Bleecker ER. Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease.[comment]. [Clinical Trial. Journal Article] <i>New England Journal of Medicine</i> ; 2002; 346:967-74, 2002  |
| ADH3 | 4q22          |     |       |            | Hines L. M., Stampfer M. J., Ma J., Gaziano J. M., Ridker P. M., Hankinson S. E., Sacks F., Rimm E. B., Hunter D. J. N. Genetic Variation in Alcohol Dehydrogenase and the Beneficial Effect of Moderate Alcohol Consumption on Myocardial Infarction <i>Engl J Med</i> 2001; 344:549-555.   |
| GCLM | 1p22.1        |     |       |            | Nakamura S, Kugiyama K, Sugiyama S, Miyamoto S, Koide S, Fukushima H, Honda O, Yoshimura M, Ogawa H. Polymorphism in the 5'-flanking region of human glutamate-cysteine ligase modifier subunit gene is associated with myocardial infarction. <i>Circulation</i> . 2002; 105:2968-73.   |
| MMP9 | 20q11.2-q13.1 |     |       |            | Morgan AR, Zhang B, Tapper W, Collins A, Ye S. Haplotypic analysis of the MMP-9 gene in relation to coronary artery disease. <i>J Mol Med</i> . 2003; 81:321-6.<br>Blankenberg S, Rupprecht HJ, Poirier O, Bickel C, Smieja M, Hafner G, Meyer J, Cambien F, Tiret L; AtheroGene Investigators. Plasma concentrations and genetic variation of matrix metalloproteinase 9 and prognosis of patients with cardiovascular disease. <i>Circulation</i> . 2003; 107:1579-85.   |
| IL6  | 7p21          |     |       |            | Jerrard-Dunne P, Sitzer M, Risley P, Steckel DA, Buehler A, von Kegler S, Markus HS; Carotid Atherosclerosis Progression Study. Interleukin-6 promoter polymorphism modulates the effects of heavy alcohol consumption on early carotid artery atherosclerosis: the Carotid Atherosclerosis Progression Study (CAPS). <i>Stroke</i> . 2003; 34:402-7.<br>Basso F, Lowe GD, Rumley A, McMahon AD, Humphries SE. Interleukin-6 -174G>C polymorphism and risk of coronary heart disease in West of Scotland coronary prevention study (WOSCOPS). <i>Arterioscler Thromb Vasc Biol</i> . 2002; 22:599-604.<br>Jenny NS, Tracy RP, Ogg MS, Luongle A, Kuller LH, Arnold AM, Sharrett AR, Humphries SE. In the elderly, interleukin-6 plasma levels and the -174G>C polymorphism are associated with the development of cardiovascular disease. <i>Arterioscler Thromb Vasc Biol</i> . 2002; 22:2066-71. |

| Gen  | Cromo-soma | SNP | Diana | Evi-den-cia | Referencias  |
|------|------------|-----|-------|-------------|--|
| CD14 | 5q31.1     |     |       |             | Risley P, Jerrard-Dunne P, Sitzler M, Buehler A, von Kegler S, Markus HS; Carotid Atherosclerosis Progression Study. Promoter polymorphism in the endotoxin receptor (CD14) is associated with increased carotid atherosclerosis only in smokers: the Carotid Atherosclerosis Progression Study (CAPS). <i>Stroke</i> . 2003; 34:600-4.  |
| TLR4 | 9q32-q33   |     |       |             | Boekholdt SM, Agema WR, Peters RJ, Zwinderman AH, van der Wall EE, Reitsma PH, Kastelein JJ, Jukema JW; REGression GRowth Evaluation Statin Study Group. Variants of toll-like receptor 4 modify the efficacy of statin therapy and the risk of cardiovascular events. <i>Circulation</i> . 2003; 107:2416-21.   |
| NOS3 | 7q36       |     |       |             | Rossi GP, Cesari M, Zanchetta M, Colonna S, Maiolino G, Pedon L, Cavallin M, Maiolino P, Pessina AC. The T-786C endothelial nitric oxide synthase genotype is a novel risk factor for coronary artery disease in Caucasian patients of the GENICA study. <i>J Am Coll Cardiol</i> . 2003; 41:930-7.<br><br>Gomma AH, Elrayess MA, Knight CJ, Hawe E, Fox KM, Humphries SE. The endothelial nitric oxide synthase (Glu298Asp and -786T>C) gene polymorphisms are associated with coronary in-stent restenosis. <i>Eur Heart J</i> . 2002; 23:1955-62. |

#### 4.2.3.2 Obtención de la lista de genes

Dado que el número de genes identificados en el proceso preliminar fueron más de cien, y los polimorfismos asociados podían subir a centenares, se decidió empezar con un subconjunto de los 20 genes más significativos y/o más conocidos, con los cuales se confeccionó una tabla como la Tabla 4.4.

**Tabla 4.4. Actividad 2: Lista de 20 genes (parcial)**

| Gen    | Cromosoma   | SNP | Diana | Evidencia |
|--------|-------------|-----|-------|-----------|
| APOA1  | 11q21       |     |       |           |
| APOA4  | 11q21       |     |       |           |
| APOC3  | 11q21       |     |       |           |
| APOE   | 19          |     |       |           |
| APOA5  | 11q21       |     |       |           |
| LIPC   | 15q21-q23   |     |       |           |
| LPL    | 8p22        |     |       |           |
| CETP   | 16q21       |     |       |           |
| SCARB1 | 12q         |     |       |           |
| PPARA  | 22q12-q13.1 |     |       |           |
| PPARG  | 3p25        |     |       |           |
| ABCA1  | 9q22-q31    |     |       |           |
| ESR1   | 6q25.1      |     |       |           |
| ADH3   | 4q22        |     |       |           |
| .....  | .....       |     |       |           |

#### 4.2.3.3 Identificación de los polimorfismos

Esta actividad se efectuó con la función *POLYMORFISM IDENTIFICATION* del proyecto. Se tomó como línea base la lista de la Tabla 4.4 y la información asociada a cada gen. A continuación se identificaron los polimorfismos en las bases de datos de NCBI GenBank: Nucleotide, SNP, Protein, OMIM, etc. Las tareas ejecutadas para cada uno de los genes fueron las siguientes:

1. Se exploraron las evidencias publicadas por cada gen buscando los datos siguientes:
  - Identificador de cada polimorfismo.
  - Fenotipo asociado.
  - Variación de nucleótidos en ARN: A/G, T/C.
  - Variación de aminoácidos en proteína: Ala/Glu.



- Obtención del SNP de cada polimorfismo en la secuencia de nucleótidos del gen en NCBI SNP.

El resultado se muestra en la Tabla 4.5.

**Tabla 4.5. Actividad 3: Lista de 20 genes (parcial)**

| Gen    | Cromosoma   | SNP                         | Diana | Evidencia |
|--------|-------------|-----------------------------|-------|-----------|
| APOA1  | 11q21       | -75G/A                      |       |           |
| APOA4  | 11q21       | G360H<br>Ser347Tre          |       |           |
| APOC3  | 11q21       | SstI                        |       |           |
| APOE   | 19          | E2/E3/E4                    |       |           |
| APOA5  | 11q21       | 56C/G<br>-1131T/C           |       |           |
| LIPC   | 15q21-q23   | C-514T                      |       |           |
| LPL    | 8p22        | S447X<br>D9N,<br>N291S      |       |           |
| CETP   | 16q21       | Taq1B                       |       |           |
| SCARB1 | 12q         | Exon1,<br>Intron5,<br>exon8 |       |           |
| PPARA  | 22q12-q13.1 | L162V                       |       |           |
| PPARG  | 3p25        | Pro12Ala                    |       |           |
| ABCA1  | 9q22-q31    | R219K                       |       |           |
| ESR1   | 6q25.1      | IVS1-401                    |       |           |
| ADH3   | 4q22        | Gamma1/g<br>amma2           |       |           |
| .....  | .....       |                             |       |           |

Los resultados parciales se guardaron en las tablas Gene-CodeInfo, Gene-SNPInfo y Gene-Polymorphism base de datos.

#### 4.2.3.4 Documentación de los polimorfismos

Esta actividad se llevó a cabo con la función *POLYMORFISM & SNP DOCUMENTATION* del proyecto. Se partió de la información de la Tabla 4.5 y se documentaron los polimorfismos y los SNP asociados mediante las bases de datos de NCBI GenBank: Nucleotide, SNP, Protein, OMIM, etc. Las tareas ejecutadas para cada uno de los polimorfismos fueron las siguientes:

- Justificación de la selección del polimorfismo: Objetivo o Diana.
- Descripción de la Evidencia.
- Inclusión de las referencias bibliográficas.

## 4. Obtención de los datos del SNP de cada polimorfismo:

- dbSNP.
- Posición según contig.
- Determinación de la subcadena que va a contener el SNP:
  - nucleótido de más a la izquierda.
  - nucleótido de más a la derecha.
- Determinación de las direcciones de acceso.

En la Figura 4.13 se muestra la interfaz de la base de datos que se diseñó y utilizó en las tareas para la identificación de polimorfismos.

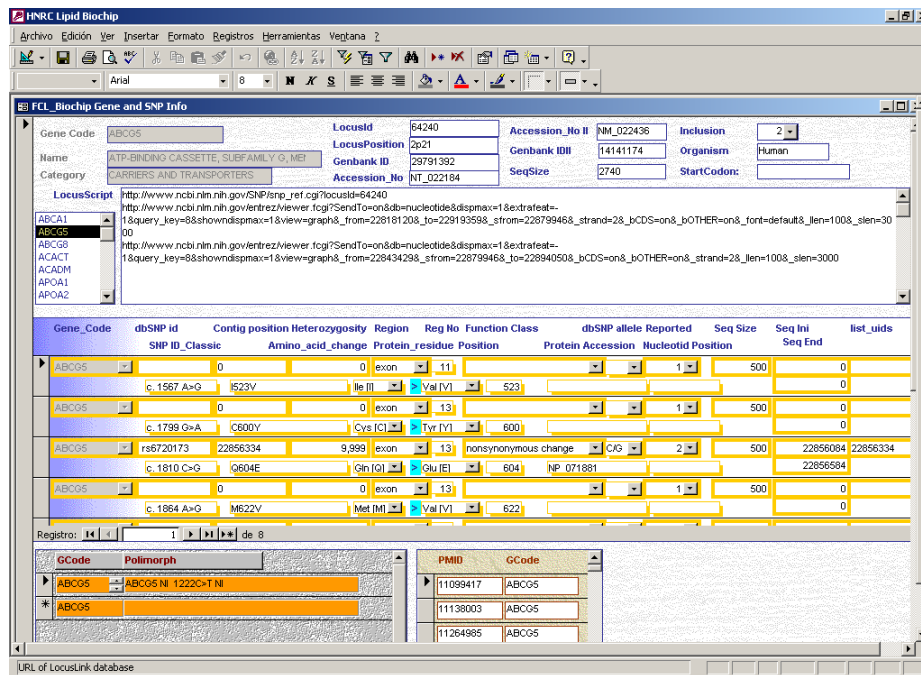


Figura 4.13. Interfaz de la base de datos para la documentación de polimorfismos

Posteriormente, una vez completada la lista de polimorfismos se procedió a ejecutar las tareas siguientes:

1. Se confeccionó una Tabla Resumen (Tabla 4.6).
2. Se completó la información de la Base de Datos del Modelo.

3. Se diseñó la estructura de ficheros de los resultados.

**Tabla 4.6. Actividad 4: Lista de 20 genes (parcial)**

| Gen    | Cromosoma   | SNP                         | Diana                         | Evidencia  |
|--------|-------------|-----------------------------|-------------------------------|--|
| APOA1  | 11q21       | -75G/A                      | Respuesta a terapia dietética | Determina la respuesta de los niveles de C-HDL a cambios en el consumo de ácidos grasos poliinsaturados.                 |
| APOA4  | 11q21       | G360H<br>Ser347Tre          | Respuesta a terapia dietética | Determina la respuesta de C-LDL a cambios en el consumo de colesterol. También afecta el estado de lipemia postprandial. |
| APOC3  | 11q21       | SstI                        | Riesgo                        | Niveles de triglicéridos, potencialmente asociado con un aumento del riesgo cardiovascular                               |
| APOE   | 19          | E2/E3/E4                    | Riesgo y respuesta            | Riesgo Cardiovascular. Respuesta a terapias (Dieta, estatinas).  |
| APOA5  | 11q21       | 56C/G<br>-1131T/C           | Riesgo                        | Hipertrigliceridemia   |
| LIPC   | 15q21-q23   | C-514T                      | Riesgo y respuesta            | Niveles de C-HDL, Riesgo cardiovascular y respuesta a terapia  |
| LPL    | 8p22        | S447X<br>D9N,<br>N291S      | Respuesta a terapia dietética | Determina la respuesta de los niveles de C-HDL a cambios en el consumo de ácidos grasos poliinsaturados.                 |
| CETP   | 16q21       | Taq1B                       | Riesgo                        | c-HDL, Riesgo cardiovascular   |
| SCARB1 | 12q         | Exon1,<br>Intron5,<br>exon8 | Riesgo                        | c-LDL, c-HDL   |
| PPARA  | 22q12-q13.1 | L162V                       | Riesgo y respuesta            | c-LDL, c-HDL, respuesta a niveles de ácidos grasos poliinsaturados en la dieta.  |
| PPARG  | 3p25        | Pro12Ala                    | Riesgo y respuesta            | Riesgo, c-HDL  |
| ABCA1  | 9q22-q31    | R219K                       | Riesgo                        | Riesgo, c-HDL  |
| ESR1   | 6q25.1      | IVS1-401                    | Respuesta en Mujeres          | HDL, y respuesta a terapia hormonal sustitutiva.   |
| ADH3   | 4q22        | Gamma1/g<br>amma2           | Riesgo y respuesta            | c-HDL, riesgo cardiovascular y respuesta al consumo de alcohol.  |
| .....  | .....       |                             |                               |  |

Los resultados parciales se guardaron en las tablas Los resultados parciales se guardaron en las tablas Gene-CodeInfo, Gene-SNPInfo y Gene-Polymorphism base de datos.

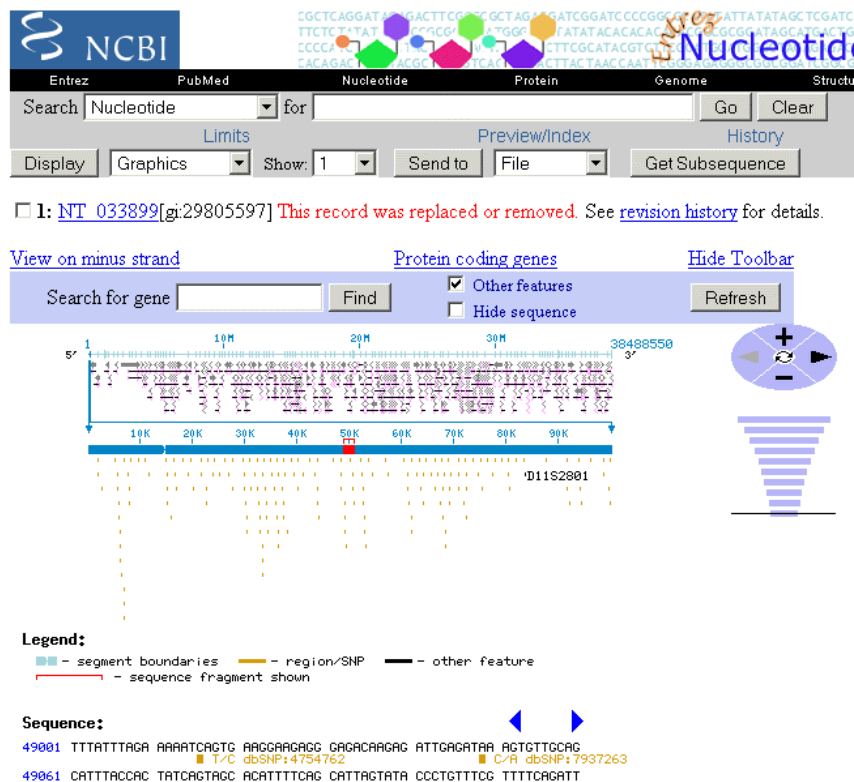
#### 4.2.3.5 Extracción de las subsecuencias de SNP de los polimorfismos

Esta actividad se realizó con la función *POLYMORFISM SNP-SEQUENCE EXTRACTION*. Se partió de la lista de SNP de la Tabla 4.6 y de la información sobre las cadenas de ADN en las tablas de la base de datos. Con toda esta información se obtuvieron las subcadenas de longitud 501 bases para cada SNP

identificado en formato FASTA(Figura 4.14), mediante la aplicación del *NCBI Sequence Viewer* sobre las bases de datos de *NCBI Genbank LocusLink* (Figura 4.15).

```
>gi|29805597:20249928-20250398 Homo sapiens chromosome 11 genomic contig
CTACTTGAGCCTTGCTAAGGCAGACTCTACTCCCTCACTCGTAAGCAGCCAGGCGTCCAGCAGGTCCTCC
AACGTCGATCTTGGCCCTAAGACGTCCAGTCTGGGCACGGAGTTGTTGAGATCCGGCAGGAAGTCCCTGC
TCCAGGGCCAAAGGCCCAACCCGGGCTCCCCGGATGTCCCCGCACCCCTCTATTCTCCCAAAGAAA
GAAGCTGCTTCCCACTTTGGAAACGTTTATTCTGAGCACCGGGAAGGGGGCGGGCGGGGCGCCTCACT
GGGTGTTGAGCTTCTTAGTGTACTCCTCGAGAGCGCTCAGGAAGCTGACCTTGAAGCTCTCCAGCACGGG
CAGCAGGCCTTGGCGGAGTCTCGAGCGGGCTTGGCCTTCTCGCTGAGCGTGCTCAGATGCTCGGTG
GCCTTGGCGTGGTACTCGGCCAGTCTGGCGCCCGCTTCTCCTTGAAGGCC
```

Figura 4.14. Composición de las secuencias de SNP en formato FASTA



The screenshot displays the NCBI Sequence Viewer interface. At the top, there is a search bar with 'Nucleotide' selected and a 'Go' button. Below the search bar, there are options for 'Display' (Graphics) and 'Show' (1). A message indicates that the record for NT\_033899[gi:29805597] has been replaced or removed. The main view shows a genomic track with a highlighted region from approximately 49K to 50K. The track includes a scale from 10M to 38488550 and a legend for segment boundaries, region/SNP, and other features. The sequence is shown below the track, with two dbSNPs highlighted: T/C dbSNP:4754762 and C/A dbSNP:7937263.

Legend:

- segment boundaries
- region/SNP
- other feature
- sequence fragment shown

Sequence:

```
49001 TTTATTTAGA AAATCAGTC AAGGAGAGG CAGACACAGC ATTGAGATTA ACTGTTGCAG
      T/C dbSNP:4754762
49061 CATTTACCAC TATCAGTAGC ACATTTTCAG CATTAGTATA CCCTGTTTCG TTTTCAGATT
```

Figura 4.15. Interfaz de la base de datos para la documentación de polimorfismos

Las secuencias y sus descriptores se han guardado en la base de datos y al mismo tiempo se han generado los ficheros de texto para cada secuencia. Los nombres de los ficheros se han construido en la forma indicada en la parte de metodología:

```
“GID” & < GenBank ID > & “_” & <dbSNP id> & “_” & <SNP contig  
position> & “.snp”
```

Finalmente, se procedió al empaquetamiento de base de datos y de los ficheros de secuencias.

#### 4.2.4 Manejo y compresión de secuencias (Squeezer)

Se describen en esta subsección los resultados del desarrollo de unas utilidades de compresión de secuencias de nucleótidos y visualización de ficheros de grandes dimensiones con dichas secuencias. Las utilidades de compresión (y descompresión) se han denominado “Squeezer2B” y “Squeezer3B”, cuya justificación va implícita en la descripción de su funcionamiento que viene a continuación.

##### 4.2.4.1 Las utilidades de compresión

Se trata de dos programas implementados mediante el lenguaje de programación Java, que permiten comprimir y descomprimir cadenas usando el formato de compresión BSD (*binary squeezed dna*) y basados en la utilización de grupos de bits en lugar de bytes. Actualmente las cadenas de ADN y ARN en ficheros de texto convencionales se almacenan empleando caracteres de 8 bits para codificar cada nucleótido. Teniendo en cuenta que sólo hay 4 tipos de nucleótidos, “A”, “C”, “G” y “T” para el caso de ADN, y “A”, “C”, “G” y “U” para el caso de ARN, emplear 8 bits por nucleótido es un uso ineficiente de espacio cuando con dos o tres bits es suficiente.

Con dos bits, sólo se puede almacenar cuatro elementos de información, suficientes para las cadenas de ADN o ARN. Sin embargo, con tres bits se puede almacenar hasta ocho elementos de información, lo cual resulta interesante a la hora de respetar ciertos caracteres de representación, como el espacio en blanco, el tabulador, el retorno de carro o el carácter de nueva línea, que hacen más fácil la lectura de las cadenas.

Estos programas toman una secuencia de ADN, o ARN, codificada con caracteres de ocho bits y la transforman en su equivalente de dos bits, lo que hace el Squeezer2B, o tres bits, lo que hace el Squeezer3B, para después aplicar el

algoritmo GZIP sobre el resultante codificado y, de esta manera, lograr altos ratios de compresión, en función de la homogeneidad de la secuencia a comprimir. Independientemente de si se emplea dos o tres bits, estos programas toman el fichero fuente por secciones de 8 MB. A cada sección se le aplica la transformación de dos o tres bits, reduciendo su tamaño significativamente. A la sección transformada se le aplica el algoritmo de compresión GZIP y se guarda en un archivo. De este modo, los archivos con este formato, denominado BSD, presentan el aspecto interno que se muestra en la Figura 4.16.

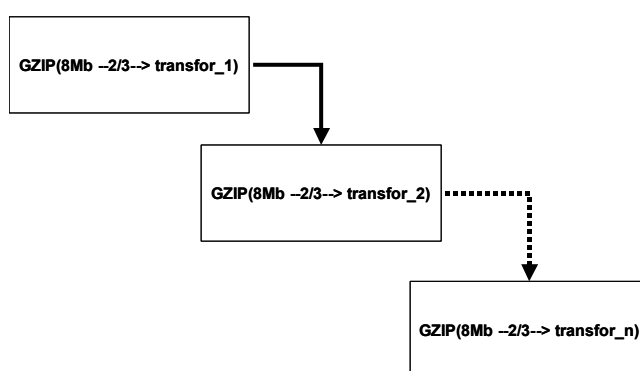


Figura 4.16. Estructura interna del formato BSD

Estos ficheros se componen así de bloques en formato zip de secuencias codificadas en dos o tres bits que al descomprimirse generan bloques de 8 Mb como máximo, teniendo en cuenta que el último bloque puede tener un tamaño menor.

En los Anexos 12.5, 12.6 y 12.7 se muestran los algoritmos principales y la estructura de datos básica de las herramientas Squeezer2B y Squeezer3B.

#### 4.2.4.2 Las interfaces

En la Figura 4.17 se muestra la interfaz de la utilidad Squeezer2B. La de Squeezer3B es idéntica salvo la cabecera que distingue una de otra.

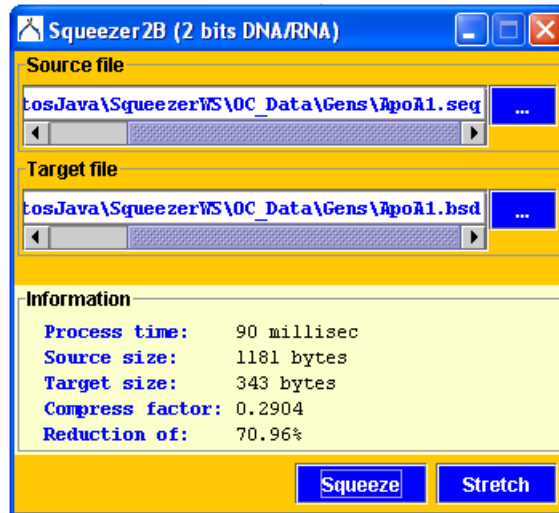


Figura 4.17. Interfaz de la utilidad Squeezer2B

Se puede ver que se permite la selección de la operación a realizar: compresión (*squeeze*) o descompresión (*stretch*). En función de la operación seleccionada, el fichero fuente será el fichero de secuencia a comprimir y el fichero destino será el fichero comprimido, para el caso de la compresión; o viceversa para el caso de la descompresión. En otro panel se muestran los resultados del proceso: tiempo de proceso, tamaño del fichero fuente (Size(source)), tamaño del fichero destino (Size(target)), factor de compresión y tasa de reducción. Las ecuaciones son las siguientes:

$$\text{factor de compresión} = \frac{\text{Size(target)}}{\text{Size(source)}} \quad (\text{E: 4.11})$$

$$\text{tasa de reducción} = \frac{|\text{Size(source)} - \text{Size(target)}|}{\text{Size(source)}} \times 100 \quad (\text{E: 4.12})$$

Estos parámetros son válidos en los dos tipos de operaciones.

#### 4.2.4.3 Ventajas y desventajas

Los programas Squeezer2B y Squeezer3B sirven para procesar ficheros que sólo contengan código genético perteneciente a ADN o ARN y también son capaces de filtrar la primera cabecera FASTA presente en el archivo fuente. Sin embargo, no son capaces por el momento de filtrar las cabeceras que preceden a cada una de las siguientes regiones. Resulta que como cualquiera de las cabeceras puede contener “T” o “U”, que se pueden confundir con los nucleótidos Timina o Uracilo, los programas fallan (no puede haber “T” en ARN ni “U” en ARN).

Se han comparado los resultados de estos dos programas con otro que implementa el algoritmo de Huffman y se han calculado las complejidades espaciales y temporales. En primer lugar, los cálculos arrojan una estimación de que la complejidad espacial es la siguiente:

- **Squeezer2B**: al aprovechar  $\frac{1}{4}$  de bit por cada base, frente a 8 bits del fichero texto convencional, la ocupación de espacio es de  $n/4$ . Al aplicarle el algoritmo gzip a los bytes, y en función de la variabilidad del resultado del primer paso, la ocupación de espacio puede ir desde  $(n/4 * 0,10)$  a  $(n/4 * 0,90)$ . En la descompresión se obtienen los valores inversos. La tasa de compresión va del 70 al 90%.
- **Squeezer3B**: al aprovechar  $\frac{3}{8}$  de bit por cada base, frente a 8 bits del fichero texto convencional, la ocupación de espacio es de  $3n/8$ . Al aplicarle el algoritmo gzip a los bytes, y en función de la variabilidad del resultado del primer paso, la ocupación de espacio puede ir desde  $(3n/8 * 0,10)$  a  $(3n/8 * 0,90)$ . En la descompresión se obtienen los valores inversos. En todo caso, como la compresión se hace por bloque del orden de MB, la pérdida de bits por byte es módulo de 7 por en número de bloques que se han leído del fichero. La tasa de compresión va del 70 al 90%.
- **Huffman**: la ocupación de espacio es aproximadamente  $2 \log n$ . En cadenas cortas, el fichero resultante de la compresión es mucho mayor que en los dos casos anteriores. En ficheros grandes, la tasa de compresión es similar a Squeezer#B.

Y la complejidad temporal es la siguiente:

- **Squeezer2B**: es lineal y proporcional a  $n/4$ . En la descompresión se obtienen valores similares.
- **Squeezer3B**: es lineal y proporcional a  $3n/8$ . En la descompresión se obtienen valores similares.



- **Huffman**: en la compresión, puesto que realiza varias etapas (recorrido bloque, confección árbol, codificación diccionario, compresión) el tiempo es dos órdenes por encima de Squeezer#B. En la descompresión, los tiempos son menores que para la misma operación en Squeezer#B.

Por lo tanto, la utilidad de ambas estrategias es parcial según la aplicación que se le quiera dar. Si se trata de manejar secuencias de bases sin caracteres distintos de los cuatro nucleótidos, la propuesta de esta tesis puede ser válida. En cambio, para manejar ficheros FASTA, puede ser más útil Huffman. Sin embargo, cuando se trata de tener procesos rápidos en la compresión, Huffman no es muy adecuado por su lentitud.

De todas formas, esta propuesta sigue en estudio y se ha previsto diseñar un formato de ficheros que permita el proceso lineal de las secuencias sin perder las cabeceras basado en XML.

### 4.3 Análisis estadístico de los resultados

En esta sección se describen los resultados obtenidos en los trabajos sobre los seis genes: CETP, APOE, APOA1, LIPC, SR-BI y PLIN. En cada una de las correspondientes subsecciones se da la referencia de la revista y del artículo en el que fueron publicados. Como se ha planteado en el Capítulo 2 al expresar los objetivos generales y particulares, esta sección describe los resultados correspondientes a la parte de Epidemiología genética.

#### 4.3.1 Resultados del análisis del gen CETP

Estos resultados se han publicado en la revista *Arteriosclerosis, Thrombosis and Cardiovascular Diseases* en el año 2001 con el título “Association of The Ceta Polymorphism with Variations in Lipoprotein Subclasses and Coronary Heart Disease Risk: The Framingham Study” (Ordovás et al., 2000). A continuación se presenta un resumen de los principales hallazgos.

Se analizaron un total de 2.876 individuos (1.411 hombres y 1.505 mujeres) participantes en el *Framingham Offspring Study* y que tenían datos de lípidos válidos y que, al mismo tiempo, no estaban tomando medicación que alterara las concentraciones de lípidos. En la Tabla 4.7 se presenta un resumen de las características demográficas, genotípicas y bioquímicas de los participantes según el género. La media de edad para los sujetos incluidos fue de 51,6 y 51,2 respectivamente. Aunque una proporción similar de hombres y mujeres eran

fumadores (23,4% y 22,8%, respectivamente), los hombres fumaban más cigarrillos por día ( $5,8 \pm 12,5$ ) que las mujeres ( $4,7 \pm 10,3$ ;  $p < 0,016$ ), y algo más de la mitad de las mujeres participantes (54,2%) eran posmenopáusicas.

**Tabla 4.7. Características demográficas, genotípicas y bioquímicas de los participantes del FOS según su sexo**

|                                | Hombres         | Mujeres         | P                                   |
|--------------------------------|-----------------|-----------------|-------------------------------------|
|                                | (n=1411)        | (n=1505)        | Comparación entre hombres y mujeres |
| Genotipo TaqIB CETP            |                 |                 |                                     |
| B1B1 (%)                       | 428 (30,3)      | 477 (31,7)      |                                     |
| B1B2 (%)                       | 713 (50,6)      | 754 (50,1)      |                                     |
| B2B2 (%)                       | 270 (19,1)      | 274 (18,2)      |                                     |
| Frecuencia del alelo B2        | 0,444           | 0,433           |                                     |
| Alelos de ApoE                 |                 |                 |                                     |
| E2 (%)                         | 12,0            | 14,7            |                                     |
| E3 (%)                         | 67,2            | 62,9            |                                     |
| E4 (%)                         | 20,8            | 22,4            |                                     |
| Edad (años)                    | $51,6 \pm 10,1$ | $51,2 \pm 9,7$  | 0,247                               |
| IMC (kg/m <sup>2</sup> )       | $27,6 \pm 3,9$  | $25,9 \pm 5,3$  | < 0,001                             |
| TC (mmol/L)                    | $5,28 \pm 0,96$ | $5,30 \pm 1,01$ | 0,394                               |
| LDL-C (mmol/L)                 | $3,47 \pm 0,85$ | $3,28 \pm 0,93$ | < 0,001                             |
| HDL-C (mmol/L)                 | $1,12 \pm 0,29$ | $1,45 \pm 0,39$ | < 0,001                             |
| HDL <sub>2</sub> -C (mmol/L)   | $0,13 \pm 0,10$ | $0,26 \pm 0,15$ | < 0,001                             |
| HDL <sub>3</sub> -C (mmol/L)   | $0,99 \pm 0,23$ | $1,20 \pm 0,28$ | < 0,001                             |
| TG (mmol/L)                    | $1,54 \pm 1,12$ | $1,23 \pm 1,14$ | < 0,001                             |
| ApoAI (g/L)                    | $1,34 \pm 0,24$ | $1,55 \pm 0,31$ | < 0,001                             |
| ApoB (g/L)                     | $1,02 \pm 0,24$ | $0,95 \pm 0,26$ | < 0,001                             |
| TC/HDL ratio                   | $5,00 \pm 1,50$ | $3,90 \pm 1,50$ | < 0,001                             |
| Glucosa (mmol/L)               | $5,41 \pm 1,48$ | $5,03 \pm 1,26$ | < 0,001                             |
| Alcohol (onzas/semana)         | $4,0 \pm 5,3$   | $1,8 \pm 2,9$   | < 0,001                             |
| Cigarrillos/día (en fumadores) | $5,8 \pm 12,5$  | $4,7 \pm 10,3$  | 0,016                               |
| Posmenopáusica (%)             | -               | 54,2            |                                     |
| Con estrógeno Rx* (%)          | -               | 12,9            |                                     |

No se detectaron diferencias significativas en las frecuencias del alelo B2 entre hombre y mujeres y la distribución de los alelos era consistente con el principio de equilibrio de Hardy-Weinberg. Los niveles de consumo de alcohol, IMC, LDL-C en plasma, apolipoproteína B total, triglicéridos y glucosa eran significativamente altos en los hombres comparados con las mujeres. Y las concentraciones totales de HDL-C, HDL<sub>2</sub>-C y HDL-C eran significativamente altas

en las mujeres participantes. La distribución del genotipo de APOE era similar en hombres y mujeres ( $P = 0,398$ ).

A continuación se describen los resultados con respecto a las asociaciones entre el polimorfismo TaqIB y la actividad de CETP, el tamaño de partículas de la lipoproteína y el riesgo cardiovascular.

#### **4.3.1.1 Asociación del polimorfismo TaqIB con las variaciones en los niveles de plasma de lípidos, lipoproteínas, apolipoproteínas y con la actividad de CETP**

En la Tabla 4.8 se puede ver que los tres grupos de fenotipos en hombres y mujeres eran equivalentes con respecto a la edad y el IMC. Los hombres homocigotos para el alelo B1 tenían niveles bajos de HDL-C ( $1,07 \pm 0,27$  mmol/L) en comparación con los individuos B1B2 ( $1,14 \pm 0,28$  mmol/L) y B2B2 ( $1,18 \pm 0,34$  mmol/L);  $p < 0,001$ . De la misma forma, las mujeres homocigotos para alelo B1 tenían niveles bajos de HDL-C ( $1,40 \pm 0,38$  mmol/L) en comparación con los individuos B1B2 ( $1,46 \pm 0,39$  mmol/L) y B2B2 ( $1,53 \pm 0,40$  mmol/L);  $p < 0,001$ . Se encontraron asociaciones similares para los valores de apolipoproteína A-I. Los niveles altos de HDL-C asociados con el alelo B2 eran debidos a los incrementos de las subfracciones de HDL2-C y HDL3-C.

También se detectó una asociación significativa entre el genotipo TaqIB y la actividad de CETP. Los hombres y mujeres portadores del alelo B2 tenían significativamente una baja actividad de CETP en comparación con los homocigotos para el alelo B1. En ambos sexos no se detectaron diferencias significativas entre los grupos de genotipos con respecto a los niveles en plasma de colesterol total, LDL-C y apolipoproteína B. Estos resultados se confirmaron mediante el enfoque de componentes de varianza y revelaron que el polimorfismo TaqIB influye en aproximadamente un 1% de la variabilidad de HDL-C.

Para validar la consistencia de la asociación entre el genotipo TaqIB de CETP y los niveles de HDL-C se aplicó un análisis de sensibilidad mediante un modelo de regresión lineal tal y como se ha descrito en la sección de Métodos, en el apartado correspondiente al gen CETP. La Figura 4.18 muestra los coeficientes de regresión y los intervalos de confianza al 95% para los respectivos genotipos B1B2 y B2B2 comparados con B1B1 cuando cada variable indicada se ha ido incluyendo en los modelos de regresión lineal (modelos 1 a 6). En primer lugar, se ha procedido a incluir solamente las variables mudas para el genotipo TaqIB (modelo 1). Este factor genético influyen en 1% en la variabilidad del HDL-C ( $p < 0,001$ ). Los coeficientes de regresión lineal iniciales para B1B2 y B2B2, después de haber controlado el efecto del sexo (modelo 2), han resultado ser respectivamente 0,06

(95%IC: 0,03-0,09) mmol/L,  $p < 0,001$ ; y 0,14 (95%IC: 0,09-0,18) mmol/L,  $p < 0,001$ .

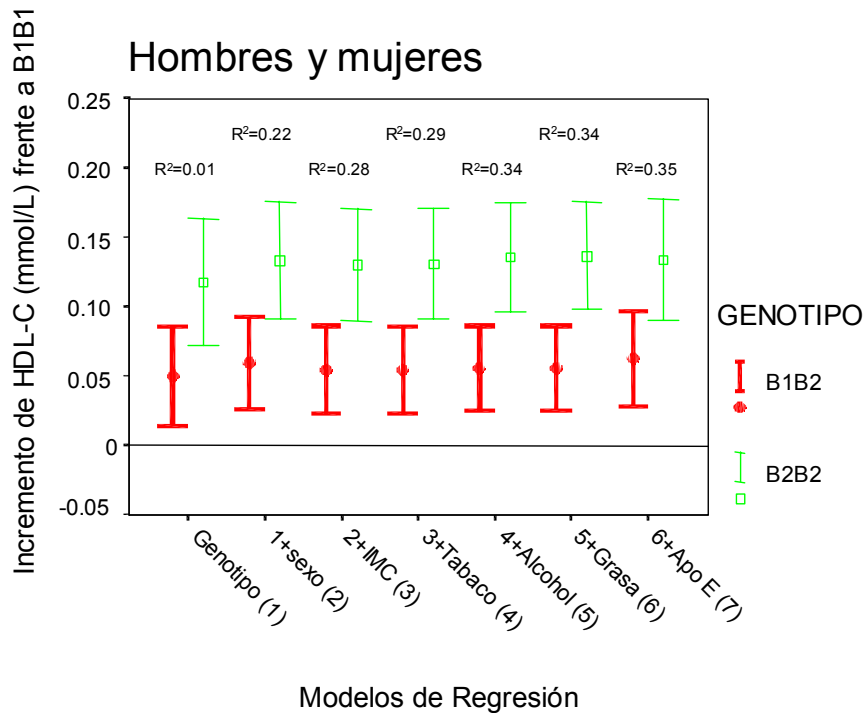
**Tabla 4.8. Niveles en plasma de lípidos, lipoproteínas y apolipoproteínas de los sujetos del FOS según los genotipos TaqIB CETP**

|                          | B1B1        | B1B2         | B2B2         | P*      | P†      |
|--------------------------|-------------|--------------|--------------|---------|---------|
| <b>Hombres</b>           |             |              |              |         |         |
| n                        | 428         | 713          | 270          |         |         |
| Edad (años)              | 51,2 ± 10,3 | 52 ± 10,0    | 51,3 ± 10,1  | 0,313   |         |
| IMC (kg/m <sup>2</sup> ) | 27,9 ± 4,0  | 27,50 ± 3,80 | 27,6 ± 3,8   | 0,169   |         |
| TC (mmol/L)              | 5,28 ± 0,93 | 5,25 ± 0,96  | 5,22 ± 0,96  | 0,639   | 0,889   |
| LDL-C (mmol/L)           | 3,49 ± 0,83 | 3,47 ± 0,88  | 3,41 ± 0,85  | 0,288   | 0,363   |
| HDL-C (mmol/L)           | 1,07 ± 0,27 | 1,14 ± 0,28‡ | 1,18 ± 0,34§ | < 0,001 | < 0,001 |
| HDL2-C (mmol/L)          | 0,12 ± 0,09 | 0,14 ± 0,10  | 0,15 ± 0,11§ | < 0,001 | 0,033   |
| HDL3-C (mmol/L)          | 0,95 ± 0,21 | 1,00 ± 0,22‡ | 1,03 ± 0,26§ | < 0,001 | < 0,001 |
| TG (mmol/L)              | 1,63 ± 1,16 | 1,52 ± 1,14  | 1,45 ± 0,95  | 0,059   | 0,098   |
| ApoAI (g/L)              | 1,32 ± 0,25 | 1,35 ± 0,23  | 1,37 ± 0,24§ | 0,017   | 0,025   |
| ApoB (g/L)               | 1,03 ± 0,25 | 1,02 ± 0,24  | 1,00 ± 0,25  | 0,135   | 0,662   |
| TC/HDL ratio             | 5,3 ± 1,5   | 4,9 ± 1,5‡   | 4,8 ± 1,6§   | < 0,001 | 0,011   |
| CETP (nM/L.hr)           | 160 ± 10,0  | 156 ± 10,0   | 139 ± 9,0    | 0,026   | 0,045   |
| <b>Mujeres</b>           |             |              |              |         |         |
| n                        | 477         | 754          | 274          |         |         |
| Edad (años)              | 51,2 ± 9,7  | 50,8 ± 9,4   | 51,3 ± 10,1  | 0,413   |         |
| IMC (kg/m <sup>2</sup> ) | 25,6 ± 5,4  | 25,8 ± 5,1   | 26,5 ± 5,5   | 0,081   |         |
| TC (mmol/L)              | 5,28 ± 0,98 | 5,30 ± 1,0   | 5,33 ± 1,03  | 0,901   | 0,794   |
| LDL-C (mmol/L)           | 3,34 ± 0,93 | 3,28 ± 0,9   | 3,23 ± 0,98  | 0,297   | 0,383   |
| HDL-C (mmol/L)           | 1,40 ± 0,38 | 1,46 ± 0,39‡ | 1,53 ± 0,40§ | < 0,001 | < 0,001 |
| HDL2-C (mmol/L)          | 0,24 ± 0,15 | 0,26 ± 0,1   | 0,28 ± 0,17§ | 0,008   | < 0,001 |
| HDL3-C (mmol/L)          | 1,16 ± 0,28 | 1,20 ± 0,3   | 1,25 ± 0,29§ | < 0,001 | < 0,001 |
| TG (mmol/L)              | 1,21 ± 0,86 | 1,24 ± 1,4   | 1,23 ± 0,84  | 0,834   | 0,646   |
| ApoAI (g/L)              | 1,32 ± 0,25 | 1,35 ± 0,2   | 1,37 ± 0,24  | 0,040   | 0,097   |
| ApoB (g/L)               | 0,95 ± 0,24 | 0,94 ± 0,3   | 0,95 ± 0,28  | 0,775   | 0,648   |
| TC/HDL ratio             | 4,0 ± 1,5   | 3,9 ± 1,5    | 3,7 ± 1,3§   | 0,006   | < 0,001 |
| CETP (nM/L.hr)           | 178 ± 11,0  | 159 ± 10,0‡  | 148 ± 11,0§  | < 0,001 | < 0,001 |

Los resultados se muestran como medias ± Desv. Std. Los valores P se han obtenido en la comparación entre genotipos CETP: \* Después de ajustar por las relaciones familiares. † Después de ajustar por las relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de beta bloqueantes, (estado menopáusico y terapia estrógena en mujeres) y APOE. Diferencias significativas entre los grupos B1B1 y B1B2 (‡); B1B1 y B2B2 (§); y B1B2 y B2B2 (□); después de ajustar por las relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de beta bloqueantes, (estado menopáusico y terapia estrógena en mujeres) y APOE.

Cuando se han ido incluyendo progresivamente a modelo nuclear otras variables: IMC, consumo de tabaco, consumo de alcohol y los genotipos de apoE, solamente se observó una ligera variación de los valores de los coeficientes de regresión estimados inicialmente, revelando una asociación independiente del polimorfismo TaqIB con los niveles de HDL-C con una fuerte consistencia sea cual

sea el factor adicional, ambiental o genético, a considerar. Así, el modelo final explicaba un 35% de la variabilidad del HDL-C en la población, y los coeficientes de regresión para B1B2 y B2B2 han resultado ser respectivamente de 0,07 (95%IC: 0,03-0,10) mmol/L y 0,14 (95% IC: 0,09-0,18) mmol/l, ( $p < 0,001$ ).



*Figura 4.18. Análisis de Sensibilidad. Coeficientes de regresión e intervalos de confianza al 95% para los genotipos B1B2 y B2B2 respectivamente, comparados con B1B1 cuando cada variable indicada se ha ido incluyendo progresivamente en los modelos de regresión lineal. Modelo 1: genotipo CETP. Modelo 2: modelo 1 + sexo. Modelo 3: modelo 2 + IMC. Modelo 4: modelo 3 + consumo de tabaco. Modelo 5: modelo 4 + consumo de alcohol. Modelo 6: modelo 5 + genotipo APOE. Se ha incluido R<sup>2</sup> en la figura para mostrar la variabilidad considerada para cada modelo de regresión*

Con el propósito de comprender mejor la base metabólica de la asociación de los niveles altos de HDL-C con el alelo B2 en hombre y mujeres, se han medido los perfiles de subclases de lipoproteínas mediante espectroscopia NMR automatizada. Como se muestra en la Tabla 4.9, esta asociación se debía específicamente al incremento significativo de las subfracciones de partículas grandes de HDL. Además, se ha observado una asociación significativa entre este

polimorfismo y las subfracciones de LDL en hombres. El alelo B2 se ha asociado con niveles incrementados de subfracciones de partículas grandes de LDL ( $1,17\pm 0,89$  y  $1,94\pm 0,88$  mmol/L para B1B2 y B2B2 respectivamente) en comparación con los sujetos B1B1 ( $1,64\pm 0,86$  mmol/L). En otras palabras, los hombres B1B1 han incrementado los niveles de las fracciones de partículas pequeñas de LDL ( $0,86\pm 0,65$  mmol/L) en comparación con B1B2 ( $0,79\pm 0,60$  mmol/L) y B2B2 ( $0,80\pm 0,65$  mmol/L) ( $p = 0,031$ ). En las mujeres se ha observado un efecto similar, pero sin significación estadística.

#### **4.3.1.2 Asociación del polimorfismo CETP TaqIB con las variaciones en el tamaño de las partículas de lipoproteínas**

También se ha investigado el efecto del polimorfismo de CETP sobre el tamaño de las partículas de lipoproteínas (Tabla 4.10). En los hombres, el alelo B2 se ha asociado con incrementos de tamaño de partícula de HDL y LDL después de haber ajustado por relaciones familiares, edad, IMC, tabaco, consumo de alcohol, uso de betabloqueantes y genotipo de APOE. En las mujeres, se ha demostrado un efecto similar para el tamaño de partículas de HDL después de haber ajustado por las variables indicadas anteriormente más el estado menopáusico y la terapia estrógena. Sin embargo, no se han observado diferencias para los tamaños de partículas de LDL.

#### **4.3.1.3 El genotipo CETP TaqIB y el riesgo de ECV**

En este conjunto, las ECV estaban presentes en 163 hombres y 62 mujeres. Debido al reducido número de casos en mujeres, sólo se ha comprobado en los hombres la asociación entre el polimorfismo TaqIB y el riesgo de ECV. Cuando se ha examinado la prevalencia de ECV en el ciclo 5 frente a la ausencia o presencia del alelo B2 mediante análisis de Chi-cuadrado, se ha demostrado una significativa ( $p = 0,035$ ) baja frecuencia de portadores del alelo B2 (58,7 vs. 70,6 %) entre los sujetos con ocurrencia de ECV. De forma similar, el odds ratio para ECV asociada con el alelo B2 ha sido 0,696 (95% IC: 0,50-0,98;  $p = 0,035$ ). Después de haber ajustado por edad, IMB, presión sanguínea sistólica, diabetes, tabaco, consumo de alcohol, uso de betabloqueantes y fármacos de control del colesterol, TC y HDL-C, este odds ratio ha resultado ser de 0,735 (95% IC: 0,46-1,162;  $p = 0,188$ ).

**Tabla 4.9. Distribuciones de las subclases de lipoproteínas de los sujetos del FOS según los genotipos de TaqIB CETP**

|                | B1B1        | B2B2        | B2B2        | P*      | P†      |
|----------------|-------------|-------------|-------------|---------|---------|
| <b>Hombres</b> |             |             |             |         |         |
| <b>VLDL</b>    |             |             |             |         |         |
| Grande         | 0,14 ± 0,27 | 0,11 ± 0,18 | 0,10 ± 0,16 | 0,247   | 0,248   |
| Intermedia     | 0,90 ± 0,78 | 0,83 ± 0,70 | 0,76 ± 0,64 | 0,226   | 0,516   |
| Pequeña        | 0,22 ± 0,15 | 0,23 ± 0,15 | 0,23 ± 0,15 | 0,189   | 0,128   |
| <b>LDL</b>     |             |             |             |         |         |
| Grande         | 1,64 ± 0,86 | 1,77 ± 0,89 | 1,94 ± 0,88 | < 0,001 | 0,012   |
| Intermedia     | 0,97 ± 0,63 | 0,95 ± 0,62 | 0,87 ± 0,68 | 0,070   | 0,095   |
| Pequeña        | 0,86 ± 0,65 | 0,79 ± 0,60 | 0,80 ± 0,65 | 0,004   | 0,031   |
| <b>HDL</b>     |             |             |             |         |         |
| Grande         | 0,31 ± 0,27 | 0,37 ± 0,29 | 0,45 ± 0,37 | < 0,001 | < 0,001 |
| Intermedia     | 0,56 ± 0,17 | 0,55 ± 0,17 | 0,54 ± 0,18 | 0,384   | 0,668   |
| Pequeña        | 0,22 ± 0,14 | 0,22 ± 0,14 | 0,22 ± 0,13 | 0,577   | 0,742   |
| <b>Mujeres</b> |             |             |             |         |         |
| <b>VLDL</b>    |             |             |             |         |         |
| Grande         | 0,06 ± 0,18 | 0,07 ± 0,21 | 0,07 ± 0,37 | 0,073   | 0,458   |
| Intermedia     | 0,56 ± 0,59 | 0,56 ± 0,64 | 0,57 ± 0,54 | 0,862   | 0,369   |
| Pequeña        | 0,24 ± 0,15 | 0,25 ± 0,15 | 0,24 ± 0,15 | 0,894   | 0,530   |
| <b>LDL</b>     |             |             |             |         |         |
| Grande         | 2,13 ± 0,89 | 2,15 ± 0,86 | 2,19 ± 0,86 | 0,648   | 0,172   |
| Intermedia     | 0,86 ± 0,65 | 0,79 ± 0,60 | 0,80 ± 0,65 | 0,104   | 0,093   |
| Pequeña        | 0,51 ± 0,59 | 0,48 ± 0,51 | 0,47 ± 0,46 | 0,395   | 0,610   |
| <b>HDL</b>     |             |             |             |         |         |
| Grande         | 0,76 ± 0,43 | 0,81 ± 0,42 | 0,87 ± 0,44 | < 0,001 | < 0,001 |
| Intermedia     | 0,54 ± 0,18 | 0,53 ± 0,18 | 0,54 ± 0,17 | 0,449   | 0,596   |
| Pequeña        | 0,14 ± 0,13 | 0,14 ± 0,12 | 0,14 ± 0,11 | 0,641   | 0,388   |

Los resultados se muestran como medias ± Desv. Std. Los valores P se han obtenido en la comparación entre genotipos CETP: \* Después de ajustar por las relaciones familiares. † Después de ajustar por las relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de beta bloqueantes, (estado menopáusico y terapia estrógena en mujeres) y APOE. Diferencias significativas (p < 0,05) entre los grupos B1B1 y B1B2 (‡); B1B1 y B2B2 (§); y B1B2 y B2B2 (¶); después de ajustar por las relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de beta bloqueantes, (estado menopáusico y terapia estrógena en mujeres) y APOE.

**Tabla 4.10. Diámetros de lipoproteínas (nm) de los sujetos del FOS según los genotipos del TaqIB CETP**

|                | B1B1          | B1B2          | B2B2          | P*      | P†      |
|----------------|---------------|---------------|---------------|---------|---------|
| <b>Hombres</b> |               |               |               |         |         |
| VLDL           | 49,12 ± 10,24 | 48,52 ± 9,23  | 47,34 ± 8,58  | 0,054   | 0,649   |
| LDL            | 20,56 ± 0,60  | 20,69 ± 0,58‡ | 20,8 ± 0,53§  | < 0,001 | < 0,001 |
| HDL            | 8,83 ± 0,37   | 8,92 ± 0,40‡  | 8,98 ± 0,45§  | < 0,001 | < 0,001 |
| <b>Mujeres</b> |               |               |               |         |         |
| VLDL           | 43,99 ± 8,59  | 44,11 ± 8,40  | 45,81 ± 8,89§ | 0,019   | 0,129   |
| LDL            | 21,05 ± 0,52  | 21,07 ± 0,46  | 21,09 ± 0,41  | 0,547   | 0,194   |
| HDL            | 9,35 ± 0,45   | 9,4 ± 0,43‡   | 9,44 ± 0,46§  | 0,027   | < 0,001 |

Los resultados se muestran como medias ± Desv. Std. Los valores P se han obtenido en la comparación entre genotipos CETP: \* Después de ajustar por las relaciones familiares. † Después de ajustar por las relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de beta bloqueantes, (estado menopáusico y terapia estrógena en mujeres) y APOE. Diferencias significativas (p < 0,05) entre los grupos B1B1 y B1B2 (‡); B1B1 y B2B2 (§); y B1B2 y B2B2 (□); después de ajustar por las relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de beta bloqueantes, (estado menopáusico y terapia estrógena en mujeres) y APOE.



### 4.3.2 Resultados del análisis del gen APOE

Estos resultados se han publicado en la revista *American Journal of Clinical Nutrition* en el año 2000 con el título “Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study” (Corella et al., 2001). A continuación se presenta un resumen de los principales hallazgos.

Se analizaron un total de 2.147 individuos (1.014 hombres y 1.133 mujeres) participantes en el *Framingham Offspring Study*. En la Tabla 4.11 se presenta un resumen de las características demográficas y antropométricas, y de los datos dietéticos y plasmáticos de los participantes según el género y el régimen de bebida. Todas las variables incluidas, a excepción de edad, concentraciones de LDL-C y consumo de tabaco, tienen diferencias significativas entre hombres y mujeres.

La Tabla 4.12 muestra las concentraciones en plasma de lípidos, consumo de alcohol, tabaco y grasa en dieta según el sexo para los portadores del alelo  $\epsilon 2$  ( $\epsilon 2/\epsilon 3$  y  $\epsilon 2/\epsilon 2$ ); los homocigotos para el alelo  $\epsilon 3$ ; y los portadores del alelo  $\epsilon 4$  ( $\epsilon 3/\epsilon 4$  y  $\epsilon 4/\epsilon 4$ ). Los sujetos portadores del genotipo  $\epsilon 2/\epsilon 4$  (15 hombres y 18 mujeres) se han excluido del análisis. El grupo APOE tenía un efecto significativo en las concentraciones de colesterol total y LDL en hombre y mujeres ( $p < 0,001$ ). Los sujetos  $\epsilon 2$  tenían más bajas concentraciones de TC y LDL-C que los sujetos  $\epsilon 3$  y  $\epsilon 4$ . Después de haber aplicado la corrección de Bonferroni, no se han observado diferencias significativas en las medias entre sujetos  $\epsilon 3$  y  $\epsilon 4$ . Los valores medios para la ingesta dietética de grasa total, grasa saturada, colesterol y alcohol no difieren significativamente entre los grupos de APOE.

Se han llevado a cabo correlaciones de Pearson entre la grasa total y saturada en dieta, colesterol, aporte de energía y consumo de alcohol (g/día) en hombre y mujeres por el genotipo de APOE. Habían coeficientes de correlación ligeramente positivos y significativos entre la grasa, el colesterol y el aporte de energía en hombres y mujeres, sin embargo no se apreciaron diferencias significativas en las ingestas entre los grupos de APOE (Tabla 4.13). El consumo de alcohol no mostraba asociaciones significativas con los nutrientes, y una asociación positiva laxa con el aporte de energía. El coeficiente de correlación más alto para el consumo de alcohol y el aporte de energía ( $r = 0,27$ ;  $p < 0,01$ ) se ha observado en los hombres con el alelo  $\epsilon 2$ .

Con el propósito de examinar las interacciones entre los subgrupos de APOE y las variables de dieta, se han categorizado los consumos de alcohol e ingesta de grasa saturada (altamente correlacionados con la ingesta de grasa total) como se describe en el correspondiente apartado de los métodos de análisis (3.4.5.2). El número de sujetos bebedores ( $p < 0,001$ ) era más alto en hombres (80,6%) que en mujeres (69,8%). Los grandes bebedores ( $>26,4$  g/día en hombres y  $>13,2$  g/día en mujeres) eran más frecuentes entre los hombres (21,2%) que entre las mujeres (17,5%) ( $p < 0,001$ ). El número de fumadores activos no era diferente en ambos sexos (19%). La ingesta de grasa saturada, expresada como porcentaje del aporte de energía, era más baja en mujeres que en hombres ( $p < 0,02$ ). En los hombres, se tenía, que un 41.2% consumían bajas cantidades ( $\leq 10\%$  de energía proveniente de la grasa saturada), un 45.7% cantidades intermedias (entre 10% y 14%) y un 13.1% altas cantidades ( $\geq 14\%$  de energía). En las mujeres, los valores correspondientes eran respectivamente de 45.7%, 43.6% y 10.7%.

**Tabla 4.11. Características demográficas, bioquímicas y dietéticas de la población (media y desviación estándar)**

|  | HOMBRES (n=1014)  |                      | MUJERES (n=1133)  |                      |
|--|-------------------|----------------------|-------------------|----------------------|
|  | Bebedores (n=817) | No bebedores (n=197) | Bebedores (n=791) | No bebedores (n=342) |
| Edad (años)                                  | 54,0 (9,7)        | 54,0 (10,1)          | 53,4 (9,5)        | 55,5 (9,8)*          |
| Índice de masa corporal (kg/m <sup>2</sup> ) | 28,1 (4,0)        | 28,1 (4,2)           | 26,1 (5,3)        | 27,5 (5,5)*          |
| Total-C (mmol/L)                             | 5,23 (0,87)       | 5,38 (0,97)*         | 5,33 (0,93)       | 5,38 (0,97)          |
| LDL-C (mmol/L)                               | 3,34 (0,79)       | 3,21 (0,80)*         | 3,17 (0,88)       | 3,28 (0,84)*         |
| HDL-C (mmol/L)                               | 1,17 (0,29)       | 1,01 (0,25)*         | 1,53 (0,39)       | 1,33 (0,35)*         |
| Triglicéridos (mmol/L)                       | 1,58 (0,83)       | 1,69 (0,91)          | 1,37 (0,71)       | 1,56 (0,75)*         |
| Grasa total (g/día)                          | 65,9 (26,0)       | 73,2 (30,1)*         | 56,7 (23,6)       | 56,3 (22,5)          |
| Grasa saturada (g/día)                       | 23,7 (10,3)       | 26,4 (11,5)*         | 20,1 (9,1)        | 20,2 (8,7)           |
| Colesterol (mg/día)                          | 242,0 (90,2)      | 205,5 (94,9)         | 211,4 (90,3)      | 205,5 (94,9)         |
| Aporte de energía (MJ/día)                   | 8,35 (2,64)       | 8,68 (2,79)          | 7,28 (2,40)       | 7,08 (2,37)          |
| Alcohol (g/día)                              | 19,0 (20,2)       | 0,0 (0,0)*           | 15,5 (13,0)       | 0,0 (0,0)*           |
| PSS (mmHg)                                   | 128 (16)          | 129 (19)             | 123 (19)          | 125 (21)             |
| PSD (mmHg)                                   | 82 (10)           | 81 (11)              | 79 (10)           | 77 (10)              |
| Fumadores actuales, n (%)                    | 146 (17,9%)       | 43 (21,8%)           | 150 (19,0%)       | 70 (21,5%)           |

\*: Significativamente diferentes en los bebedores (P < 0,05)

PSS: Presión sanguínea sistólica. PSD: Presión sanguínea diastólica

Aporte de energía expresados en mega Julios (MJ) por día. El factor de conversión es 1MJ=239.23 kcal

**Tabla 4.12. Concentraciones en plasma de lípidos, consumo de alcohol, tabaco y grasa en dieta según los genotipos de APOE y sexo**

|                              |         | E2                         |                 | E3                       |                 | E4                       |                 | P*      | P trend† |
|------------------------------|---------|----------------------------|-----------------|--------------------------|-----------------|--------------------------|-----------------|---------|----------|
|                              |         | hombres<br>n = 125 (12,3%) | n = 170 (15,0%) | n = 691 (68,1%)          | n = 736 (65,0%) | n = 198 (19,5%)          | n = 227 (20,0%) |         |          |
|                              |         | Mean (SD)                  |                 | Mean (SD)                |                 | Mean (SD)                |                 |         |          |
| Edad (años)                  | hombres | 54,1 (9,4)                 |                 | 54,2 (9,8)               |                 | 53,8 (9,9)               |                 | 0,856   | 0,783    |
|                              | mujeres | 54,4 (9,4)                 |                 | 54,1 (9,8)               |                 | 53,6 (9,1)               |                 | 0,671   | 0,399    |
| IMC (kg/m <sup>2</sup> )     | hombres | 28,6 (4,0)                 |                 | 28,0 (4,1)               |                 | 28,0 (4,0)               |                 | 0,856   | 0,783    |
|                              | mujeres | 26,3 (5,5)                 |                 | 26,4 (5,3)               |                 | 27,0 (5,8)               |                 | 0,671   | 0,399    |
| Total-C (mmol/L)             | hombres | 4,92 (0,96) <sup>b,c</sup> |                 | 5,22 (0,85) <sup>a</sup> |                 | 5,34 (0,97) <sup>a</sup> |                 | < 0,001 | < 0,001  |
|                              | mujeres | 5,09 (1,06) <sup>b,c</sup> |                 | 5,38 (0,95) <sup>a</sup> |                 | 5,49 (0,91) <sup>a</sup> |                 | < 0,001 | < 0,001  |
| LDL-C (mmol/L)               | hombres | 2,92 (0,86) <sup>b,c</sup> |                 | 3,37 (0,77) <sup>a</sup> |                 | 3,43 (0,83) <sup>a</sup> |                 | < 0,001 | < 0,001  |
|                              | mujeres | 2,88 (0,90) <sup>b,c</sup> |                 | 3,25 (0,85) <sup>a</sup> |                 | 3,35 (0,82) <sup>a</sup> |                 | < 0,001 | < 0,001  |
| HDL-C (mmol/L)               | hombres | 1,14 (0,32)                |                 | 1,14 (0,29)              |                 | 1,08 (0,28)              |                 | 0,043   | 0,066    |
|                              | mujeres | 1,51 (0,43)                |                 | 1,47 (0,39)              |                 | 1,43 (0,40)              |                 | 0,101   | 0,032    |
| Triglicéridos (mmol/L)       | hombres | 1,99 (1,43) <sup>b,c</sup> |                 | 1,71 (1,22) <sup>a</sup> |                 | 1,87 (1,31) <sup>a</sup> |                 | 0,014   | 0,438    |
|                              | mujeres | 1,51 (0,74)                |                 | 1,46 (0,87)              |                 | 1,61 (1,17)              |                 | 0,146   | 0,314    |
| Grasa total (g/día)          | hombres | 70,7 (27,0)                |                 | 66,7 (27,0)              |                 | 67,8 (28,0)              |                 | 0,293   | 0,329    |
|                              | mujeres | 57,4 (22,5)                |                 | 56,7 (23,3)              |                 | 55,9 (23,7)              |                 | 0,815   | 0,526    |
| Grasa saturada (g/día)       | hombres | 25,8 (10,8)                |                 | 24,0 (10,5)              |                 | 24,0 (10,9)              |                 | 0,212   | 0,137    |
|                              | mujeres | 20,0 (8,5)                 |                 | 20,2 (9,0)               |                 | 20,0 (9,5)               |                 | 0,950   | 0,980    |
| Colesterol en dieta (mg/día) | hombres | 250,9 (93,0)               |                 | 243,7 (116,8)            |                 | 249,3 (113,5)            |                 | 0,709   | 0,895    |
|                              | mujeres | 218,2 (101,3)              |                 | 209,5 (91,3)             |                 | 203,4 (85,0)             |                 | 0,279   | 0,110    |
| Aporte de energía (MJ/día)   | hombres | 8,7 (2,8)                  |                 | 8,4 (2,7)                |                 | 8,4 (2,7)                |                 | 0,626   | 0,438    |
|                              | mujeres | 7,3 (2,3)                  |                 | 7,3 (2,4)                |                 | 7,1 (2,4)                |                 | 0,508   | 0,286    |
| Alcohol (g/día)              | hombres | 13,3 (17,9)                |                 | 15,6 (19,9)              |                 | 16,0 (20,5)              |                 | 0,398   | 0,551    |
|                              | mujeres | 8,1 (12,3)                 |                 | 6,7 (11,3)               |                 | 8,7 (13,3)               |                 | 0,063   | 0,603    |
| SBP (mmHg)                   | hombres | 127 (16)                   |                 | 128 (17)                 |                 | 128 (17)                 |                 | 0,743   | 0,441    |
|                              | mujeres | 122 (18)                   |                 | 124 (20)                 |                 | 124 (20)                 |                 | 0,391   | 0,230    |
| DBP (mmHg)                   | hombres | 82 (10)                    |                 | 81 (10)                  |                 | 82 (10)                  |                 | 0,414   | 0,542    |
|                              | mujeres | 76 (10) <sup>c</sup>       |                 | 77 (10)                  |                 | 79 (10) <sup>a</sup>     |                 | 0,022   | 0,010    |
| No bebedores, n (%)          | hombres | 20 (16,0%)                 |                 | 129 (18,7%)              |                 | 48 (24,2%)               |                 | 0,127   | 0,135    |
|                              | mujeres | 46 (27,1%)                 |                 | 241 (32,7%)              |                 | 55 (24,4%)               |                 | 0,087   | 0,363    |
| Fumadores actuales, n (%)    | hombres | 24 (19,2%)                 |                 | 127 (18,4%)              |                 | 38 (19,2%)               |                 | 0,955   | 0,677    |
|                              | mujeres | 30 (17,6%)                 |                 | 142 (19,3%)              |                 | 48 (21,1%)               |                 | 0,677   | 0,378    |

Para propósitos de análisis, los genotipos 2/2 y 2/3 se agruparon como E2. Los genotipos 3/4 y 4/4, se agruparon como E4. El genotipo 3/3, como E3. Se han excluido del análisis los casos con genotipo 2/4.

SD: Desviación estándar. Aporte de energía expresado en mega Julios (MJ) por día. El factor de conversión es 1MJ=239.23 kcal

\*: P obtenido en los tests ANOVA o Chi-cuadrado para la comparación global de medias entre los genotipos de APOE

†: P para tendencias lineales entre categorías calculado también con ANOVA. a, b, c: P obtenidas con el ajuste de Bonferroni para múltiples comparaciones: a: P < 0.05 comparada con el genotipo E2, b: P < 0.05 comparada con el genotipo E3, c: P < 0.05 comparada con el genotipo E4

**Tabla 4.13. Coeficientes de correlación de Pearson entre grasa de la dieta y el aporte de energía y el consumo de alcohol por sexo y genotipo de APOE**

|                                   | HOMBRES (n=1014) |                |               |                   | MUJERES (n=1133) |                |               |                   |
|-----------------------------------|------------------|----------------|---------------|-------------------|------------------|----------------|---------------|-------------------|
|                                   | Grasa total      | Grasa saturada | Colesterol    | Aporte de energía | Grasa total      | Grasa saturada | Colesterol    | Aporte de energía |
| <b>Grasa saturada (g/día)</b>     | <b>0.947*</b>    |                |               |                   | <b>0.947*</b>    |                |               |                   |
| E2                                | 0.905*           |                |               |                   | 0.958*           |                |               |                   |
| E3                                | 0.944*           |                |               |                   | 0.949*           |                |               |                   |
| E4                                | 0.941*           |                |               |                   | 0.943*           |                |               |                   |
| <b>Colesterol (mg/día)</b>        | <b>0.718*</b>    | <b>0.733*</b>  |               |                   | <b>0.738*</b>    | <b>0.741*</b>  |               |                   |
| E2                                | 0.804*           | 0.830*         |               |                   | 0.721*           | 0.715*         |               |                   |
| E3                                | 0.702*           | 0.723*         |               |                   | 0.752*           | 0.755*         |               |                   |
| E4                                | 0.712*           | 0.735*         |               |                   | 0.780*           | 0.766*         |               |                   |
| <b>Aporte de energía (MJ/día)</b> | <b>0.846*</b>    | <b>0.784*</b>  | <b>0.628*</b> |                   | <b>0.838*</b>    | <b>0.770*</b>  | <b>0.648*</b> |                   |
| E2                                | 0.845*           | 0.780*         | 0.688*        |                   | 0.843*           | 0.783*         | 0.703*        |                   |
| E3                                | 0.847*           | 0.784*         | 0.607*        |                   | 0.837*           | 0.766*         | 0.643*        |                   |
| E4                                | 0.872*           | 0.815*         | 0.638*        |                   | 0.864*           | 0.806*         | 0.696*        |                   |
| <b>Alcohol (g/día)</b>            | <b>-0.062†</b>   | <b>-0.055</b>  | <b>-0.019</b> | <b>0.102*</b>     | <b>0.004</b>     | <b>-0.014</b>  | <b>0.012</b>  | <b>0.082*</b>     |
| E2                                | 0.111            | 0.097          | 0.103         | 0.268*            | 0.007            | -0.003         | 0.038         | 0.083             |
| E3                                | -0.038           | -0.025         | 0.003         | 0.152*            | 0.014            | 0.004          | 0.006         | 0.122*            |
| E4                                | -0.094           | -0.090         | -0.028        | 0.085             | -0.001           | -0.009         | -0.038        | -0.033            |

\*: P < 0.01 ;†: P < 0.05

A continuación se describen los resultados más destacados con respecto a las interacciones de consumo de alcohol, alelos de APOE y concentraciones de LDL-C.

En los hombres, las medias de las concentraciones de LDL-C eran más altas en bebedores ( $3,34 \pm 0,79$  mmol/L) que en no bebedores ( $3,21 \pm 0,80$  mmol/L) ( $p = 0,018$ ). Por el contrario, en las mujeres estos valores eran más altos en no bebedores ( $3,28 \pm 0,84$  mmol/L) que en bebedores ( $3,17 \pm 0,88$  mmol/L) ( $p = 0,021$ ). Se ha evaluado si había diferencias en las concentraciones de LDL-C por genotipos de APOE con las categorías de consumo de alcohol (bebedores y no bebedores).

La Figura 4.19 muestra las concentraciones de LDL-C medias según los grupos de APOE y el consumo de alcohol en hombres (A) y en mujeres (B). Para hombres no bebedores, las concentraciones de LDL-C medias no difieren estadísticamente ( $p = 0,321$ ) entre los grupos de APOE. El efecto decreciente esperado del alelo  $\epsilon 2$  estaba ausente y el efecto creciente usual del alelo  $\epsilon 4$  estaba invertido. En los bebedores masculinos se observaron diferencias estadísticamente significativas ( $p < 0,001$ ) en LDL-C, con los sujetos  $\epsilon 2$  mostrando las concentraciones más bajas y los sujetos  $\epsilon 4$  mostrando las concentraciones más altas.

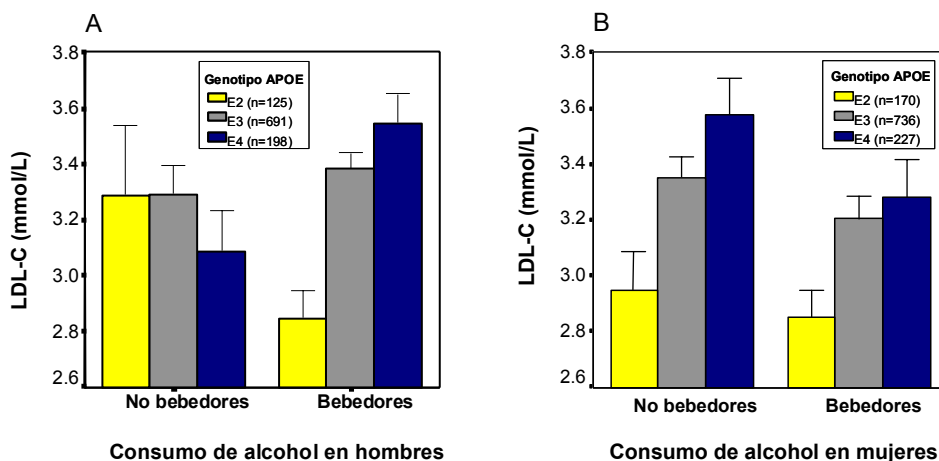


Figura 4.19. Concentraciones de LDL-C medias según los genotipos de APOE y el consumo de alcohol en hombres (A) y en mujeres (B). Los valores de  $P$  se han obtenido en los tests ANOVA para la comparación de las medias entre genotipos por consumo de alcohol. Las barras de error muestran el error estándar de las medias

Cuando se han comparado las concentraciones de LDL-C entre los subgrupos de APOE por tipos de bebedores, se ha observado que las concentraciones de LDL-C en los bebedores  $\epsilon 2$  eran significativamente más bajas que en los no bebedores  $\epsilon 2$  (2,84 mmol/L vs. 3,28 mmol/L;  $p < 0,05$ ). Por el contrario, en los sujetos  $\epsilon 4$  las concentraciones de LDL-C medias eran más altas en bebedores que en no bebedores (3,54 mmol/L vs. 3,09 mmol/L;  $p < 0,001$ ). En los sujetos  $\epsilon 3$  las concentraciones de LDL-C medias no diferían significativamente según el consumo de alcohol (3,29 mmol/L en no bebedores y 3,38 mmol/L en bebedores;  $p = 0,63$ ). En las mujeres el efecto esperado de los alelos de APOE sobre las concentraciones de LDL-C estaba presente en bebedores ( $p < 0,001$ ) y no bebedores ( $p < 0,001$ ). Se han obtenido similares resultados cuando las comparaciones de las medias de LDL-C se estratificaron por grupos de edad (menores y mayores de 47 años).

Con el propósito de examinar si las asociaciones observadas entre los alelos de APOE y el consumo de alcohol estaban presentes después de haber controlado por posibles factores de confusión, se han utilizado modelos de regresión multivariada. En primer lugar, para comprobar la hipótesis nula de que no hay interacciones entre el genotipo de APOE y los factores ambientales, se ha empleado un análisis de covarianza con la concentración de LDL-C como variable de salida; el consumo de alcohol (2 categorías) y los grupos de APOE (3 categorías) como factores fijos (efectos principales e interacciones de dos vías); la edad, el IMC y el aporte total de energía como covariantes continuas; y la ingesta de grasa saturada (3 categorías) y el consumo de tabaco (2 categorías) con factores de control.

En segundo lugar, para estimar los coeficientes de regresión se han ajustado análisis de regresión lineal múltiple con las mismas variables. Para calcular los coeficientes de regresión para los alelos de APOE, se ha considerado respectivamente el alelo  $\epsilon 3$ , los bebedores y el consumo de grasa saturada más alto como grupos de referencia. La Tabla 4.14 muestra el resumen de estos modelos ajustados para hombres y mujeres por separado. Se ha observado una altamente significativa interacción ( $p = 0,001$ ) entre el consumo de alcohol y el efecto de APOE en las concentraciones de LDL-C en hombres. Este término de interacción no era estadísticamente significativa. Estos modelos han confirmado el efecto observado con las comparaciones de las medias (Figura 1). El término de interacción entre alcohol y APOE permanecía estadísticamente significativo en hombres después de haber controlado por edad, IMC, grasa saturada, aporte de energía y consumo de tabaco. Los coeficientes de regresión para los términos de interacción indicaban que el efecto decreciente del alelo  $\epsilon 2$  sobre el LDL-C ( $-0,532$  mmol/L;  $p < 0,001$ , Tabla 4.14) estaba ausente en bebedores masculinos ( $+0,538$  mmol/L;  $p < 0,01$ ). De la misma forma, el usual efecto del alelo  $\epsilon 4$  sobre LDL-C ( $+0,157$  mmol/L;  $p < 0,05$ ) estaba también ausente en bebedores

masculinos ( $-0,339$  mmol/L;  $p < 0,05$ ). La interacción descrita entre el alcohol y el genotipo de APOE estaba presente en todas las categorías de ingesta de grasa saturada. El término de interacción entre APOE y grasa saturada no era significativamente estadístico cuando se ha incluido en estos modelos, por lo que se ha descartado en subsiguientes análisis.

Estos efectos se han analizado posteriormente para tener en cuenta el nivel de alcohol consumido. Para este propósito, se han dividido los bebedores en las categorías de moderados y grandes bebedores, según el criterio descrito en el apartado de métodos. La Tabla 4.15 muestra las concentraciones de LDL-C medias en hombres y mujeres por grupos de APOE y niveles de alcohol (sin consumo, moderado y alto). Los efectos de APOE no eran diferentes para los niveles de consumo de alcohol moderado y alto. Para probar la presencia de interacciones se han ajustado modelos de regresión lineal, como se ha descrito antes, con la variable de alcohol de tres categorías y controlando por edad, IMC, tabaco e ingesta de grasa por dieta. En los hombres, el término de interacción entre APOE y nivel de consumo de alcohol permanecía estadísticamente significativo ( $p = 0,004$ ). Sin embargo, sólo las comparaciones de no bebedores con consumo moderado o alto eran estadísticamente significativas y no se observaron diferencias entre moderados y grandes bebedores. En las mujeres, el término de interacción entre el grupo de APOE y alcohol no era estadísticamente significativo ( $p = 0,53$ ).

Finalmente, para comprobar la asociación entre el consumo de alcohol como una variable continua (g/día) con LDL-C, se han calculado modelos de regresión lineal múltiple estratificando por sexo y APOE, y controlando por edad, IMC, tabaco, grasa y energía. La Tabla 4.16 muestra los coeficientes de regresión parciales entre el consumo de alcohol (raíz cuadrada de g/día) y concentraciones de LDL-C. En los hombres portadores del alelo  $\epsilon 2$ , se ha encontrado una asociación negativa y estadísticamente significativa ( $r = -0,279$ ,  $B = -0,104$ ;  $p = 0,002$ ) entre la ingesta de alcohol y las concentraciones de LDL-C. En los hombres portadores del alelo  $\epsilon 4$ , se ha encontrado una asociación positiva y estadísticamente significativa ( $r = +0,198$ ,  $B = +0,064$ ;  $p = 0,008$ ) entre la ingesta de alcohol y las concentraciones de LDL-C. Las asociaciones estadísticas en hombres y mujeres  $\epsilon 3$  no eran significativas. Cuando se han excluido los no bebedores en los análisis previos solamente la asociación negativa de consumo de alcohol y concentraciones de LDL-C observada para los hombres  $\epsilon 2$  permanecía estadísticamente significativa ( $r = -0,214$ ,  $B = -0,082$ ;  $p < 0,05$ ).



**Tabla 4.14. Efecto del genotipo de APOE, alcohol, grasa saturada y la interacción APOE\*alcohol en el LDL-C del plasma . Analisis de regresión lineal por sexo**

|                                      | HOMBRES (n=1014) |                  |                   | MUJERES (n=1133) |                  |                   |
|--------------------------------------|------------------|------------------|-------------------|------------------|------------------|-------------------|
|                                      | B (SE)           | 95% CI           | p                 | B (SE)           | 95% CI           | p                 |
| <b>Genotipo APOE</b>                 |                  |                  | <b>0,036</b>      |                  |                  | <b>&lt; 0,001</b> |
| E3                                   | (referencia)     |                  |                   | (referencia)     |                  |                   |
| E2                                   | -0,532 (0,083)   | (-0,695; -0,369) | < 0,001           | -0,370 (0,080)   | (-0,527; -0,213) | < 0,001           |
| E4                                   | 0,157 (0,070)    | (0,020; 0,294)   | 0,028             | 0,066 (0,070)    | (-0,081; 0,193)  | 0,436             |
| <b>Ingesta de alcohol</b>            |                  |                  | <b>0,161</b>      |                  |                  | <b>0,104</b>      |
| Consumption                          | (referencia)     |                  |                   | (referencia)     |                  |                   |
| No consumption                       | -0,107 (0,075)   | (-0,254; 0,040)  | 0,161             | 0,066 (0,065)    | (-0,061; 0,193)  | 0,104             |
| <b>Ingesta de grasa saturada</b>     |                  |                  | <b>0,077</b>      |                  |                  | <b>0,593</b>      |
| >= 14 %                              | (referencia)     |                  |                   | (referencia)     |                  |                   |
| 10 - 14 %                            | -0,171 (0,078)   | (-0,324; -0,018) | 0,027             | -0,031 (0,083)   | (-0,194; 0,132)  | 0,713             |
| <= 10 %                              | -0,159 (0,075)   | (-0,306; -0,012) | 0,045             | -0,072 (0,083)   | (-0,235; 0,091)  | 0,389             |
| <b>Interacción APOE x Alcohol</b>    |                  |                  | <b>0,001</b>      |                  |                  | <b>0,474</b>      |
| Genotipo E3 x Consumo de alcohol     | (referencia)     |                  |                   | (referencia)     |                  |                   |
| Genotipo E2 x Sin consumo de alcohol | 0,538 (0,204)    | (0,138; 0,938)   | 0,009             | -0,036 (0,153)   | (-0,336; 0,264)  | 0,814             |
| Genotipo E4 x Sin consumo de alcohol | -0,339 (0,147)   | (-0,627; -0,051) | 0,024             | 0,160 (0,140)    | (-0,114; 0,434)  | 0,257             |
| <b>R<sup>2</sup> del modelo*</b>     | <b>0,071</b>     |                  | <b>&lt; 0,001</b> | <b>0,135</b>     |                  | <b>&lt; 0,001</b> |

Variable dependiente: LDL-C (mmol/L). B= Coeficientes de regresión. SE= Error estándar. CI= Intervalos de confianza  
 \*: Modelos ajustados adicionalmente por edad, IMC, tabaco y aporte de energía.

**Tabla 4.15. Concentraciones medias de LDL-C plasmático según el genotipo de APOE, estratificadas por el consumo de alcohol**

| Genotipos | HOMBRES    |                            |                            |                            | MUJERES    |                            |                            |                            |                            |      |
|-----------|------------|----------------------------|----------------------------|----------------------------|------------|----------------------------|----------------------------|----------------------------|----------------------------|------|
|           | N          | Sin consumo                |                            | Alto                       |            | N                          | Sin consumo                |                            | Alto                       |      |
|           |            | Media (SD)                 | Moderado                   | Media (SD)                 | Alto       |                            | Media (SD)                 | Moderado                   | Media (SD)                 | Alto |
| E2        | (n = 125)  | 2.92 (0.86) <sub>b,c</sub> | 2.88 (0.82) <sub>b,c</sub> | 2.68 (0.67) <sub>b,c</sub> | (n = 170)  | 2.88 (0.90) <sub>b,c</sub> | 2.95 (0.86) <sub>b,c</sub> | 2.90 (0.95) <sub>b,c</sub> | 2.71 (0.82) <sub>b,c</sub> |      |
| E3        | (n = 691)  | 3.37 (0.77) <sub>a</sub>   | 3.40 (0.78) <sub>a</sub>   | 3.35 (0.71) <sub>a</sub>   | (n = 736)  | 3.25 (0.85) <sub>a</sub>   | 3.35 (0.86) <sub>a</sub>   | 3.22 (0.83) <sub>a</sub>   | 3.18 (0.88) <sub>a</sub>   |      |
| E4        | (n = 198)  | 3.43 (0.83) <sub>a</sub>   | 3.56 (0.77) <sub>a</sub>   | 3.49 (0.92) <sub>a</sub>   | (n = 227)  | 3.35 (0.82) <sub>a</sub>   | 3.57 (0.91) <sub>a</sub>   | 3.23 (0.76) <sub>a</sub>   | 3.40 (0.84) <sub>a</sub>   |      |
| Total     | (n = 1014) | 3.33 (0.80)                | 3.35 (0.81)                | 3.33 (0.78)                | (n = 1133) | 3.22 (0.86)                | 3.33 (0.88)                | 3.17 (0.84)                | 3.16 (0.88)                |      |
| P*        |            | <0.001                     | 0,341                      | <0.001                     | P*         | <0.001                     | 0,001                      | 0,004                      | 0,002                      |      |

Valores de las medias expresados en mmol/L. SD: Desviación estándar

\*: P obtenido en los tests ANOVA para comparación global de las medias entre genotipos de APOE

a, b, c: P obtenidos con ajuste de Bonferroni para comparaciones múltiples:

a: P < 0.05 comparado con el genotipo E2

b: P < 0.05 comparado con el genotipo E3

c: P < 0.05 comparado con el genotipo E4

**Tabla 4.16. Asociación entre ingesta de alcohol y concentraciones de LDL-C plasmático por genotipo de APOE ajustado por edad, IMC, grasa aporte de energía y tabaco**

|                | n    | B      | SE    | r      | p     |
|----------------|------|--------|-------|--------|-------|
| <b>HOMBRES</b> |      |        |       |        |       |
| <b>Total</b>   | 1014 | 0,018  | 0,009 | 0,057  | 0,049 |
| E2             | 125  | -0,104 | 0,033 | -0,279 | 0,002 |
| E3             | 691  | 0,013  | 0,012 | 0,042  | 0,292 |
| E4             | 198  | 0,064  | 0,024 | 0,198  | 0,008 |
| <b>MUJERES</b> |      |        |       |        |       |
| <b>Total</b>   | 1133 | -0,012 | 0,012 | -0,026 | 0,299 |
| E2             | 170  | -0,012 | 0,034 | -0,026 | 0,732 |
| E3             | 736  | -0,026 | 0,017 | -0,056 | 0,122 |
| E4             | 227  | -0,013 | 0,027 | -0,031 | 0,631 |

Variable dependiente: LDL-C (mmol/L)

Coeficientes estimados por análisis de regresión lineal múltiple y de correlación parcial

B: coeficiente de regresión parcial para ingesta de alcohol (raíz cuadrada de g/día) controlado por edad, IMC, grasa, energía y tabaco

SE: Error estándar

r: coeficiente de correlación parcial controlado por edad, IMC, grasa, energía y tabaco

### 4.3.3 Resultados del análisis del gen APOA1

Estos resultados se han publicado en la revista *American Journal of Clinical Nutrition* en el año 2002 con el título “Polyunsaturated fatty acids modulate the effect of the APOA1-75(G/A) polymorphism on HDL-C levels in a specific fashion: The Framingham Study” (Ordovás et al., 2002). A continuación se presenta un resumen de los principales hallazgos.

Se analizaron un total de 1.597 individuos (755 hombres y 822 mujeres) participantes en el *Framingham Offspring Study*. En la Tabla 4.17 se presenta un resumen de los datos antropométricos, bioquímicos y dietéticos de los participantes. Todas las variables incluidas, a excepción de la edad, tienen diferencias significativas entre hombres y mujeres. La distribución del genotipo de

APOA1 no difiere para hombres y mujeres ( $p = 0,753$ ) y no se ha observado ninguna desviación del equilibrio de Hardy-Weinberg ( $\text{Chi-cuadrado} = 1,74$ ;  $p > 0,05$ ). Las frecuencias alélicas calculadas para este polimorfismo han sido de 0,835 (95% CI: 0,822-0,848) para el alelo G, y de 0,165 (95% CI: 0,152-0,178) para el alelo A.

**Tabla 4.17. Características demográficas, bioquímicas y dietéticas de la población**

|  | HOMBRES (n=755)   | MUJERES (n=822)   |
|--|-------------------|-------------------|
|  | Media (SD) o n(%) | Media (SD) o n(%) |
| Edad (años)                                  | 55,01 (9,38)      | 54,28 (9,52)      |
| Índice de masa corporal (kg/m <sup>2</sup> ) | 28,17 (4,07)      | 26,45 (5,36)*     |
| Total-C (mg/dL)                              | 201,20 (35,25)    | 207,56 (37,83)*   |
| LDL-C (mg/dL)                                | 128,44 (31,55)    | 124,70 (33,40)*   |
| HDL-C (mg/dL)                                | 43,26 (11,19)     | 56,61 (15,62)*    |
| Triglicéridos (mg/dL)                        | 157,84 (126,96)   | 131,22 (91,28)*   |
| Apo A-I (mg/dL)                              | 141,70 (21,19)    | 162,85 (25,81)*   |
| Apo B (mg/dL)                                | 116,80 (57,63)    | 108,72 (27,68)*   |
| Grasa total (g/día)                          | 67,97 (27,75)     | 56,54 (24,08)*    |
| Grasa saturada (g/día)                       | 24,39 (10,87)     | 20,16 (9,41)*     |
| Grasa monoinsaturada (g/día)                 | 26,17 (11,06)     | 21,27 (9,43)*     |
| Grasa poliinsaturada (g/día)                 | 12,79 (5,63)      | 11,46 (5,26)*     |
| Aporte de energía (Kcal/día)                 | 2.018,7 (656,8)   | 1.723,9 (570,2)*  |
| Alcohol (g/día)                              | 15,92 (20,10)     | 7,18 (11,75)*     |
| Alcohol en bebedores† (g/día)                | 19,56 (20,35)     | 10,12 (12,70)*    |
| Consumo de tabaco (cig/día) en fumadores‡    | 23,72 (12,78)     | 19,38 (10,40)*    |
| Genotipo Apo A-I (-75 bp)                    |                   |                   |
| G/G  | 530 (70,2%)       | 581 (70,7%)       |
| G/A  | 197 (26,1%)       | 214 (26,0%)       |
| A/A  | 28 (3,7%)         | 27 (3,3%)         |

Los valores se muestran como Media (desviación estándar, SD) o como número (n) y porcentaje (%) para genotipos

\*: Diferentes significativamente para los hombres ( $P < 0,05$ )

†: 80.0% en hombres y 73.3% en mujeres

‡: 20.9% en hombres y 18.8% en mujeres

SD: Desviación estándar. El aporte de energía se expresa en kilocalorías (Kcal) por día. El factor de conversión es 1 Kcal = 4,18 KJ (Joules)

Debido al pequeño tamaño de muestra del grupo A/A (28 hombres y 27 mujeres), se han combinado los genotipos G/A y A/A y se han comparado como los homocigotos G/G. La Tabla 4.18 muestra los niveles de lípidos plasmáticos, lipoproteínas y apolipoproteínas, y la ingesta de grasa en dieta según el sexo en los homocigotos G/G y en los portadores del alelo A.

**Tabla 4.18. Concentraciones en plasma de lípidos, lipoproteínas, apolipoproteínas, e ingesta de grasa en dieta según los genotipos de APOA1 y sexo**

|                                      |         | G/G        |                 | G/A + A/A       |                 | P*    |
|--------------------------------------|---------|------------|-----------------|-----------------|-----------------|-------|
|                                      |         | hombres    | n = 530 (70,2%) | n = 225 (29,8%) | n = 241 (29,3%) |       |
|                                      |         | Media (SD) |                 | Media (SD)      |                 |       |
| Edad (años)                          | hombres | 55,01      | (9,44)          | 55,02           | (9,25)          | 0,986 |
|                                      | mujeres | 54,28      | (9,56)          | 54,28           | (9,46)          | 0,995 |
| IMC (kg/m <sup>2</sup> )             | hombres | 28,13      | (4,06)          | 28,28           | (3,96)          | 0,650 |
|                                      | mujeres | 26,53      | (5,26)          | 26,26           | (5,28)          | 0,521 |
| Total-C (mg/dL)                      | hombres | 201,11     | (36,20)         | 201,40          | (33,27)         | 0,960 |
|                                      | mujeres | 207,12     | (38,57)         | 207,69          | (34,99)         | 0,842 |
| LDL-C (mg/dL)                        | hombres | 128,33     | (31,91)         | 128,70          | (30,75)         | 0,884 |
|                                      | mujeres | 124,49     | (34,16)         | 124,93          | (34,63)         | 0,865 |
| HDL-C (mg/dL)                        | hombres | 43,11      | (11,05)         | 43,58           | (11,72)         | 0,597 |
|                                      | mujeres | 56,55      | (15,56)         | 56,73           | (15,79)         | 0,881 |
| LDL-C/HDL-C ratio                    | hombres | 3,14       | (1,09)          | 3,12            | (1,07)          | 0,827 |
|                                      | mujeres | 2,37       | (0,95)          | 2,40            | (1,07)          | 0,655 |
| Apo A-I (mg/dL)                      | hombres | 141,32     | (20,67)         | 142,86          | (22,41)         | 0,402 |
|                                      | mujeres | 162,40     | (25,62)         | 163,81          | (26,28)         | 0,517 |
| Apo B (mg/dL)                        | hombres | 117,24     | (66,59)         | 115,74          | (25,24)         | 0,875 |
|                                      | mujeres | 108,77     | (28,01)         | 108,62          | (26,95)         | 0,623 |
| Triglicéridos (mg/dL)                | hombres | 157,92     | (123,19)        | 155,72          | (109,59)        | 0,943 |
|                                      | mujeres | 130,75     | (75,42)         | 131,70          | (101,05)        | 0,658 |
| Grasa saturada (% energía/día)       | hombres | 10,68      | (2,82)          | 11,11           | (3,20)          | 0,070 |
|                                      | mujeres | 10,32      | (2,78)          | 10,67           | (3,09)          | 0,187 |
| Grasa monoinsaturada (% energía/día) | hombres | 11,55      | (2,54)          | 11,76           | (2,67)          | 0,275 |
|                                      | mujeres | 10,91      | (2,67)          | 11,21           | (2,55)          | 0,116 |
| Grasa poliinsaturada (% energía/día) | hombres | 5,68       | (1,65)          | 5,79            | (1,98)          | 0,427 |
|                                      | mujeres | 5,92       | (1,67)          | 6,06            | (1,71)          | 0,326 |

SD: Desviación estándar.

\*: P obtenido en el test T de Student en la comparación entre los genotipos de APOA1 (-75 bp)

No se han observado diferencias para edad, IMC y para cualquier variable de lípidos, lipoproteínas y apolipoproteínas, entre los homocigotos G/G y los portadores del alelo A. En las mujeres, los niveles medios de HDL-C en G/G y G/A+A/A eran respectivamente de  $56,55 \pm 15,56$  mg/dL y  $56,73 \pm 15,79$  mg/dL, ( $p=0,881$ ). De la misma forma, los homocigotos G/G masculinos tenían los mismos niveles medios de HDL-C ( $43,58 \pm 11,17$  mg/dL) que los portadores de A ( $43,11 \pm 11,05$  mg/dL);  $p = 0,597$ . Se han obtenido similares resultados para los niveles de apoA-I. No difieren significativamente las medias entre los grupos de APOA1 para la ingesta en dieta de grasa total, grasa saturada (SAFA), grasa monoinsaturada (MUFA) y grasa poliinsaturada (PUFA).

A continuación se describen los resultados encontrados para las interacciones entre la ingesta en dieta de grasa, los alelos de APOA1 y los niveles de HDL-C; y entre la ingesta en dieta de grasa, los alelos de APOA1 y los niveles de ApoA-I.

#### 4.3.3.1 Ingesta en dieta de grasa, alelos de APOA1 y niveles de HDL-C

Para contrastar la hipótesis de que la ingesta en dieta de grasa puede afectar de formas diferentes los niveles de HDL-C en los portadores G y A, se han ajustado, separadamente para hombres y mujeres, varios modelos de regresión lineal multivariados, controlados por posibles factores de confusión, incluyendo un término de interacción para la ingesta de grasa en dieta. La Tabla 4.19 muestra los resultados para las mujeres.

En primer lugar, se ha ajustado un modelo básico (Modelo 1). Este modelo incluía HDL-C como variable de salida y el genotipo de APOA1 (G/G como referencia); y como variables predictivas, el consumo de alcohol (bebedores y no bebedores), consumo de tabaco (fumadores y no fumadores), edad, IMC e aporte total de energía. En el Modelo 1, el genotipo de APOA1 ajustado por los otros factores no estaba significativamente relacionado con los niveles de HDL-C ( $B = 0,28$  mg/dL para los portadores de A en comparación con los homocigotos G/G;  $p = 0,797$ ). No se han obtenido términos de interacción estadísticamente significativos en este modelo básico.

En el Modelo 2, se ha añadido al Modelo 1 la ingesta de grasa total (expresada como porcentaje de energía en continuo) y el término de interacción con el genotipo de APOA1. Este término de interacción estaba cerca de la significación estadística ( $p = 0,13$ ) y el genotipo de APOA1 incrementaba su coeficiente de regresión ( $B = -7,50$  mg/dl, para los portadores de A en comparación con los homocigotos G/G), pero sin alcanzar la significación estadística ( $p = 0,15$ ). En el Modelo 3, se ha añadido al Modelo 1 la ingesta de grasa saturada y el valor de  $p$  para el término de interacción no era significativo ( $p = 0,653$ ). En el Modelo 4, se

ha añadido al Modelo 1 la ingesta de grasa monoinsaturada (MUFA) y el valor de p para el término de interacción era de  $p = 0,128$ .

Finalmente, en el Modelo 5, se ha añadido al Modelo 1 la ingesta de grasa poliinsaturada (PUFA) y el valor de p para el término de interacción sí que era significativo ( $p = 0,005$ ). Según este último modelo las mujeres portadoras del alelo A tenían un decremento medio de  $-10,69 \pm (4,01)$  mg/dL en los niveles de HDL-C en comparación con los homocigotos G/G ( $p = 0,008$ ). Este efecto genético está influido por la ingesta de PUFA en continuo. La regresión predice que cada 1% de incremento de ingesta de PUFA provoca un incremento en HDL-C en 1,79 (95% CI: 0,54-3,04) mg/dL,  $p = 0,005$ , para las mujeres portadoras del alelo A en comparación con los homocigotos G/G.

**Tabla 4.19. Interacción del consumo de grasa con los efectos de los alelos de APOA1 (-75 bp) en los niveles de HDL-C, controlados por edad, IMC, consumo de alcohol y de tabaco. Se aplica análisis de regresión lineal múltiple en mujeres**

|  | Modelo básico (1) |        | (1)+Grasa total (Modelo 2) |        | (1)+SATFAT (Modelo 3) |        | (1)+MUFA (Modelo 4) |        | (1)+PUFA (Modelo 5) |        |
|--|-------------------|--------|----------------------------|--------|-----------------------|--------|---------------------|--------|---------------------|--------|
|  | B(SE)             | p      | B(SE)                      | p      | B(SE)                 | p      | B(SE)               | p      | B(SE)               | p      |
| Genotipo APOA1 (-75) (GA+AA vs G/G)          | 0,28 (1,10)       | 0,797  | -7,50 (5,11)               | 0,146  | -1,60 (4,02)          | 0,691  | -6,94 (4,79)        | 0,148  | -10,60 (4,01)       | 0,008  |
| Alcohol: No consumidores vs consumidores.    | -4,62 (1,10)      | <0,001 | -4,77 (1,10)               | <0,001 | -4,90 (1,10)          | <0,001 | -4,77 (1,11)        | <0,001 | -4,39 (1,10)        | <0,001 |
| Consumo de tabaco vs no fumadores            | -4,73 (1,23)      | <0,001 | -5,42 (1,29)               | <0,001 | -5,50 (1,30)          | <0,001 | -5,17 (1,30)        | <0,001 | -4,75 (1,30)        | <0,001 |
| IMC (Kg/m2)                                  | -1,10 (0,10)      | <0,001 | -1,14 (1,11)               | <0,001 | -1,13 (0,09)          | <0,001 | -1,13 (0,09)        | <0,001 | -1,10 (0,09)        | <0,001 |
| Grasa total (% energía)                      |                   |        | 0,16 (0,10)                | 0,099  |                       |        |                     |        |                     |        |
| Interacción Grasa total*APOA1 (GA+AA vs G/G) |                   |        | 0,26 (0,17)                | 0,130  |                       |        |                     |        |                     |        |
| SATFAT (% energía)                           |                   |        |                            |        | 0,48 (0,22)           | 0,028  |                     |        |                     |        |
| Interacción SATFAT*APOA1 (GA+AA vs G/G)      |                   |        |                            |        | 0,16 (0,37)           | 0,653  |                     |        |                     |        |
| MUFA (% energía)                             |                   |        |                            |        |                       |        | 0,27 (0,23)         | 0,220  |                     |        |
| Interacción MUFA*APOA1 (GA+AA vs G/G)        |                   |        |                            |        |                       |        | 0,64 (0,42)         | 0,128  |                     |        |
| PUFA (% energía)                             |                   |        |                            |        |                       |        |                     |        | -0,18 (0,35)        | 0,605  |
| Interacción PUFA*APOA1 (GA+AA vs G/G)        |                   |        |                            |        |                       |        |                     |        | 1,79 (0,64)         | 0,005  |
| R2 del modelo *                              | 0,17              | <0,001 | 0,19                       | <0,001 | 0,18                  | <0,001 | 0,18                | <0,001 | 0,19                | <0,001 |

Variable dependiente: HDL-C (mg/dL)

B = coeficiente de regresión. SE = error estándar

\*: Se ajustan adicionalmente los modelos por edad y aporte de energía.



**Tabla 4.20. Interacción del consumo de grasa con los efectos de los alelos de APOA1 (-75 bp) en los niveles de HDL-C, controlados por edad, IMC, consumo de alcohol y de tabaco. Se aplica análisis de regresión lineal múltiple en hombres**

|  | Modelo básico (1) |        | (1)+Grasa total (Modelo 2) |        | (1)+SATFAT (Modelo 3) |        | (1)+MUFA (Modelo 4) |        | (1)+PUFA (Modelo 5) |        |
|--|-------------------|--------|----------------------------|--------|-----------------------|--------|---------------------|--------|---------------------|--------|
|  | B(SE)             | p      | B(SE)                      | p      | B(SE)                 | p      | B(SE)               | p      | B(SE)               | p      |
| Genotipo APOA1 (-75) (GA+AA vs G/G)          | 0,64 (0,85)       | 0,452  | 2,97 (3,96)                | 0,454  | 2,77 (3,17)           | 0,382  | 3,42 (3,75)         | 0,362  | -0,02 (2,80)        | 0,994  |
| Alcohol:Nonconsumption vs consump.           | -6,20 (0,98)      | <0,001 | -6,60 (1,00)               | <0,001 | -6,71 (0,99)          | <0,001 | -6,60 (1,00)        | <0,001 | -6,23 (0,98)        | <0,001 |
| Consumo de tabaco vs no fumadores            | -1,10 (0,99)      | 0,275  | -1,25 (0,99)               | 0,210  | -1,50 (1,00)          | 0,140  | -1,25 (1,00)        | 0,212  | -1,14 (1,00)        | 0,254  |
| IMC (Kg/m2)                                  | -0,62 (0,09)      | <0,001 | -0,63 (0,10)               | <0,001 | -0,63 (0,10)          | <0,001 | -0,63 (0,10)        | <0,001 | -0,62 (0,09)        | <0,001 |
| Grasa total (% energía)                      |                   |        | 0,14 (0,08)                | 0,063  |                       |        |                     |        |                     |        |
| Interacción Grasa total*APOA1 (GA+AA vs G/G) |                   |        | -0,07 (0,13)               | 0,533  |                       |        |                     |        |                     |        |
| SATFAT (% energía)                           |                   |        |                            |        | 0,42 (0,17)           | 0,013  |                     |        |                     |        |
| Interacción SATFAT*APOA1 (GA+AA vs G/G)      |                   |        |                            |        | -0,21 (0,28)          | 0,456  |                     |        |                     |        |
| MUFA (% energía)                             |                   |        |                            |        |                       |        | 0,30 (0,19)         | 0,100  |                     |        |
| Interacción MUFA*APOA1 (GA+AA vs G/G)        |                   |        |                            |        |                       |        | -0,24 (0,31)        | 0,439  |                     |        |
| PUFA (% energía)                             |                   |        |                            |        |                       |        |                     |        | -0,17 (0,28)        | 0,410  |
| Interacción PUFA*APOA1 (GA+AA vs G/G)        |                   |        |                            |        |                       |        |                     |        | 0,12 (0,46)         | 0,799  |
| R2 del modelo *                              |                   |        |                            |        | 0,11                  | <0,001 | 0,11                | <0,001 | 0,10                | <0,001 |
|  |                   |        |                            |        |                       |        |                     |        | 0,10                | <0,001 |

Variable dependiente: HDL-C (mg/dL)

B = coeficiente de regresión. SE = error estándar

\*: Se ajustan adicionalmente los modelos por edad y aporte de energía.

**Tabla 4.21. HDL-C plasmático y niveles de apo A-I por los genotipos de APOA1, ingesta de grasa poliinsaturada y sexo**

|                 |       | <=4% PUFA      |         | 4%-8% PUFA     |         | >=8 % PUFA     |        | P†    | P‡    |
|-----------------|-------|----------------|---------|----------------|---------|----------------|--------|-------|-------|
|                 |       | n = 107        |         | n = 624        |         | n = 91         |        |       |       |
|                 |       | Media (SD)     | n = 118 | Media (SD)     | n = 585 | Media (SD)     | n = 52 |       |       |
| <b>Mujeres</b>  |       |                |         |                |         |                |        |       |       |
| HDL-C (mg/dL)   | G/G   | 58,50 (16,81)  |         | 56,59 (15,54)  |         | 53,96 (13,96)  |        | 0,085 |       |
|                 | AG+AA | 50,87 (13,70)  |         | 57,10 (15,51)  |         | 61,11 (18,54)  |        | 0,015 |       |
|                 |       | p= 0,027       |         | p= 0,713       |         | p= 0,048       |        |       |       |
|                 |       | p*= 0,041      |         | p*= 0,795      |         | p*= 0,049      |        |       | 0,012 |
| Apo A-I (mg/dL) | G/G   | 168,29 (27,55) |         | 161,76 (25,28) |         | 160,13 (25,10) |        | 0,086 |       |
|                 | AG+AA | 155,93 (24,43) |         | 165,10 (25,40) |         | 164,87 (33,30) |        | 0,220 |       |
|                 |       | p= 0,038       |         | p= 0,157       |         | p= 0,499       |        |       |       |
|                 |       | p*= 0,153      |         | p*= 0,177      |         | p*= 0,512      |        |       | 0,079 |
| <b>Hombres</b>  |       |                |         |                |         |                |        |       |       |
| HDL-C (mg/dL)   | G/G   | 44,25 (12,29)  |         | 42,88 (10,90)  |         | 43,08 (9,87)   |        | 0,604 |       |
|                 | AG+AA | 45,67 (15,21)  |         | 42,78 (10,69)  |         | 47,00 (12,59)  |        | 0,699 |       |
|                 |       | p= 0,588       |         | p= 0,920       |         | p= 0,227       |        |       |       |
|                 |       | p*= 0,536      |         | p*= 0,975      |         | p*= 0,298      |        |       | 0,653 |
| Apo A-I (mg/dL) | G/G   | 145,09 (21,69) |         | 140,76 (20,56) |         | 139,81 (19,58) |        | 0,236 |       |
|                 | AG+AA | 143,58 (28,39) |         | 141,63 (19,99) |         | 154,21 (30,02) |        | 0,129 |       |
|                 |       | p= 0,763       |         | p= 0,665       |         | p= 0,058       |        |       |       |
|                 |       | p*= 0,588      |         | p*= 0,920      |         | p*= 0,227      |        |       | 0,408 |

†: Valor de P para tendencia obtenido en el test ANOVA para la comparación global entre genotipos

\*: Valores de P en la comparación entre portadores G/G y G/A, ajustados por edad, IMC, consumo de tabaco, consumo de alcohol, e ingesta de grasa saturada, monoinsaturada y poliinsaturada.

‡: Valor de P para el término de interacción genotipo APOA1 X PUFA (3 categorías) en el modelo de regresión lineal ajustado por edad, IMC, alcohol, consumo de tabaco, MUFA, PUFA y SATFAT.

Esta predicción es consistente con los resultados observados. Considerando que la ingesta media de PUFA en las mujeres es alrededor de un 6% de la energía por día, el efecto decreciente del alelo A (-10,60 mg/dL) estaba contrarrestado por la interacción de PUFA:  $((1,79 \text{ mg/dL HDL} / 1\% \text{PUFA}) \times 6\% \text{PUFA} = 10,7 \text{ mg/dL})$ .

Tal y como se presenta en la Tabla 4.19, los niveles medios de HDL-C en los homocigotos G/G y portadores de A eran respectivamente de  $56,55 \pm 15,56$  y  $56,73 \pm 15,79$  mg/dL ( $p = 0,881$ ). Según los cálculos anteriores, por debajo del 6% de ingesta de PUFA los valores medios de HDL-C deberían ser más bajos en los homocigotos G/G; y por encima del 6% de ingesta de PUFA, los valores medios de HDL-C deberían ser más altos en las mujeres portadoras del alelo A.

Cuando en el Modelo 5 se ha sustituido la variable continua de PUFA por una dicotómica (PUFA  $\leq 6\%$  y PUFA  $> 6\%$ ), su interacción con el genotipo de APOA1 era estadísticamente significativa ( $p = 0,049$ ). En las mujeres que consumían menos de un 6% de PUFA por día, los niveles medios de HDL-C para las G/G y portadoras de A eran respectivamente de  $57,44 \pm 15,54$  y  $55,54 \pm 15,55$  mg/dL. En las mujeres que consumían más de un 6% de PUFA por día, estos valores eran de  $55,72 \pm 15,29$  mg/dL en las homocigotos G/G y  $57,73 \pm 15,29$  mg/d en las portadoras del alelo A.

La Tabla 4.20 muestra los resultados del análisis de regresión para los hombres. Aunque se ha observado un incremento en la magnitud del coeficiente de regresión para el genotipo de APOA1, ninguno de los términos de interacción para la ingesta de grasa era estadísticamente significativo. Cuando se ha dicotomizado la variable de PUFA continua según el consumo medio, el término de interacción entre esta variable y el genotipo de APOA1 era también no significativo ( $p = 0,757$ ) en los hombres, y no se han detectado diferencias estadísticamente significativas en los niveles de HDL-C a través de los estratos de PUFA.

Aunque la interacción con la ingesta de PUFA era la única estadísticamente significativa, es difícil distinguir las asociaciones de varios tipos de grasa con los niveles de lípidos debido a la multicolinealidad entre los tipos de grasa. En este estudio, los coeficientes de correlación entre PUFA y MUFA eran respectivamente de  $r = 0,537$  y  $r = 0,508$  en hombres y mujeres. Entre PUFA y SAFA eran de  $r = 0,217$  en mujeres y de  $r = 0,140$  en hombres. Estos valores han permitido incluir y retener estas variables en los modelos de regresión multivariada sin problemas de multicolinealidad, y obtener así un estimador más independiente del coeficiente de regresión para la ingesta de PUFA. Entonces, el Modelo 5 (modelo PUFA) se ha ajustado adicionalmente para SAFA y MUFA, obteniendo el Modelo 6. El término

de interacción entre el genotipo de APOA1 y PUFA permanecía estadísticamente significativa en mujeres ( $B = 1,73$ ;  $p = 0,007$ ) y no significativa en hombres ( $B = 0,08$ ;  $p = 0,857$ ). Después de haber probado en el Modelo 6 otras interacciones binarias, ternarias y cuaternarias entre el genotipo de APOA1, el alcohol, el tabaco y la ingesta de PUFA, no se han observado términos de interacción estadísticamente significativos en las mujeres. Sin embargo, en los hombres, se ha encontrado una interacción cuaternaria estadísticamente significativa ( $p = 0,023$ ) entre el genotipo de APOA1, el alcohol, el tabaco y la ingesta de PUFA. El tamaño de la muestra en hombres en este estudio no tiene la suficiente potencia estadística para analizar este efecto en los distintos estratos.

#### 4.3.3.2 Ingesta en dieta de grasa, alelos de APOA1 y niveles de ApoA-I

En un paso posterior del análisis se han considerado los niveles de apoA-I como variable de salida, y se han ajustado los seis modelos de regresión lineal multivariada anteriores para hombres y mujeres por separado. En las mujeres, los valores  $p$  para los términos de interacción fueron los siguientes: 0,543 para grasa total, 0,351 para MUFA, 0,837 para SAFA y 0,160 para PUFA. Aunque el término de interacción entre el genotipo de APOA1 y la ingesta de PUFA era no estadísticamente significativo en la predicción de los niveles de apoA-I en mujeres, los efectos iban en el mismo sentido que los observados por los niveles de HDL-C. No se han obtenido interacciones de orden superior con el tabaco o el alcohol. En los hombres, los valores de  $p$  para las interacciones APOA1 x grasa eran las siguientes: 0,232 para grasa total, 0,525 para MUFA, 0,544 para SAFA y 0,09 para PUFA. Cuando también se han considerado en el Modelo 6 las interacciones con alcohol y tabaco, se ha obtenido una interacción ternaria ( $p = 0,035$ ) entre PUFA, consumo de alcohol y genotipo de APOA1. El efecto de esta interacción era similar a la descrita para HDL-C en las mujeres: los portadores del alelo A tenían niveles de apoA-I menores que los homocigotos G/G, pero estos valores se incrementaban linealmente con la ingesta de PUFA con diferencias entre bebedores y no bebedores. En los no bebedores, la pendiente era mayor que en los no bebedores.

Adicionalmente a los ya demostrados efectos lineales de la ingesta de PUFA, con el propósito de obtener valores medios que permitan la comparación directa con otros estudios, se han dividido los hombre y mujeres en tres categorías de ingesta de PUFA como se ha descrito en el correspondiente apartado de métodos. La Tabla 4.21 muestra los valores medios de HDL-C y apoA-I por el genotipo de APOA1 y sexo a través de las tres categorías de ingesta de PUFA (<4%, 4 a 8% y  $\geq 8\%$  de la energía diaria). Los valores de  $p$  se presentan en crudo y ajustados por edad, IMC, consumo de alcohol, tabaco e ingesta de grasa. En las mujeres, se ha obtenido un efecto claramente diferente del genotipo de APOA1 a través de las distintas categorías de PUFA. Cuando la ingesta de PUFA era baja, los

homocigotos G/G tenían niveles de HDL-C un 14% más altos que los portadores del alelo A ( $p < 0,05$ ). Cuando la ingesta de PUFA era alta, los niveles medios de HDL-C en los portadores de A eran un 13% más altos que en los homocigotos G/G. Para los niveles de apoA-I, la tendencia era similar sin alcanzar la significación estadística. En los hombres, ha sido necesaria una estratificación adicional por consumo de alcohol o tabaco para obtener diferencias significativamente estadísticas entre los homocigotos G/G y los portadores de A.

Finalmente, para probar el efecto dosis gen de APOA1 por la interacción de PUFA en mujeres, se ha dividido el grupo de portadores de A en G/A y A/A. La Figura 4.20 muestra los niveles medios estimados de HDL-C (A) o de apoA-I (B) en mujeres G/G, G/A y A/A a través de las tres categorías de ingesta de PUFA, ajustando por edad, IMC, tabaco, consumo de alcohol, SAFA, MUFA y la interacción entre el genotipo de APOA1 (G/G, G/A y A/A) y tres categorías de ingesta de PUFA (<4%, 4-8% y >8%). Para HDL-C, el término de interacción era estadísticamente significativo ( $p = 0,010$ ) y se ha observado un claro efecto de dosificación genética con los individuos A/A presentando los valores de HDL-C más altos en la categoría de PUFA  $\geq 8\%$ , y los valores de HDL-C más bajos en la categoría de PUFA <4%. Es más, cuando la ingesta de PUFA se ha considerado en continuo, el coeficiente de regresión para los sujetos A/A ajustado por todas las covariantes era dos veces y media mayor que el correspondiente a los heterocigotos A/G ( $p < 0,05$ ). En el modelo de regresión ajustado por los niveles de apoA-I (B), también se ha observado un efecto similar de dosificación genética, sin alcanzar la significación estadística ( $p = 0,092$ ). En los hombres, debido a las interacciones de orden superior, no se probó este efecto de la PUFA.

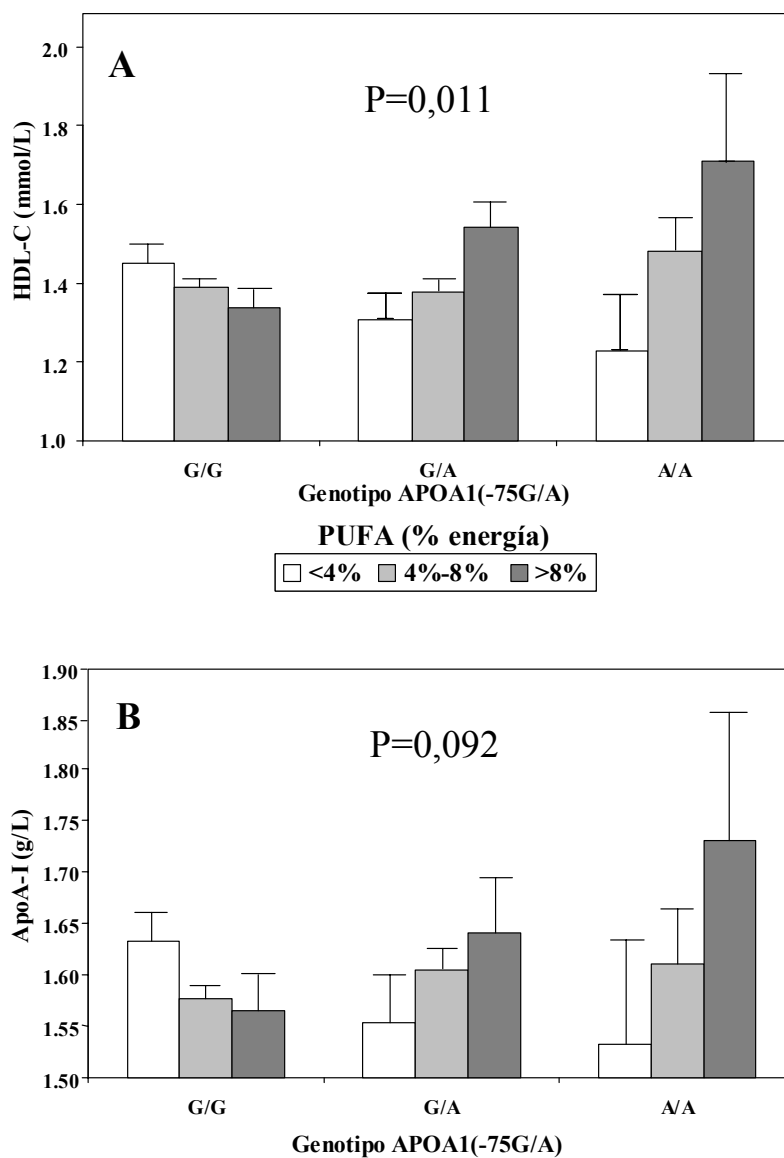


Figura 4.20. Medias ajustadas de HDL-C (A) y apolipoproteína A-I (B) según el genotipo de APOA1 y el consumo de PUFA ( $\leq 4\%$ , 4-8% y  $>8\%$ ) en mujeres. Las medias se han ajustado por edad, IMC, consumo de alcohol, tabaco, energía, consumo de SAFA, MUFA y PUFA. Los valores de P se han obtenido para el término de interacción entre el genotipo de APOA1 y PUFA en los modelos de regresión lineal multivariada adaptados como se ha indicado en los métodos

#### 4.3.4 Resultados del análisis del gen Lipasa Hepática

Estos resultados se han publicado en la revista *Circulation* en el año 2002 con el título “Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on HDL metabolism: Evidence of a strong dose-effect in this gene-nutrient interaction in the Framingham Study” (Ordovás et al., 2002). A continuación se presenta un resumen de los principales hallazgos.

Se analizaron un total de 1.597 individuos (755 hombres y 822 mujeres) participantes en el *Framingham Offspring Study*. En la Tabla 4.22 se presenta un resumen de las características demográficas y bioquímicas y de los datos dietéticos y genéticos de los participantes según el sexo. La prevalencia de sujetos diabéticos era del 5,5%. Las frecuencias genotípicas no se desviaban del equilibrio de Hardy-Weinberg. Las concentraciones de HDL-C eran altas y el tamaño de HDL se incrementaba en los portadores del alelo T (Tabla 4.23), como se ha descrito previamente en esta población (Couture et al., 2000). No se han encontrado diferencias significativas en LDL-C y TG, y en grasa en dieta o consumo de alcohol con relación a los genotipos de LIPC.

Para comprobar si la grasa en dieta modifica el efecto del polimorfismo de LIPC, en primer lugar se ha examinado las categorías de ingesta de grasa total según la media de la población (30% de energía). Se ha encontrado una interacción fuerte entre el polimorfismo de LIPC y el consumo de grasa total sobre las concentraciones de HDL-C (la P para la interacción era  $< 0,001$ ), según muestra la Figura 4.21. El alelo T se ha asociado con un incremento de HDL-C (8,7% en CT y 15,9% en TT,  $p < 0,001$ ) en los sujetos que consumían menos del 30% de energía a partir de la grasa. Sin embargo, las concentraciones medias de HDL-C eran mas bajas entre los portadores del genotipo TT que consumían  $< 30\%$  de grasa total (-9,5%;  $p < 0,05$ ), y no se han observado diferencias entre los individuos CC y CT. Se han encontrado similares interacciones gen  $\times$  nutriente para el HDL<sub>2</sub>-C ( $p < 0,001$ ), subfracciones grandes de HDL ( $p < 0,001$ ) y tamaño de HDL ( $p = 0,001$ ).

**Tabla 4.22. Características demográficas y bioquímicas y de los datos dietéticos y genéticos de los participantes según el sexo**

|                                     | <b>HOMBRES (n=1020)</b> | <b>MUJERES (n=1110)</b> |
|-------------------------------------|-------------------------|-------------------------|
|                                     | Media (SD) o n(%)       | Media (SD) o n(%)       |
| Edad (años)                         | 55,4 (9,9)              | 54,8 (9,6)              |
| IMC (kg/m <sup>2</sup> )            | 28,2 (4,0)              | 26,8 (5,4)*             |
| LDL colesterol (mmol/L)             | 3,27 (0,79)             | 3,23 (0,88)             |
| HDL colesterol (mmol/L)             | 1,12 (0,29)             | 1,45 (0,40)*            |
| Triglicéridos (mmol/L)              | 1,69 (1,32)             | 1,53 (0,98)*            |
| Glucosa (mmol/L)                    | 5,39 (1,29)             | 5,11 (1,21)*            |
| Ingesta Energía (MJ/d)              | 8.376 (2.731)           | 7.231 (2.377)*          |
| Grasa Total (g/d)                   | 67,1 (28,1)             | 56,5 (23,9)*            |
| SFA (% energía)                     | 10,7 (3,0)              | 10,3 (2,8)*             |
| MUFA (% energía)                    | 11,5 (2,7)              | 11,0 (2,6)*             |
| PUFA (% energía)                    | 5,7 (1,7)               | 6,0 (1,7)*              |
| Grasa Animal (% energía)            | 16,1 (5,5)              | 15,2 (5,0)*             |
| Grasa Vegetal (% energía)           | 13,9 (4,7)              | 14,0 (4,4)              |
| Carbohidratos (% energía)           | 50,0 (8,4)              | 51,9 (8,3)*             |
| Fibra (g/d)                         | 19,1 (8,5)              | 18,9 (8,2)              |
| Alcohol (g/d)                       | 15,4 (20,4)             | 7,2 (8,3)*              |
| Bebedores (n, %)                    | 736 (72,2)              | 760 (69,1)*             |
| Fumadores (n, %)                    | 193 (18,9)              | 204 (18,4)              |
| Sí Beta-bloqueantes Rx (n, %)       | 110 (10,7)              | 98 (8,8)                |
| Sí estrógeno Rx (n, %) <sup>†</sup> | 0 (0,0)                 | 229 (20,0)              |
| Genotipo lipasa hepática (n, %)     |                         |                         |
| CC                                  | 653 (64,0)              | 706 (63,6)              |
| CT                                  | 342 (33,5)              | 356 (32,1)              |
| TT                                  | 25 (2,5)                | 48 (4,3)                |

Los valores se muestran en formato "media (desviación estándar, SD)" or "número (n) y porcentaje (%)".

\*: Significativamente diferentes de los hombres (P < 0,05)

La ingesta de energía se expresa en MJ/días por día. El factor de conversión es 1 MJ = 239,23 kcal.

SFA, ácidos grasos saturados; MUFA, ácidos grasos monoinsaturados; PUFA, ácidos grasos poliinsaturados.

<sup>†</sup>Incluye terapia de sustitución hormonal y el uso de anticonceptivos orales.



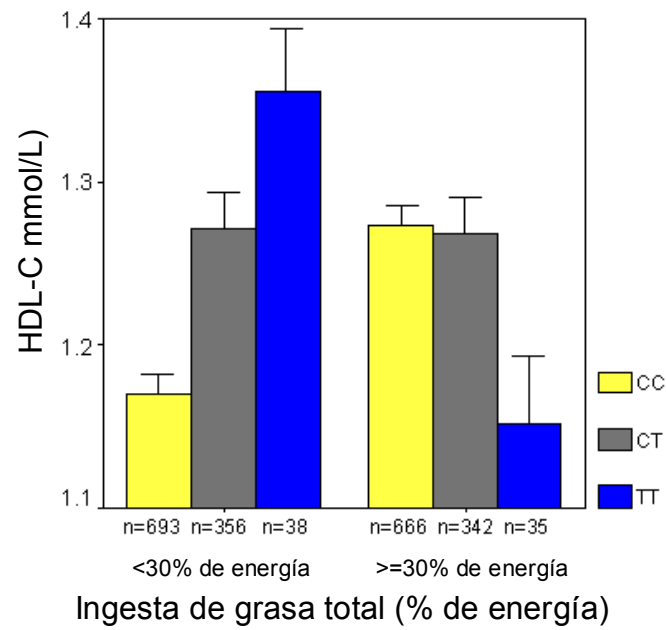


Figura 4.21. Concentraciones medias de HDL-C según el polimorfismo de LIPC y las categorías de ingesta de grasa total. Se han ajustado las medias por relaciones familiares, sexo, edad, IMC, tabaco, alcohol, estrógenos, Betabloqueantes y energía. Las barras de error indican en error estándar de la media

**Tabla 4.23. Media de lípidos plasmáticos e ingesta dietaria según los genotipos de LIPC**

|                              | CC                     | CT                    | TT                   | P*     |
|------------------------------|------------------------|-----------------------|----------------------|--------|
|                              | n = 1359<br>Media (SD) | n = 698<br>Media (SD) | n = 73<br>Media (SD) |        |
| Edad (años)                  | 54,6 (9,7)             | 55,0 (9,8)            | 53,3 (9,5)           | 0,303  |
| IMC (kg/m <sup>2</sup> )     | 27,5 (5,0)             | 27,4 (4,7)            | 27,4 (4,5)           | 0,824  |
| HDL-C (mmol/L)               | 1,29 (0,39)            | 1,33 (0,39)           | 1,37 (0,46)          | 0,034  |
| HDL <sub>2</sub> -C (mmol/L) | 0,23 (0,15)            | 0,25 (0,16)           | 0,29 (0,19)          | <0,001 |
| HDL <sub>3</sub> -C (mmol/L) | 1,06 (0,26)            | 1,08 (0,26)           | 1,08 (0,30)          | 0,205  |
| HDL grande (mmol/L)          | 0,60 (0,43)            | 0,64 (0,44)           | 0,74 (0,52)          | 0,010  |
| HDL intermedia (mmol/L)      | 0,54 (0,17)            | 0,54 (0,18)           | 0,51 (0,17)          | 0,383  |
| HDL pequeña (mmol/L)         | 0,18 (0,13)            | 0,17 (0,13)           | 0,16 (0,14)          | 0,455  |
| HDL diámetro (nm)            | 9,17 (0,47)            | 9,22 (0,49)           | 9,30 (0,58)          | 0,014  |
| Triglicéridos (mmol/L)       | 1,60 (1,10)            | 1,63 (1,19)           | 1,86 (1,13)          | 0,204  |
| LDL-C (mmol/L)               | 3,25 (0,84)            | 3,30 (0,85)           | 3,13 (0,74)          | 0,107  |
| Glucosa (mmol/L)             | 5,18 (1,14)            | 5,22 (1,33)           | 5,38 (1,64)          | 0,255  |
| Energía (MJ/d)               | 7,82 (2,64)            | 7,91 (2,65)           | 7,51 (2,37)          | 0,411  |
| Grasa Total (% energía)      | 29,9 (6,5)             | 29,8 (6,0)            | 28,3 (6,7)           | 0,129  |
| SFA (% energía)              | 10,7 (2,9)             | 10,5 (2,7)            | 10,1 (2,9)           | 0,084  |
| MUFA (% energía)             | 11,3 (2,7)             | 11,3 (2,5)            | 10,7 (2,8)           | 0,117  |
| PUFA (% energía)             | 5,8 (1,7)              | 5,9 (1,7)             | 5,7 (1,8)            | 0,291  |
| Alcohol (g/d)                | 11,3 (16,9)            | 10,7 (16,6)           | 9,9 (15,4)           | 0,587  |

SD: Desviación estándar.

\*: Valor de P obtenido en el test ANOVA en la comparación con los genotipos de HL después de ajustar por relaciones familiares.

SFA, ácidos grasos saturados; MUFA, ácidos grasos monoinsaturados; PUFA, ácidos grasos poliinsaturados.

**Tabla 4.24. Interacción del consumo de grasa con los efectos de los genotipos de lipasa hepática sobre las concentraciones de HDL-C. Análisis de regresión combinado múltiple (centrado y no centrado por grasa) por el tipo de grasa consumida**

|  | Total fat*     |        |  | SFA*           |        |  | MUFA*          |       |  | PUFA*          |       |  |
|--|----------------|--------|--|----------------|--------|--|----------------|-------|--|----------------|-------|--|
|  | B (SE)         | p      |  | B (SE)         | p      |  | B (SE)         | p     |  | B (SE)         | p     |  |
| CT (en el modelo centrado en grasa)†                     | 0,053 (0,015)  | 0,001  |  | 0,054 (0,015)  | <0,001 |  | 0,053 (0,014)  | 0,001 |  | 0,052 (0,016)  | 0,001 |  |
| (en el modelo no centrado)‡                              | 0,221 (0,074)  | 0,003  |  | 0,186 (0,059)  | 0,002  |  | 0,196 (0,067)  | 0,005 |  | 0,098 (0,054)  | 0,077 |  |
| TT (en el modelo centrado en grasa)†                     | 0,042 (0,040)  | 0,290  |  | 0,052 (0,041)  | 0,193  |  | 0,042 (0,041)  | 0,297 |  | 0,047 (0,042)  | 0,217 |  |
| (en el modelo no centrado)‡                              | 0,567 (0,168)  | 0,001  |  | 0,341 (0,139)  | 0,015  |  | 0,479 (0,154)  | 0,002 |  | 0,236 (0,130)  | 0,071 |  |
| Ingesta de grasa (% energía)§                            | 0,009 (0,001)  | 0,001  |  | 0,017 (0,000)  | 0,001  |  | 0,018 (0,003)  | 0,001 |  | 0,012 (0,005)  | 0,027 |  |
| Término de interacción: Lipasa hepática x tipo de grasa§ |                | <0,001 |  |                | 0,002  |  |                | 0,001 |  |                | 0,360 |  |
| CT x Ingesta de grasa                                    | -0,006 (0,002) | 0,021  |  | -0,012 (0,005) | 0,021  |  | -0,013 (0,006) | 0,021 |  | -0,008 (0,009) | 0,379 |  |
| TT x Ingesta de grasa                                    | -0,018 (0,006) | 0,002  |  | -0,027 (0,013) | 0,040  |  | -0,039 (0,014) | 0,002 |  | -0,032 (0,022) | 0,144 |  |

La variable dependiente era HDL colesterol (mmol/L). B, coeficiente de regresión; SE, error estándar.

SFA, ácidos grasos saturados; MUFA, ácidos grasos monoinsaturados; PUFA, ácidos grasos poliinsaturados.

\*: Se han adaptado dos modelos de regresión separados (uno centrado en grasa y otro no centrado) para cada tipo de grasa (como variables continuas). Los modelos se han ajustado adicionalmente por relaciones familiares, sexo, edad, IMC, tabaco, ingesta de alcohol, terapia de entérogénos en mujeres, uso de betabloqueantes e ingesta de energía.

Los enfoques centrado y no centrado proporcionan idénticos estadísticos y test de modelos de regresión lineal para los efectos de interacción.

†: En el modelo centrado en grasa, la media de cada grasa se ha restado de la correspondiente variable de grasa original, de forma que los términos de efectos principales son interpretables como las concentraciones de HDL-C en las correspondientes medias de cada grasa.

‡: En los modelos no centrados, no se han transformado las variables.

§: Los coeficientes de regresión y la significación para estos términos eran los mismos en los modelos centrados y no centrados.

Las pendientes para los genotipos CC, CT y TT para cada tipo de grasa se pueden obtener fácilmente: el término de ingesta de grasa es la pendiente de CC (0,009, para grasa total); el término [CT x ingesta de grasa + pendiente de CC] es la pendiente de CT (0,003, para grasa total); y el término [TT x ingesta de grasa + pendiente de CC] es la pendiente de TT (-0,009, para grasa total).

**Tabla 4.25. Valores de P para los términos de interacción entre ingesta de grasa (en continuo) y el genotipo de la lipasa hepática (tres categorías) en la determinación de las concentraciones de HDL-C y la distribución de subclases. Análisis de regresión multivariados por el tipo de grasa consumida**

| Términos de interacción en los modelos de regresión* | Variable dependiente |                     |                      |              |
|--|----------------------|---------------------|----------------------|--------------|
|  | HDL-C (mmol/L)       | HDL grande (mmol/L) | HDL int+peq (mmol/L) | tam HDL (nm) |
| Grasa total (% energía) x lipasa hepática            |                      |                     |                      |              |
| Multivariado†  | <0.001               | <0.001              | 0,069                | 0,007        |
| Ajuste adicional para carbohidratos‡                 |                      |                     |                      |              |
| SFA (% energía) x lipasa hepática                    | <0.001               | <0.001              | 0,064                | 0,009        |
| Multivariado†  | 0,002                | <0.001              | 0,083                | 0,002        |
| Ajuste adicional para carbohidratos, MUFA y PUFA‡    | 0,002                | <0.001              | 0,078                | 0,004        |
| MUFA (% energía) x lipasa hepática                   |                      |                     |                      |              |
| Multivariado†  | 0,001                | <0.001              | 0,015                | 0,003        |
| Ajuste adicional para carbohidratos, SFA y PUFA‡     | 0,001                | <0.001              | 0,016                | 0,004        |
| PUFA (% energía) x lipasa hepática                   |                      |                     |                      |              |
| Multivariado†  | 0,379                | 0,809               | 0,988                | 0,546        |
| Ajuste adicional para carbohidratos, SFA y MUFA‡     | 0,208                | 0,814               | 0,994                | 0,684        |

\*Se han adaptado modelos de regresión con términos de interacción para cada tipo de grasa dietaria (en continuo) y para cada variable dependiente.

†: Los modelos multivariados incluyen relaciones familiares, sexo, edad, IMC, tabaco, ingesta de alcohol, terapia de estrógenos en mujeres, uso de betabloqueantes, ingesta de energía, genotipo de lipasa hepática e ingesta de grasa específica.

‡: Los modelos multivariados se han ajustado adicionalmente por carbohidratos (% energía) y por SFA, MUFA o PUFA (en continuo) con el propósito de ajustar mutuamente las ingestas de tipos específicos de ácidos grasos por las variables anteriores. SFA, ácidos grasos saturados; MUFA, ácidos grasos monoinsaturados; PUFA, ácidos grasos poliinsaturados.

Posteriormente, se ha examinado el efecto de la grasa total y los ácidos grasos específicos (FA) como variables continuas. La Tabla 4.23 muestra los coeficientes de regresión para los efectos principales y los términos de interacción entre la grasa total o los ácidos grasos específicos y los genotipos de LIPC en la determinación de las concentraciones de HDL-C. Con el propósito de incrementar la interpretabilidad, se han adaptado por separado dos modelos de regresión (centrados y no centrados en la grasa) para cada tipo de grasa. Se ha observado una interacción significativa ( $p < 0,001$ ) entre la grasa total y el genotipo LIPC. Los coeficientes de regresión parcial para el efecto de interacción en los individuos de los alelos CT y TT eran ambos estadísticamente significativos en la comparación con los individuos del alelo CC. También se ha observado la significación estadística de las interacciones con SFAT ( $p = 0,002$ ) y MUFA ( $p = 0,001$ ). Sin embargo, no había interacciones significativas con PUFA ( $p = 0,36$ ). Según el modelo de regresión no centrado de la Tabla 4.23, y para destacar el efecto de interacción, la Figura 4.22 muestra la modificación del efecto del polimorfismo de LIPC sobre HDL-C (A) y el tamaño de HDL (B) por medio de la grasa total consumida. Las diferencias en la pendiente de la regresión indican que el efecto de la grasa total sobre las concentraciones o el tamaño de HDL-C depende del genotipo de LIPC. La baja ingesta de grasa se ha asociado con más altas concentraciones de HDL-C en los individuos TT. Sin embargo, en estos sujetos el HDL-C decrecía cuando la ingesta de grasa total se incrementaba. Se encontró el efecto opuesto en los individuos CC. Además, en los sujetos CC, el tamaño de partícula de HDL se ha asociado con mayor ingesta de grasa, mientras que el aumento de ingesta de grasa provocaba la reducción en el tamaño de partícula en los individuos CT y TT. Estos efectos en el metabolismo de HDL se han encontrado consistentes sobre diferentes subfracciones de HDL incluso después de haber controlado por variables de confusión potenciales (Tabla 4.24).

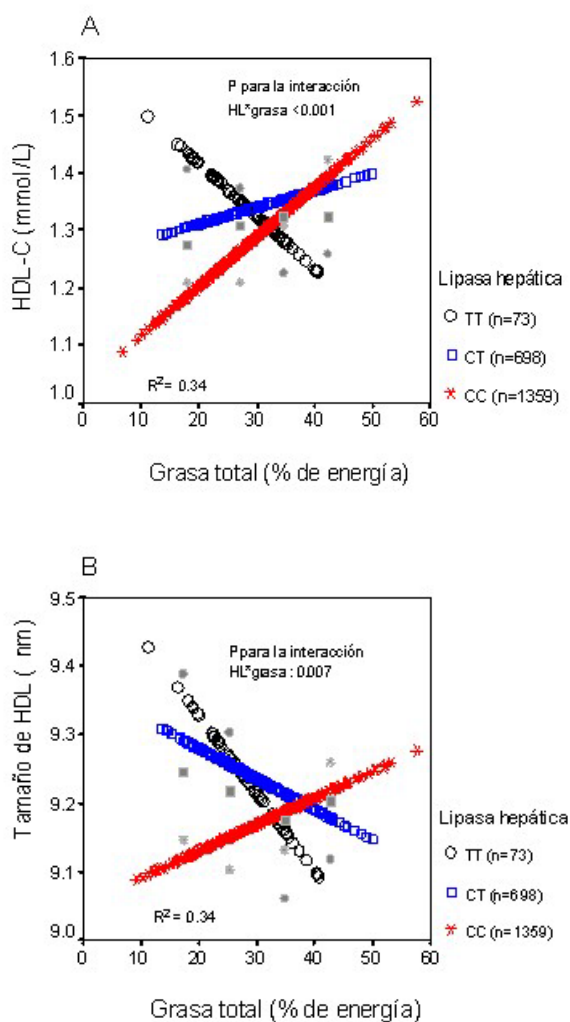


Figura 4.22. Valores predichos (símbolos abiertos) del tamaño de HDL-C (A) y HDL según los genotipos de HL dependientes de la grasa total consumida (en continuo). Los valores predichos se han calculado a partir de los modelos de regresión que contienen la ingesta de grasa total, el polimorfismo de LIPC, su término de interacción y las variables de confusión potenciales. Los valores de  $P$  y  $R^2$  se han obtenido de los modelos de regresión. Los símbolos sólidos representan la media de la variable dependiente por genotipo en las cuatro categorías de grasa total (valores medios:  $<20\%$  [ $n = 166$ ],  $20$  a  $30\%$  [ $n = 951$ ],  $30$  a  $40\%$  [ $n = 882$ ] y  $\geq 40\%$  [ $n = 125$ ])

Como resulta que las ingestas de los tipos específicos de grasa tienen correlación entre sí, se ha examinado esta correlación tomando como fuente la grasa (Tabla 4.26). Los MUFA y PUFA estaban altamente correlacionados con la grasa animal debido a las fuentes compartidas de SAFA y MUFA en la población de la que se tomó la muestra. En consecuencia, se han llevado a cabo análisis posteriores considerando la fuente de grasa (animal o vegetal). Se han adaptado modelos de regresión multivariada para grasa animal y vegetal. Estos modelos se han ajustado adicionalmente por carbohidratos y por grasa vegetal o animal. Para la grasa animal, la significación estadística de los términos de interacción con el polimorfismo de LIPC ha sido la siguiente en las determinaciones correspondientes: HDL-C,  $p < 0,001$ ; tamaño de partícula de HDL,  $p = 0,003$ ; subfracción mayor de HDL,  $p < 0,001$ ; y subfracciones intermedia y menores de HDL,  $p = 0,048$ .

**Tabla 4.26. Coeficientes de correlación de Pearson entre tipos de grasa dietaria en los participantes del estudio**

|               | Grasa total | Grasa animal | Grasa vegetal | SFA    | MUFA   |
|---------------|-------------|--------------|---------------|--------|--------|
| Grasa animal  | 0.721*      |              |               |        |        |
| Grasa vegetal | 0.584*      | -0.141*      |               |        |        |
| SFA           | 0.860*      | 0.860*       | 0.221*        |        |        |
| MUFA          | 0.949*      | 0.676*       | 0.565*        | 0.799* |        |
| PUFA          | 0.604*      | 0.044†       | 0.811*        | 0.183* | 0.520* |

SFA, ácidos grasos saturados; MUFA, ácidos grasos monoinsaturados; PUFA, ácidos grasos poliinsaturados.

\*:  $P < 0.01$ ; †:  $P < 0.05$

El efecto de la grasa animal sobre estas variables iba en la misma dirección así como para la grasa total. Las pendientes ajustadas (en mmol/L por % de grasa) en la estimación de HDL-C a partir de la grasa animal han sido las siguientes: 0,014,  $p < 0,001$ , en el genotipo CC; 0,008,  $p = 0,051$  para la diferencia con el genotipo CT; y -0,017,  $p < 0,001$  para la diferencia con el genotipo TT. La Figura 4.23 muestra el efecto de interacción de los polimorfismos de LIPC y la grasa animal sobre las subfracciones de HDL mayor (A) e intermedia y menor (B). Sin embargo, no se han encontrado interacciones estadísticamente significativas entre la grasa vegetal y el polimorfismo de LIPC en la determinación de HDL-C ( $p = 0,400$ ); tamaño de partícula de HDL ( $p = 0,415$ ), subfracción mayor de HDL ( $p = 0,332$ ), y subfracciones intermedia y menores de HDL ( $p = 0,738$ ).

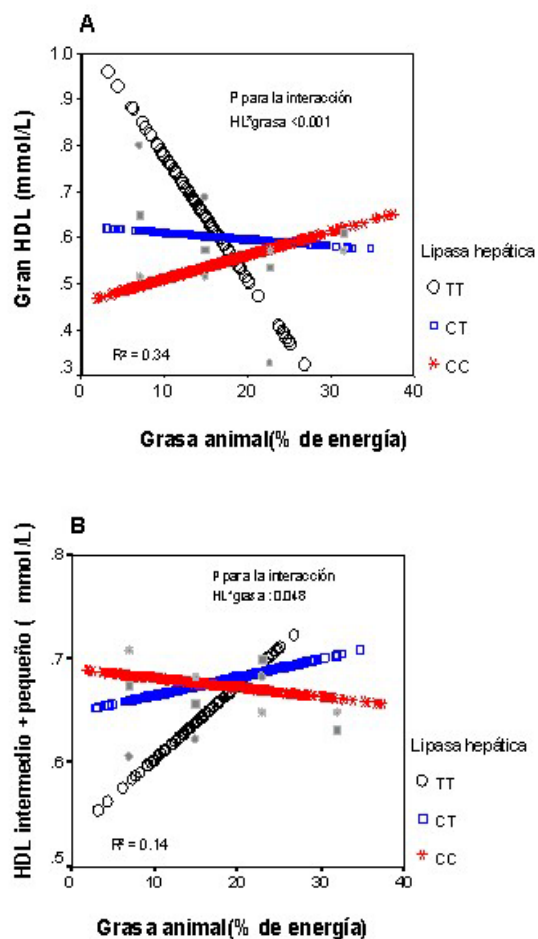


Figura 4.23. Valores predichos (símbolos abiertos) de las concentraciones de gran HDL (A) y partículas intermedias y pequeñas de HDL (B) según los genotipos de HL dependientes de la cantidad de grasa animal consumida (en continuo). Los valores predichos se han calculado a partir de los modelos de regresión que contienen grasa animal, el polimorfismo de LIPC, su término de interacción, y las variables de confusión potenciales. Los valores de P y R<sup>2</sup> se han obtenido de los modelos. Los símbolos sólidos representan las medias de la variable dependiente para el genotipo y las cuatro categorías de grasa animal (valores medios: <10% [n = 303], 10 a 20% [n = 1419], 20 a 30% [n = 369] y >=30% [n = 39]; no se han encontrado sujetos TT en la categoría de máximo consumo de grasa

Las pendientes ajustadas (en mmol/L por % de grasa) en la estimación de HDL-C a partir de la grasa vegetal han sido las siguientes: 0,001, p < 0,611, en el



genotipo CC; 0,004,  $p = 0,161$  para la diferencia con el genotipo CT; y  $-0,006$ ,  $p < 0,001$  para la diferencia con el genotipo TT; todas cercanas a 0.

Para ganar mayor perspectiva en los mecanismos por los que la grasa total modula el efecto del polimorfismo de LIPC sobre el metabolismo de HDL, se ha analizado esta interacción para las concentraciones plasmáticas de LDL-C, TG y glucosa en ayunas. Las significaciones estadísticas de los términos de interacción entre el polimorfismo de LIPC y la ingesta de grasa total después de haber controlado por variables de confusión han sido la siguiente en la determinación de: LDL-C,  $p = 0,944$ ; TG,  $p = 0,154$ ; y glucosa  $p = 0,054$ . No se ha observado ninguna modificación en LDL-C del efecto por la grasa, aunque sí se han afectado los TG y la glucosa. Los sujetos TT han mostrado las mayores concentraciones de glucosa [5,73 (0,23) mmol/L contra 4,98 (0,18)mmol/L] y TG [2,0 (1,9) mmol/L contra 1,74 (1,7) mmol/L], en el respectivo consumo superior e inferior al 30% de grasa. No se han encontrado diferencias en los individuos CC [5,19 (0,05) mmol/L contra 5,17 (0,06) mmol/L] o los individuos CT [5,20 (0,72) mmol/L contra 5,24 (0,71) mmol/L] para la glucosa, o para los TG [1,56 (0,05) mmol/L contra 1,73 (0,054) mmol/L] y [1,63 (0,07) mmol/L contra 1,72 (0,07) mmol/L], en el respectivo consumo superior e inferior al 30% de grasa. Sin embargo, cuando se ha examinado la asociación entre la diabetes y el polimorfismo de LIPC, no se han encontrado diferencias en el porcentaje de los sujetos diabéticos por genotipo (5,67% para los individuos CC; 5,44% para los individuos CT; y 5,48% para los individuos TT;  $p = 0,978$ ). Finalmente, se han considerado controles adicionales por diabetes como variable de confusión potencial. La consecuencia es que el ajuste por diabetes no ha tenido efecto en todos los resultados obtenidos previamente.

#### 4.3.5 Resultados del análisis del gen SBR-I

Estos resultados se han publicado en la revista *Journal of Clinical Endocrinology Metabolism* en el año 2003 con el título “Genetic variation at the scavenger receptor class B type I (SR-BI) gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: The Framingham Study” (Osgood et al., 2003). A continuación se presenta un resumen de los principales hallazgos.

Se analizaron un total de 1.597 individuos (755 hombres y 822 mujeres) participantes en el *Framingham Offspring Study*. En la Tabla 4.27 se presenta un resumen de las características antropométricas, bioquímicas y genéticas de los participantes según su situación diabética. Las frecuencias de los alelos menos comunes para cada uno de los tres polimorfismos en el locus del SR-BI para la

población de estudio eran las siguientes: exón 1, 0,127 (95% CI: 0,118-0,136); intrón 5, 0,088 (95% CI: 0,080-0,096); y exón 8, 0,486 (95% CI: 0,499-0,528).

**Tabla 4.27. Características demográficas, bioquímicas, dietarias y genotípicas de los participantes según el sexo y la diabetes de tipo 2**

|                                     | Hombres (n=1293)       |                   |                    | Mujeres (n=1357)       |                   |                   |
|-------------------------------------|------------------------|-------------------|--------------------|------------------------|-------------------|-------------------|
|                                     | No diabéticos (n=1173) |                   | Diabéticos (n=120) | No diabéticos (n=1290) |                   | Diabéticos (n=67) |
|                                     | Media (SD) o n(%)      | Media (SD) o n(%) | p*                 | Media (SD) o n(%)      | Media (SD) o n(%) | p*                |
| Edad (años)                         | 52,0 (10,0)            | 58,2 (8,9)        | <0,01              | 51,4 (97,9)            | 57,6 (8,1)        | <0,01             |
| IMC (kg/m <sup>2</sup> )            | 27,5 (3,7)             | 29,4 (4,6)        | <0,01              | 25,8 (5,1)             | 30,3 (7,3)        | <0,01             |
| LDL-C (mmol/L)                      | 3,47 (0,84)            | 3,30 (1,00)       | 0,09               | 3,30 (0,92)            | 3,38 (0,93)       | 0,47              |
| HDL-C (mmol/L)                      | 1,13 (0,29)            | 1,01 (0,31)       | <0,01              | 1,47 (0,39)            | 1,11 (0,35)       | <0,01             |
| Triglicéridos (mmol/L)              | 1,51 (1,12)            | 2,12 (1,39)       | <0,01              | 1,38 (0,85)            | 2,41 (1,81)       | <0,01             |
| Tamaño LDL (nm)                     | 20,70 (0,57)           | 20,50 (0,56)      | <0,01              | 21,08 (0,44)           | 20,71 (0,64)      | <0,01             |
| Tamaño HDL (nm)                     | 8,93 (0,40)            | 8,89 (0,39)       | 0,38               | 9,41 (0,45)            | 9,12 (0,42)       | <0,01             |
| Glucosa en ayunas (mmol/L)          | 5,19 (0,54)            | 10,77 (3,22)      | <0,01              | 4,95 (0,52)            | 11,20 (2,83)      | <0,01             |
| Bebedores (n, %)                    | 881 (75,4)             | 68 (57,1)         | <0,01              | 841 (65,4)             | 33 (49,3)         | <0,01             |
| Fumadores (n, %)                    | 287 (24,5)             | 23 (19,3)         | 0,21               | 314 (22,4)             | 14 (20,9)         | 0,51              |
| Sí Betabloqueantes Rx (n, %)        | 120 (10,2)             | 29 (24,4)         | <0,01              | 85 (6,6)               | 11 (16,4)         | <0,01             |
| Posmenopausia (n, %)                | 0 (0,0)                | 0 (0,0)           |                    | 692 (53,7)             | 57 (85,1)         | <0,01             |
| Sí estrógeno Rx (n, %) <sup>†</sup> | 0 (0,0)                | 0 (0,0)           |                    | 119 (9,3)              | 1 (1,5)           | 0,03              |
| Genotipo exón 1 de SR-BI (n, %)     |                        |                   | 0,72               |                        |                   | 0,69              |
| GG                                  | 838 (77,4)             | 85 (75,9)         |                    | 896 (74,3)             | 49 (76,6)         |                   |
| GA                                  | 231 (21,3)             | 26 (23,3)         |                    | 238 (24,5)             | 15 (23,4)         |                   |
| AA                                  | 14 (1,3)               | 1 (0,9)           |                    | 11 (1,2)               | 0 (0,0)           |                   |
| Genotipo intrón 1 de SR-BI (n, %)   |                        |                   | 0,16               |                        |                   | 0,46              |
| CC                                  | 915 (82,3)             | 87 (76,9)         |                    | 1.026 (84,3)           | 57 (87,7)         |                   |
| CT                                  | 188 (16,9)             | 24 (21,2)         |                    | 183 (15,0)             | 8 (12,3)          |                   |
| TT                                  | 8 (0,7)                | 2 (1,8)           |                    | 8 (0,7)                | 0 (0,0)           |                   |
| Genotipo exón 8 de SR-BI (n, %)     |                        |                   | 0,97               |                        |                   | 0,77              |
| CC                                  | 297 (27,8)             | 30 (27,8)         |                    | 325 (27,6)             | 16 (25,4)         |                   |
| CT                                  | 504 (47,1)             | 52 (48,0)         |                    | 562 (47,8)             | 29 (46,0)         |                   |
| TT                                  | 268 (25,1)             | 26 (24,1)         |                    | 289 (24,6)             | 18 (28,6)         |                   |

Los valores se muestran como media (desviación estándar, SD) o como número (n) y porcentaje (%).

<sup>†</sup>Incluye terapia de sustitución hormonal y el uso de contraceptivos orales.

\*: valor de p en la comparación entre no diabéticos y diabéticos.

**Tabla 4.28. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 1 de SR-BI para hombres y mujeres**

| EXON 1 Genotype              |                   |                      |                      |                      |
|------------------------------|-------------------|----------------------|----------------------|----------------------|
| MEN                          | GG (n=923)        | GA/AA (n=272)        | P <sup>1</sup>       | P <sup>2</sup>       |
| IMC (Kg/m <sup>2</sup> )     | 27,80 ± 0,16      | 27,57 ± 0,27         | 0,46                 | 0,42                 |
| Colesterol total (mmol/L)    | 5,20 ± 0,04       | 5,13 ± 0,06          | 0,17                 | 0,24                 |
| LDL-C (mmol/L)               | 3,37 ± 0,03       | 3,30 ± 0,06          | 0,12                 | 0,24                 |
| HDL-C (mmol/L)               | 1,11 ± 0,01       | 1,12 ± 0,02          | 0,33                 | 0,56                 |
| HDL <sub>2</sub> -C (mmol/L) | 0,13 ± 0,01       | 0,14 ± 0,01          | 0,13                 | 0,20                 |
| HDL <sub>3</sub> -C (mmol/L) | 0,99 ± 0,01       | 0,99 ± 0,01          | 0,56                 | 0,90                 |
| Triglicéridos (mmol/L)       | 1,36 ± 0,01       | 1,32 ± 0,01          | 0,46                 | 0,39                 |
| ApoA-I (g/L)                 | 1,34 ± 0,01       | 1,35 ± 0,01          | 0,41                 | 0,57                 |
| Apo B (g/L)                  | 0,99 ± 0,01       | 0,99 ± 0,01          | 0,33                 | 0,59                 |
| HDL (mmol/L)                 |                   |                      |                      |                      |
| Grande                       | 0,38 ± 0,01       | 0,38 ± 0,02          | 0,51                 | 0,95                 |
| Intermedia                   | 0,56 ± 0,01       | 0,55 ± 0,01          | 0,89                 | 0,87                 |
| Pequeña                      | 0,21 ± 0,01       | 0,22 ± 0,01          | 0,93                 | 0,70                 |
| LDL (mmol/L)                 |                   |                      |                      |                      |
| Grande                       | 1,70 ± 0,04       | 1,79 ± 0,06          | 0,13                 | 0,16                 |
| Intermedia                   | 0,95 ± 0,03       | 0,91 ± 0,04          | 0,35                 | 0,32                 |
| Pequeña                      | 0,87 ± 0,03       | 0,81 ± 0,05          | 0,10                 | 0,27                 |
| Tamaño HDL (nm)              | 8,93 ± 0,02       | 8,93 ± 0,02          | 0,31                 | 0,87                 |
| Tamaño LDL (nm)              | 20,66 ± 0,02      | 20,71 ± 0,04         | 0,14                 | 0,23                 |
| Tamaño VLDL (nm)             | 49,38 ± 0,40      | 48,96 ± 0,66         | 0,57                 | 0,56                 |
| <b>WOMEN</b>                 | <b>GG (n=945)</b> | <b>GA/AA (n=325)</b> | <b>P<sup>1</sup></b> | <b>P<sup>2</sup></b> |
| IMC (Kg/m <sup>2</sup> )     | 26,11 ± 0,22      | 26,39 ± 0,32         | 0,35                 | 0,44                 |
| Colesterol total (mmol/L)    | 5,25 ± 0,04       | 5,21 ± 0,05          | 0,86                 | 0,45                 |
| LDL-C (mmol/L)               | 3,25 ± 0,04       | 3,19 ± 0,05          | 1,00                 | 0,30                 |
| HDL-C (mmol/L)               | 1,44 ± 0,01       | 1,45 ± 0,02          | 0,71                 | 0,76                 |
| HDL <sub>2</sub> -C (mmol/L) | 0,25 ± 0,01       | 0,26 ± 0,01          | 0,82                 | 0,40                 |
| HDL <sub>3</sub> -C (mmol/L) | 1,19 ± 0,01       | 1,19 ± 0,02          | 0,39                 | 0,87                 |
| Triglicéridos (mmol/L)       | 0,97 ± 0,01       | 0,97 ± 0,01          | 0,44                 | 0,96                 |
| ApoA-I (g/L)                 | 1,54 ± 0,02       | 1,54 ± 0,02          | 0,52                 | 0,92                 |
| Apo B (g/L)                  | 0,95 ± 0,01       | 0,94 ± 0,01          | 0,77                 | 0,58                 |
| HDL (mmol/L)                 |                   |                      |                      |                      |
| Grande                       | 0,80 ± 0,02       | 0,81 ± 0,03          | 0,53                 | 0,77                 |
| Intermedia                   | 0,54 ± 0,01       | 0,54 ± 0,01          | 0,87                 | 0,90                 |
| Pequeña                      | 0,14 ± 0,01       | 0,13 ± 0,01          | 1,00                 | 0,31                 |
| LDL (mmol/L)                 |                   |                      |                      |                      |
| Grande                       | 2,11 ± 0,04       | 2,16 ± 0,05          | 0,26                 | 0,47                 |
| Intermedia                   | 0,80 ± 0,02       | 0,81 ± 0,04          | 0,49                 | 0,82                 |
| Pequeña                      | 0,49 ± 0,02       | 0,50 ± 0,04          | 0,53                 | 0,69                 |
| Tamaño HDL (nm)              | 9,40 ± 0,02       | 9,40 ± 0,03          | 0,63                 | 0,94                 |
| Tamaño LDL (nm)              | 21,06 ± 0,02      | 21,06 ± 0,03         | 0,96                 | 0,89                 |
| Tamaño VLDL (nm)             | 44,50 ± 0,34      | 44,72 ± 0,53         | 0,76                 | 0,72                 |

Los resultados se dan como media ± error estándar (del modelo 2 completamente ajustado).

<sup>1</sup>After adjustment for familial relationship

<sup>2</sup>After adjustment for familial relationship, age, BMI, smoking, alcohol intake, use of beta-bloc apoE genotype, and menopausal status and estrogen therapy (in women)

**Tabla 4.29. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 1 de SR-BI dependientes de la diabetes 2**

|                              | No diabeticos |               |              | Diabetic     |                             |                             |
|------------------------------|---------------|---------------|--------------|--------------|-----------------------------|-----------------------------|
|                              | GG (n=1734)   | GA/AA (n=555) | GG (n=134)   | GA/AA (n=42) | P1 interacción con diabetes | P2 interacción con diabetes |
| <b>HOMBRES+MUJERES</b>       |               |               |              |              |                             |                             |
| IMC (Kg/m <sup>2</sup> )     | 26,72 ± 0,14  | 26,70 ± 0,21  | 29,28 ± 0,47 | 30,60 ± 0,93 | 0,18                        | 0,20                        |
| Colesterol total (mmol/L)    | 5,23 ± 0,03   | 5,20 ± 0,04   | 5,22 ± 0,11  | 4,87 ± 0,01  | 0,12                        | 0,14                        |
| LDL-C (mmol/L)               | 3,31 ± 0,03   | 3,29 ± 0,04   | 3,19 ± 0,10  | 2,75 ± 0,01  | 0,03                        | 0,03                        |
| HDL-C (mmol/L)               | 1,29 ± 0,01   | 1,30 ± 0,02   | 1,20 ± 0,03  | 1,12 ± 0,01  | 0,04                        | 0,11                        |
| HDL <sub>2</sub> -C (mmol/L) | 0,19 ± 0,01   | 0,20 ± 0,01   | 0,17 ± 0,01  | 0,15 ± 0,01  | 0,02                        | 0,04                        |
| HDL <sub>3</sub> -C (mmol/L) | 1,10 ± 0,01   | 1,10 ± 0,01   | 1,03 ± 0,02  | 0,98 ± 0,01  | 0,17                        | 0,27                        |
| Triglicéridos (mmol/L)       | 1,01 ± 0,01   | 1,04 ± 0,01   | 1,92 ± 0,01  | 2,20 ± 0,01  | 0,06                        | 0,14                        |
| ApoA-I (g/L)                 | 1,45 ± 0,01   | 1,45 ± 0,01   | 1,40 ± 0,02  | 1,39 ± 0,04  | 0,58                        | 0,94                        |
| Apo B (g/L)                  | 0,97 ± 0,01   | 0,97 ± 0,01   | 1,01 ± 0,03  | 0,96 ± 0,04  | 0,45                        | 0,45                        |
| HDL (mmol/L)                 |               |               |              |              |                             |                             |
| Grande                       | 0,61 ± 0,01   | 0,61 ± 0,02   | 0,55 ± 0,03  | 0,46 ± 0,01  | 0,05                        | 0,08                        |
| Intermedia                   | 0,55 ± 0,01   | 0,54 ± 0,01   | 0,57 ± 0,02  | 0,56 ± 0,02  | 0,94                        | 0,85                        |
| Pequeña                      | 0,18 ± 0,01   | 0,18 ± 0,01   | 0,14 ± 0,01  | 0,14 ± 0,02  | 0,52                        | 0,80                        |
| LDL (mmol/L)                 |               |               |              |              |                             |                             |
| Grande                       | 1,92 ± 0,03   | 2,04 ± 0,04   | 1,78 ± 0,09  | 1,43 ± 0,01  | 0,02                        | 0,01                        |
| Intermedia                   | 0,87 ± 0,02   | 0,86 ± 0,03   | 0,93 ± 0,07  | 0,75 ± 0,01  | 0,36                        | 0,24                        |
| Pequeña                      | 0,66 ± 0,02   | 0,61 ± 0,03   | 0,80 ± 0,07  | 1,16 ± 0,01  | 0,06                        | 0,06                        |
| Tamaño HDL (nm)              | 9,17 ± 0,01   | 9,18 ± 0,02   | 9,13 ± 0,04  | 9,03 ± 0,08  | 0,05                        | 0,18                        |
| Tamaño LDL (nm)              | 20,88 ± 0,02  | 20,93 ± 0,02  | 20,75 ± 0,05 | 20,52 ± 0,12 | 0,03                        | 0,04                        |
| Tamaño VLDL (nm)             | 46,58 ± 0,26  | 46,32 ± 0,44  | 50,84 ± 1,04 | 50,89 ± 1,60 | 0,53                        | 0,87                        |

Los resultados se dan como media ± error estándar (del modelo completamente ajustado: después de ajustar por relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de betabloqueantes, genotipo de apoE, y estado menopáusico y terapia de estrógenos (en mujeres).

<sup>1</sup>: Valor de P para el término de interacción entre el genotipo del exón 1 y la diabetes después de ajustar por relaciones familiares.

<sup>2</sup>: Valor de P para el término de interacción entre el genotipo del exón 1 y la diabetes en el modelo completamente ajustado.

**Tabla 4.30. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 8 de SR-BI para hombres y mujeres. Interacción con la diabetes 2**

| Genotipo del exón 8          |              |              |              |                |                |   |
|------------------------------|--------------|--------------|--------------|----------------|----------------|---|
|                              | CC (n=327)   | CT (n=556)   | TT (n=294)   | P <sup>1</sup> | P <sup>2</sup> | P <sup>3</sup> interacción con diabetes |
| <b>HOMBRES</b>               |              |              |              |                |                |   |
| IMC (Kg/m <sup>2</sup> )     | 27,48 ± 0,22 | 28,00 ± 0,19 | 27,51 ± 0,23 | 0,08           | 0,07           | 0,32                                    |
| Colesterol total (mmol/L)    | 5,22 ± 0,06  | 5,21 ± 0,05  | 5,10 ± 0,06  | 0,14           | 0,21           | 0,22                                    |
| LDL-C (mmol/L)               | 3,41 ± 0,05  | 3,38 ± 0,05  | 3,26 ± 0,05  | 0,07           | 0,10           | 0,26                                    |
| HDL-C (mmol/L)               | 1,09 ± 0,02  | 1,12 ± 0,01  | 1,15 ± 0,02  | 0,07           | 0,04           | 0,20                                    |
| HDL <sub>2</sub> -C (mmol/L) | 0,13 ± 0,01  | 0,13 ± 0,01  | 0,14 ± 0,01  | 0,27           | 0,32           | 0,53                                    |
| HDL <sub>3</sub> -C (mmol/L) | 0,97 ± 0,01  | 0,99 ± 0,01  | 1,01 ± 0,02  | 0,09           | 0,05           | 0,11                                    |
| Triglicéridos (mmol/L)       | 1,39 ± 0,01  | 1,33 ± 0,01  | 1,31 ± 0,01  | 0,46           | 0,41           | 0,37                                    |
| ApoA-I (g/L)                 | 1,33 ± 0,01  | 1,34 ± 0,02  | 1,34 ± 0,02  | 0,44           | 0,77           | 0,06                                    |
| Apo B (g/L)                  | 1,01 ± 0,01  | 0,99 ± 0,02  | 0,97 ± 0,01  | 0,19           | 0,08           | 0,40                                    |
| HDL (mmol/L)                 |              |              |              |                |                |   |
| Grande                       | 0,34 ± 0,02  | 0,40 ± 0,02  | 0,43 ± 0,02  | 0,00           | 0,00           | 0,99                                    |
| Intermedia                   | 0,57 ± 0,01  | 0,55 ± 0,01  | 0,54 ± 0,01  | 0,53           | 0,23           | 0,05                                    |
| Pequeña                      | 0,22 ± 0,01  | 0,22 ± 0,01  | 0,21 ± 0,01  | 0,76           | 0,91           | 0,40                                    |
| LDL (mmol/L)                 |              |              |              |                |                |   |
| Grande                       | 1,65 ± 0,06  | 1,77 ± 0,05  | 1,72 ± 0,06  | 0,18           | 0,20           | 0,93                                    |
| Intermedia                   | 1,00 ± 0,04  | 0,90 ± 0,03  | 0,87 ± 0,04  | 0,06           | 0,05           | 0,70                                    |
| Pequeña                      | 0,92 ± 0,05  | 0,85 ± 0,03  | 0,85 ± 0,05  | 0,48           | 0,47           | 0,90                                    |
| Tamaño HDL (nm)              | 8,87 ± 0,02  | 8,96 ± 0,02  | 8,97 ± 0,03  | 0,00           | 0,00           | 0,77                                    |
| Tamaño LDL (nm)              | 20,61 ± 0,04 | 20,70 ± 0,03 | 20,69 ± 0,04 | 0,13           | 0,12           | 0,89                                    |
| Tamaño VLDL (nm)             | 49,35 ± 0,61 | 49,17 ± 0,51 | 49,13 ± 0,61 | 0,89           | 0,96           | 0,25                                    |
| <b>MUJERES</b>               |              |              |              |                |                |   |
| IMC (Kg/m <sup>2</sup> )     | 26,12 ± 0,30 | 26,24 ± 0,24 | 26,03 ± 0,34 | 0,95           | 0,85           | 0,56                                    |
| Colesterol total (mmol/L)    | 5,32 ± 0,05  | 5,25 ± 0,04  | 5,17 ± 0,06  | 0,12           | 0,13           | 0,24                                    |
| LDL-C (mmol/L)               | 3,33 ± 0,05  | 3,23 ± 0,04  | 3,16 ± 0,05  | 0,08           | 0,04           | 0,48                                    |
| HDL-C (mmol/L)               | 1,44 ± 0,02  | 1,44 ± 0,02  | 1,47 ± 0,02  | 0,49           | 0,28           | 0,33                                    |
| HDL <sub>2</sub> -C (mmol/L) | 0,25 ± 0,01  | 0,25 ± 0,01  | 0,26 ± 0,01  | 0,47           | 0,47           | 0,69                                    |
| HDL <sub>3</sub> -C (mmol/L) | 1,19 ± 0,02  | 1,18 ± 0,01  | 1,21 ± 0,02  | 0,41           | 0,17           | 0,30                                    |
| Triglicéridos (mmol/L)       | 1,06 ± 0,01  | 1,08 ± 0,01  | 1,03 ± 0,01  | 0,32           | 0,33           | 0,09                                    |
| ApoA-I (g/L)                 | 1,54 ± 0,02  | 1,53 ± 0,01  | 1,54 ± 0,02  | 1,00           | 0,79           | 0,72                                    |
| Apo B (g/L)                  | 0,94 ± 0,01  | 0,65 ± 0,01  | 0,93 ± 0,01  | 0,16           | 0,34           | 0,18                                    |
| HDL (mmol/L)                 |              |              |              |                |                |   |
| Grande                       | 0,79 ± 0,02  | 0,81 ± 0,02  | 0,84 ± 0,02  | 0,02           | 0,04           | 0,26                                    |
| Intermedia                   | 0,54 ± 0,01  | 0,54 ± 0,01  | 0,52 ± 0,01  | 0,04           | 0,11           | 0,15                                    |
| Pequeña                      | 0,15 ± 0,01  | 0,14 ± 0,01  | 0,14 ± 0,01  | 0,23           | 0,23           | 0,08                                    |
| LDL (mmol/L)                 |              |              |              |                |                |   |
| Grande                       | 2,20 ± 0,05  | 2,12 ± 0,04  | 2,09 ± 0,05  | 0,39           | 0,32           | 0,59                                    |
| Intermedia                   | 0,81 ± 0,03  | 0,82 ± 0,02  | 0,76 ± 0,03  | 0,05           | 0,07           | 0,92                                    |
| Pequeña                      | 0,49 ± 0,03  | 0,47 ± 0,02  | 0,48 ± 0,03  | 0,38           | 0,67           | 0,25                                    |
| Tamaño HDL (nm)              | 9,37 ± 0,03  | 9,39 ± 0,02  | 9,46 ± 0,03  | 0,01           | 0,03           | 0,07                                    |
| Tamaño LDL (nm)              | 21,06 ± 0,03 | 21,06 ± 0,02 | 21,08 ± 0,03 | 0,68           | 0,86           | 0,56                                    |
| Tamaño VLDL (nm)             | 44,33 ± 0,51 | 44,78 ± 0,43 | 44,61 ± 0,51 | 0,66           | 0,77           | 0,17                                    |

Los resultados se dan como media ± error estándar (a partir del modelo 2 ajustado completamente)

<sup>1</sup>Después de ajustar por relación familiar

<sup>2</sup>Después de ajustar por relación familiar, edad, IMC, tabaco, ingesta de alcohol, uso de betabloqueantes, genotipo apoE, y estado menopáusico y terapia de estrógenos en mujeres.

<sup>3</sup>Valor de P para el término de interacción entre el genotipo del exón 8 y la diabetes en el modelo 2.

Para todos los polimorfismos en hombres y mujeres la distribución de genotipos era la esperada según el equilibrio de Hardy-Weinberg puesto que en ningún caso se han detectado diferencias estadísticamente significativas ( $P > 0,05$ ) entre las frecuencias observadas y esperadas. Tampoco se han encontrado diferencias estadísticamente significativas por el estado diabético o el sexo en las frecuencias genotípicas de las variantes genéticas del SR-BI (Tabla 4.27). Un análisis de asociación combinado de la distribución de genotipos ha indicado que los SNP del intrón 5 y del exón 8 estaban asociados ( $\chi^2 = 125$ ;  $p < 0,001$ ).

El análisis de haplotipos ha estimado que el parámetro  $D'$  de desequilibrio de ligamiento tenía los valores  $-0,74$  ( $p < 0,001$ ) para el exón 8 y el intrón 5;  $0,07$  ( $p = 0,005$ ) para el exón 1 y el intrón 5; y  $-0,083$  ( $p = 0,130$ ) para el exón 1 y el exón 8. Esto indicaba que existía un fuerte desequilibrio de ligamiento entre los polimorfismos del exón 8 y del intrón 5 en la población de estudio. El alelo T del intrón 5 se ha encontrado más a menudo con el alelo C del exón 8. No se ha encontrado diferencia estadísticamente significativa entre los polimorfismos del exón 1 y exón 8.

La Tabla 4.28 muestra las medias de IMC, lípidos plasmáticos, lipoproteínas, apolipoproteínas y tamaño de partícula por el polimorfismo del exón 1 del SR-BI en hombres y mujeres. Debido al corto número de sujetos, se han agrupado los homocigotos para el alelo más raro y los heterocigotos como portadores del alelo A. No se han encontrado diferencias estadísticamente significativas para el polimorfismo del exón 1 en ninguna de las variables examinadas en hombres y mujeres, incluso después de haber hecho un ajuste multivariado.

Según la hipótesis planteada, en que el estado diabético puede interactuar con la variación del SR-BI en la determinación de las concentraciones de lípidos plasmáticos y en el perfil de las subclases, se ha evaluado, para hombres y mujeres, la significación estadística de los términos de interacción entre la diabetes y el polimorfismo del exón 1 en los modelos de regresión correspondientes. Entonces, se han observado algunas interacciones significativas en hombres, e interacciones ligeramente significativas en mujeres. El efecto de las interacciones antes mencionadas se ha examinado mediante un análisis estratificado por diabetes y sexo.

Puesto que se ha encontrado una homogeneidad consistente en los efectos de interacción por sexo y, con el propósito de incrementar la potencia estadística de las estimaciones, se han analizado conjuntamente los datos de hombres y mujeres. La Tabla 4.29 muestra las concentraciones de lípidos plasmáticos, lipoproteínas, apolipoproteínas y tamaño de partículas por el polimorfismo del exón 1 del SR-BI

en hombres y mujeres según el estado diabético. En el análisis crudo, se han encontrado términos de interacción estadísticamente significativos para LDL-C, HDL-C; HDL<sub>2</sub>-C, gran LDL-C, y tamaño de partículas de HDL y LDL. Incluso después de haber aplicado un ajuste multivariado, permanecían estadísticamente significativos LDL-C ( $p = 0,028$ ); HDL<sub>2</sub>-C ( $p = 0,036$ ); gran LDL-C ( $p = 0,014$ ) y tamaño de partícula de LDL ( $p = 0,035$ ). Esto indicaba que el estado diabético modifica el efecto del polimorfismo del exón 1 del SR-BI en las concentraciones de lípidos plasmáticos y en la distribución de tamaños de partículas. Además, un ajuste adicional por glucosa ha producido una ligera reducción en los valores de  $P$  de los términos de interacción, oscilando entre  $-0,001$  y  $-0,006$  unidades. Sin embargo, no se han notado cambios en las inferencias estadísticas.

En los no diabéticos, no se han detectado diferencias en lípidos, aunque sí se ha observado ligeros incrementos en las concentraciones de HDL-C para los portadores del raro alelo A del polimorfismo del exón 1 del SR-BI. Mientras que en los diabéticos, la presencia del alelo A se ha asociado con un decremento medio del 6% en las concentraciones de HDL-C en comparación con los homocigotos GG. Esta reducción en HDL-C era más alta y significativamente estadística para las concentraciones de HDL<sub>2</sub>-C (-15%;  $p = 0,036$ ) y consistente con la reducción en el diámetro medio de partículas de HDL-C (-0,10 nm;  $p = 0,082$ ) observada en estos individuos. No se han observado en las concentraciones de apoAI. Por tanto, la presencia del alelo del exón 1 en sujetos diabéticos parece ser compatible con un incremento de la expresión génica del SR-BI, consistente con una internalización de colesterol selectiva. Se ha encontrado también que los sujetos diabéticos portadores del alelo A en el exón 1 del gen SR-BI tienen significativamente concentraciones de LDL-C más bajas (-13,8%) y tamaño medio de partículas de LDL reducido.

No se han observado asociaciones significativas entre ninguna de las variables examinadas en el polimorfismo del intrón 5 en hombres y mujeres (Tabla 4.30). Cuando se han evaluado los efectos de interacción con el estado diabético, se ha encontrado solamente un término de interacción estadísticamente significativo entre la diabetes de tipo 2 y el polimorfismo del intrón 5 en la determinación del tamaño de partícula de HDL en los hombres ( $p = 0,037$ ).

En el análisis estratificado por estado diabético, los hombres diabéticos portadores del alelo T en el intrón 5 del SR-BI presentan menores tamaños medios de partículas de HDL ( $8,8 \pm 0,09$  nm) que los hombres diabéticos homocigóticos para el alelo C ( $9,0 \pm 0,04$  nm). Sin embargo, la ausencia de consistencia de esta interacción en la mujeres, así como la ausencia de efectos de interacción estadísticamente significativos en otros parámetros lipídicos en hombres, no permite evaluar claramente su significado fisiológico. Cuando se han estimado

posteriormente los efectos de interacción del estado diabético para el polimorfismo del intrón 5 mediante la combinación de los datos de hombres y mujeres, para incrementar la potencia estadística, no se ha observado modificación significativa del efecto.

La Tabla 4.30 muestra las concentraciones medias de lípidos plasmáticos, lipoproteínas, apolipoproteínas y tamaños de partículas para el polimorfismo del exón 8 del SR-BI según el sexo. Después del ajuste por covariantes se ha encontrado una asociación consistente entre el polimorfismo del exón 8 y el tamaño de partícula de HDL. Ha destacado así una relación entre gen y dosificación donde el alelo T estaba asociado con un incremento de el diámetro de partículas de HDL en hombres ( $8,87 \pm 0,02$  nm para CC,  $8,96 \pm 0,02$  nm para CT y  $8,97 \pm 0,03$  nm para TT;  $p = 0,004$ ) y mujeres ( $9,37 \pm 0,03$  nm para CC,  $9,39 \pm 0,02$  nm para CT y  $9,46 \pm 0,03$  nm para TT;  $p = 0,031$ ). Este efecto codominante se ha observado claramente en las mujeres para la concentración de gran HDL ( $0,79 \pm 0,02$  mmol/L para CC,  $0,81 \pm 0,02$  mmol/L para CT y  $0,84 \pm 0,02$  mmol/L para TT;  $p = 0,045$ ), sin que se haya alcanzado la significación estadística en hombres ( $p = 0,198$ ). Sin embargo, en los hombres, se ha encontrado una asociación entre gen y dosificación estadísticamente significativa en concentraciones de HDL-C plasmático total ( $1,09 \pm 0,02$  mmol/L para CC,  $1,12 \pm 0,01$  mmol/L para CT y  $1,15 \pm 0,02$  mmol/L para individuos TT;  $p = 0,044$ ). Este incremento en el HDL-C plasmático total, así como en el diámetro de partículas de HDL en los individuos portadores del alelo T, sugiere que esta variante genética puede estar relacionada con una expresión génica reducida del SR-BI. Además, para hombres y mujeres se han reducido también las concentraciones de LDL-C en portadores del alelo T; sin embargo, las diferencias entre los tres genotipos del exón 8 eran solamente estadísticamente significativas en mujeres ( $p = 0,044$ ). Estos resultados no han variado sustancialmente cuando se han evaluado de acuerdo con el estado diabético, y no se ha observado ninguna interacción estadísticamente significativa entre la diabetes de tipo 2 y el polimorfismo del exón 8 en el análisis estratificado por sexo, o cuando la modificación del efecto por diabetes se ha analizado en los datos combinados para hombres y mujeres.

Finalmente, para considerar el efecto combinado de las variantes genéticas del SR-BI se ha realizado un análisis de asociación de haplotipos en 1.607 sujetos no sin lazos familiares: 1.456 no diabéticos (694 hombres y 762 mujeres) y 151 diabéticos (93 hombres y 58 mujeres), como se ha descrito en la correspondiente subsección de métodos. La Figura 4.24 muestra el efecto de los haplotipos más comunes del intrón 5 y el exón 8 (1/1 y 1/2, con la respectiva frecuencia de 0,42 y 0,47) sobre las concentraciones y el tamaño de partícula de HDL-C y LDL-C, después de haber hecho un ajuste multivariado.



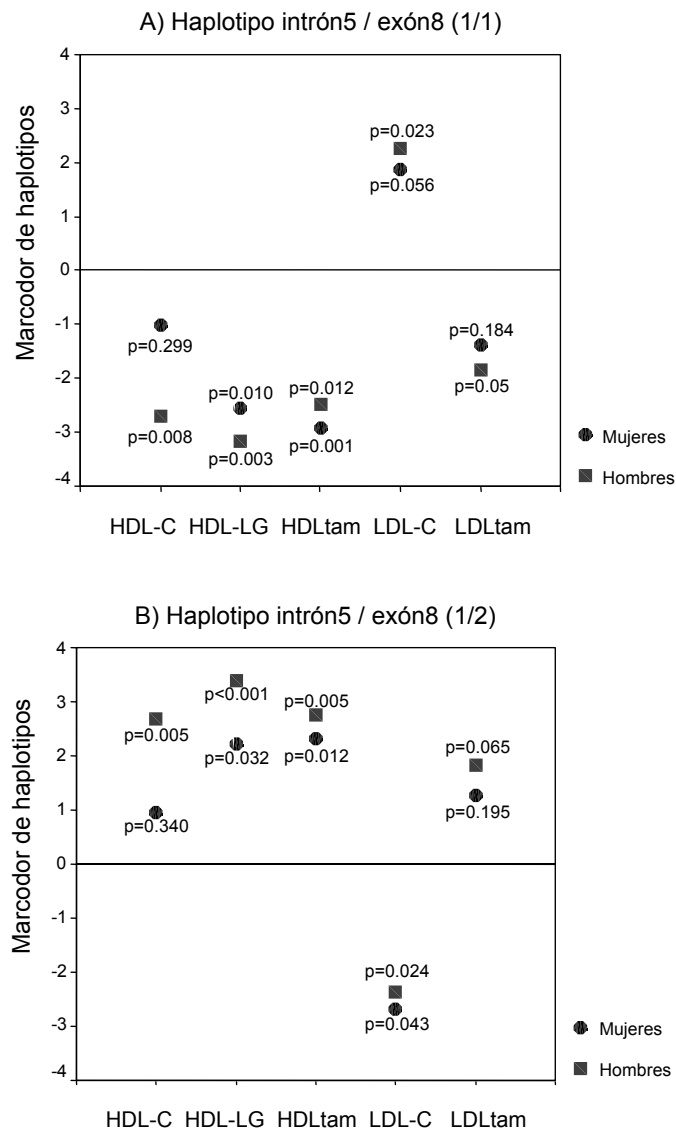


Figura 4.24. Análisis de Sensibilidad. Coeficientes de regresión e intervalos de confianza al 95% para los genotipos B1B2 y B2B2 respectivamente, comparados con B1B1 cuando cada variable indicada se ha ido incluyendo progresivamente en los modelos de regresión lineal. Modelo 1: genotipo CETP. Modelo 2: modelo 1 + sexo. Modelo 3: modelo 2 + IMC. Modelo 4: modelo 3 + consumo de tabaco. Modelo 5: modelo 4 + consumo de alcohol. Modelo 6: modelo 5 + genotipo APOE. Se ha incluido R<sup>2</sup> en la figura para mostrar la variabilidad considerada para cada modelo de regresión

Puesto que el estado diabético no interactúa con estas variantes genéticas, los sujetos diabéticos y no diabéticos se han analizado conjuntamente. Las mujeres y los hombres portadores del haplotipo 1/2 (C/T) tienen concentraciones más bajas de LDL-C y mayores tamaños de partículas de HDL estadísticamente significativas en comparación con los sujetos portadores del resto de haplotipos. Mientras que la situación opuesta es cierta para los sujetos portadores del haplotipo 1/1 (C/C).

Puesto que el estado diabético modifica el efecto del polimorfismo, se han examinado los haplotipos para las variantes genéticas del exón 1, intrón 5 y exón 8 en sujetos diabéticos y no diabéticos. El análisis de haplotipos ha confirmado los descubrimientos anteriores, sin embargo, el incremento en la especificidad del efecto estaba acompañada por una reducción en la potencia estadística para detectar diferencias estadísticamente significativas.

La Figura 4.25 muestra las concentraciones estimadas de LDL-C (A) y la tasa de gran HDL-C sobre pequeño HDL-C (B) por estado diabético dependiendo de los haplotipos más comunes exón 1/ intrón 5/ exón 8 (1/1/1, 1/1/2 y 2/1/1, con una frecuencia respectiva de 0,7, 0,42 y 0,06) después de ajuste multivariado. Se ha observado claramente la modificación del efecto de la variante genética del exón 1 sobre las concentraciones de LDL-C (a) y las concentraciones y tamaño de partículas de HDL-C (b) en sujetos diabéticos en comparación con los no diabéticos. Se han encontrado bajas concentraciones de LDL-C y una tasa reducida de gran HDL-C/ pequeño HDL-C, estadísticamente significativas en los sujetos con el haplotipo 2/1/1 (A/C/C) en comparación con los portadores del haplotipo 1/1/1 (G/C/C). Mientras que no se han encontrado diferencias cuando esta comparación se ha examinado en sujetos no diabéticos. Se ha observado claramente que la influencia de la variante genética del exón 8 sobre la concentración y tamaño de partícula de HDL-C, al examinar específicamente por el haplotipo 1/1/2 (C/C/T), en sujetos diabéticos y no diabéticos, no comporta evidencia de interacción.

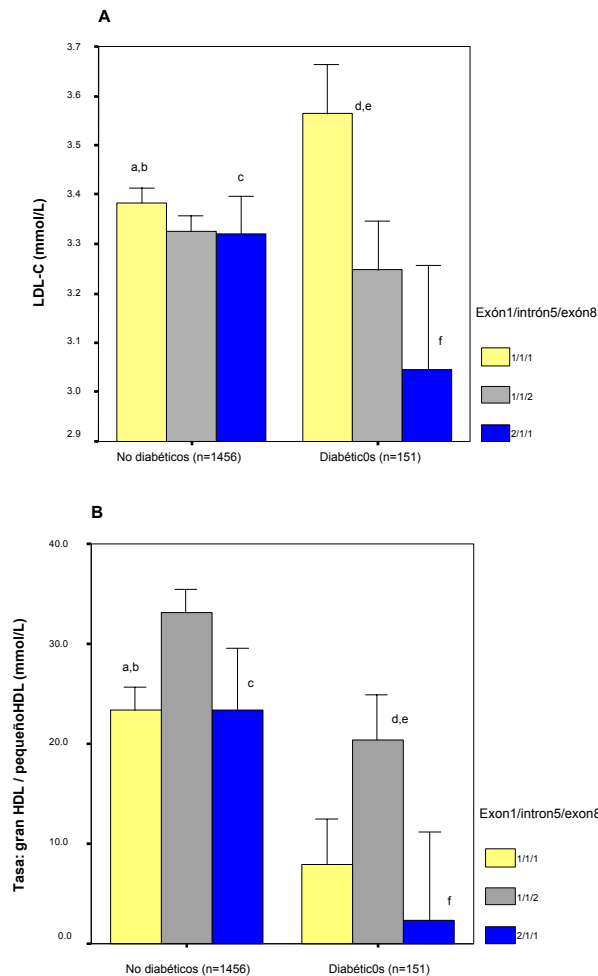


Figura 4.25. : Concentraciones estimadas de HDL-C (A) y la tasa de gran HDL-C / pequeño HDL-C (B) según la diabetes dependiente de los haplotipos más comunes exón 1/ intrón 5/ exón 8 (1/1/1, 1/1/2, y 2/1/1, con una frecuencia de 0,37, 0,42 y 0,06, respectivamente) después de ajustar por edad, sexo, IMC, tabaco, consumo de alcohol, genotipo de apoE, uso de betabloqueantes, y estado menopáusico y terapia por estrógenos en las mujeres según el procedimiento GLM.

Las barras de error son el SE de las medias. Las minúsculas indican la significación estadística de la media ajustada para cada una de las comparaciones como sigue: A. No diabéticos, a,  $P \leq 0,078$  para 1/1/2; b,  $P = 0,322$  para 2/1/1, en comparación respectiva con el tipo natural 1/1/1; c,  $P = 0,916$  para 1/1/2 en comparación con el haplotipo 2/1/1; diabéticos, d,  $P = 0,012$  para 1/1/2; y e,  $P \leq 0,045$  para 2/1/1, en comparación respectiva con el tipo natural 1/1/1; f,  $P = 0,435$  para 1/1/2 en comparación con 2/1/1. B, no diabéticos, a,  $P = 0,009$  para 1/1/2; y b,  $P = 0,917$  para 2/1/1, en comparación respectiva con el tipo natural 1/1/1; c,  $P = 0,156$  para 1/1/2 en comparación con 2/1/1; diabéticos, d,  $P = 0,041$  para 1/1/2; y e,  $P = 0,194$  para 2/1/1, en comparación con el 2/1/1.

### 4.3.6 Resultados del análisis del gen PLIN

Estos resultados se han remitido para su publicación a la revista *Clinical Genetics* en el año 2004 en un trabajo con el título “Genetic variation at the Perilipin (PLIN) locus is associated with obesity-related phenotypes in White women” (Qi et al., 2004), al que dicha revista le ha asignado el código de referencia CGE-00136-2004 y, previa aplicación de algunas modificaciones, ha aceptado finalmente con fecha de 26 de mayo de 2004. A continuación se presenta un resumen de los principales hallazgos.

Se analizaron un total de 1.589 individuos de la población general de la Comunidad Valenciana. Para buscar los polimorfismos en el locus del gen PLIN, en primer lugar se ha secuenciado la región 5' del gen PLIN en 40 sujetos sin lazos familiares para encontrar variantes potencialmente implicadas en la regulación de este gen. Se han concentrado la búsqueda específicamente en las regiones que son significativamente conservativas respecto del ratón y del hombre (Lu et al., 2001). Puesto que estos análisis no han revelado ninguna mutación común, el segundo paso ha sido la búsqueda en la base de datos pública dbSNP de los NIH estadounidenses. Se han seleccionado dianas iniciales según los criterios siguientes:

1. Se prefieren los SNP en exones frente a los de los intrones.
2. Si varios SNP se agrupan en una estrecha región, sólo uno de ellos se debe seleccionar.

Así, se han seleccionado inicialmente seis SNP (Tabla 4.31, Figura 4.26): PLIN1, PLIN2, PLIN3, PLIN4, PLIN5 y PLIN6. Dos de ellos (PLIN2 y PLIN3) no eran polimórficos con respecto a la población del estudio, por lo que los análisis se han centrado en el resto de SNP (PLIN1, PLIN4, PLIN5 y PLIN6).

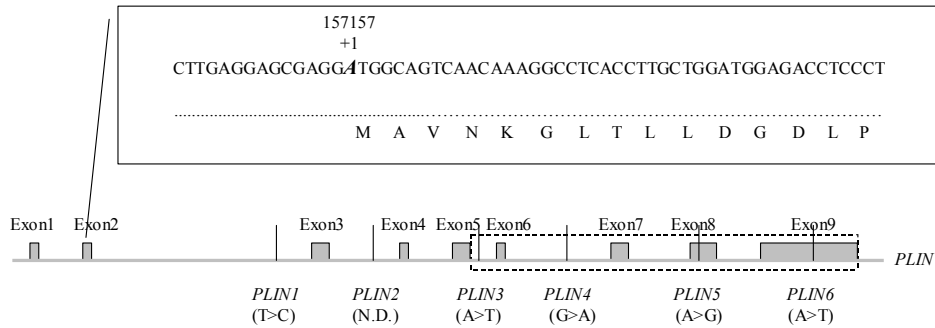


Figura 4.26. Nomenclatura de los polimorfismos de PLIN. Las posiciones de los polimorfismos examinados se indican con líneas verticales cortas, con los nombres debajo. El cuadrado debajo del diagrama del gen muestra la secuencia que acompaña al nucleótido denominado “+1” en esta nomenclatura. El A del codon iniciador ATG de la Metionina se indica en letra cursiva negrita, con la posición genómica en la secuencia de referencia (número de acceso de GenBank G12143119 etiquetado debajo). Los aminoácidos correspondientes también están incluidos. El cuadrado con barras indica la región donde puede ocurrir el splicing alternativo.



**Tabla 4.32. Características demográficas, bioquímicas y de estilo de vida de los sujetos del estudio**

|   | Población general |                 |
|---|-------------------|-----------------|
|   | Hombres (n=788)   | Mujeres (n=801) |
|   | Media (SD)        | Media (SD)      |
| Edad (años)                                     | 40,6 (11,6)       | 42,4 (14,8)*    |
| Peso corporal (kg)                              | 78,9 (11,1)       | 64,4 (12,7)*    |
| Talla (m)                                       | 1,73 (0,06)       | 1,59 (0,06)*    |
| IMC (kg/m <sup>2</sup> )                        | 26,4 (3,5)        | 25,7 (5,4)*     |
| Cintura (cm)                                    | 95,6 (11,1)       | 88,3 (15,4)*    |
| Cadera (cm)                                     | 100,8 (9,9)       | 102,0 (13,0)    |
| Razón cintura-cadera                            | 0,95 (0,07)       | 0,86 (0,07)*    |
| Glucosa en ayunas (mg/dL)                       | 92,6 (24,4)       | 96,1 (20,3)*    |
| Triglicéridos (mg/dL)                           | 129,5 (80,4)      | 94,5 (56,6)*    |
| Colesterol total (mg/dL)                        | 206,4 (38,8)      | 201,4 (38,4)*   |
| LDL-C (mg/dL)                                   | 134,7 (34,8)      | 128,1 (33,2)*   |
| HDL-C (mg/dL)                                   | 46,6 (9,8)        | 54,9 (11,5)*    |
| Presión sanguínea sistólica (mmHg)              | 124,7 (16,1)      | 123,2 (21,6)    |
| Presión sanguínea diastólica (mmHg)             | 75,6 (10,5)       | 74,6 (12,5)     |
| Obesidad (IMC≥30 kg/m <sup>2</sup> ) (%)        | 15,0              | 20,3*           |
| Sobrepeso (IMC≥25 kg/m <sup>2</sup> ) (%)       | 61,7              | 46,6*           |
| Sobrepeso (IMC≥35 kg/m <sup>2</sup> ) (%)       | 1,6               | 6,9*            |
| Obesos mórbidos (IMC≥40 kg/m <sup>2</sup> ) (%) | 0,3               | 1,9*            |
| Fumadores en activo (%)                         | 39,5              | 33,2*           |
| Consumidores de alcohol (%)                     | 90,6              | 56,8*           |
| Ejercicio físico (%)                            |                   |                 |
| Sedentarismo                                    | 36,3              | 58,4*           |
| Actividad                                       | 63,7              | 41,6            |
| Educación (%)                                   |                   |                 |
| Primaria  | 43,7              | 47,1*           |
| Secundaria                                      | 32,3              | 22,3            |
| Universitaria (ciclos I y II)                   | 24,0              | 30,5            |
| Diabetes tipo 2 (%)                             | 3,8               | 4,3             |
| Consumo de drogas lipopresoras (%)              | 5,7               | 8,1             |

SD: desviación estándar. C-Total: colesterol total. LDL-C: colesterol en lipoproteína de baja densidad. HDL-C: colesterol en lipoproteína de alta densidad. \*: P < 0,05 en la comparación entre hombres y mujeres. Test de Student para la comparación de medias y test de Chi-cuadrado para los porcentajes.

Formación universitaria por titulaciones: I: diplomado o ingeniero técnico (3 años). II: licenciado o ingeniero (5 años o más).

**Tabla 4.33. Distribución de genotipo, frecuencias alélicas y desequilibrio de ligamiento de las variantes genéticas polimórficas en el locus del PLIN en sujetos de la población general**

| Genotipos  | PLIN1 (6209T>C)     |            | PLIN4 (11482G>A)       |            | PLIN5 (13041A>G)       |            | PLIN6 (14995A>T)       |            |
|--|---------------------|------------|------------------------|------------|------------------------|------------|------------------------|------------|
|  | Hombres             | Mujeres    | Hombres                | Mujeres    | Hombres                | Mujeres    | Hombres                | Mujeres    |
|  | n (%)               | n (%)      | n (%)                  | n (%)      | n (%)                  | n (%)      | n (%)                  | n (%)      |
| 11   | 309 (40,8)          | 331 (42,4) | 405 (52,5)             | 451 (57,7) | 282 (36,2)             | 318 (40,5) | 328 (44,6)             | 346 (44,7) |
| 12   | 334 (44,1)          | 342 (43,8) | 307 (39,8)             | 271 (34,7) | 380 (48,7)             | 345 (43,9) | 321 (43,7)             | 333 (43,0) |
| 22   | 114 (15,1)          | 108 (13,8) | 60 (7,8)               | 59 (7,6)   | 118 (15,1)             | 122 (15,5) | 86 (11,7)              | 95 (12,3)  |
| <b>Frecuencia del alelo raro y 95% CI</b>                          |                     |            |                        |            |                        |            |                        |            |
| Alele 2  | 0.364 (0.347-0.381) |            | 0.262 (0.247-0.278)    |            | 0.385 (0.368-0.402)    |            | 0.337 (0.320-0.353)    |            |
| <b>Desequilibrio de ligamiento entre variantes : D', D'' y (p)</b> |                     |            |                        |            |                        |            |                        |            |
| PLIN1  | —                   |            | 0.159; 0.958 (p<0.001) |            | 0.033; 0.149 (p<0.001) |            | 0.085; 0.394 (p<0.001) |            |
| PLIN4  | —                   |            | —                      |            | 0.031; 0.191 (p<0.001) |            | 0.078; 0.453 (p<0.001) |            |
| PLIN5  | —                   |            | —                      |            | —                      |            | 0.066; 0.320 (p<0.001) |            |
| PLIN6  | —                   |            | —                      |            | —                      |            | —                      |            |

CI: intervalo de confianza.

Las diferencias por sexo entre genotipos no eran significativas para los polimorfismos PLIN1 (p = 0,727), PLIN4 (p = 0,097), PLIN5 (p = 0,142) o PLIN6 (p = 0,932). Por tanto, las frecuencias alélicas y el desequilibrio de ligamiento entre polimorfismos se ha estimado para hombres y mujeres.

D': coeficiente del desequilibrio de ligamiento.

D'': coeficiente del desequilibrio de ligamiento D estandarizado por el valor máximo que puede tomar (D/Dmax).





**Tabla 4.35. Índice de masa corporal (IMC) y fenotipos de obesidad según el tipo de portador de la variante del alelo 2 en cada uno de los polimorfismos de PLIN en sujetos de la población general. Medias en mujeres ajustadas por la edad**

|                          | PLIN1 (6209T>C) |               |       | PLIN4 (11482G>A) |               |       | PLIN5 (13041A>G) |               |       | PLIN6 (14995 A>T) |               |       |
|--------------------------|-----------------|---------------|-------|------------------|---------------|-------|------------------|---------------|-------|-------------------|---------------|-------|
|                          | 11 (n=331)      | 12+22 (n=450) |       | 11 (n=451)       | 12+22 (n=330) |       | 11 (n=318)       | 12+22 (n=467) |       | 11 (n=346)        | 12+22 (n=428) |       |
|                          | Media (SE)      | Media (SE)    | P     | Media (SE)       | Media (SE)    | P     | Media (SE)       | Media (SE)    | P     | Media (SE)        | Media (SE)    | P     |
| IMC (kg/m <sup>2</sup> ) | 26,3 (0,3)      | 25,3 (0,2)    | 0,004 | 26,1 (0,2)       | 25,2 (0,3)    | 0,004 | 25,8 (0,3)       | 25,7 (0,2)    | 0,965 | 25,9 (0,4)        | 25,7 (0,2)    | 0,456 |
| Weight (Kg)              | 65,7 (0,6)      | 63,5 (0,5)    | 0,007 | 65,4 (0,6)       | 63,2 (0,6)    | 0,011 | 64,5 (0,6)       | 64,4 (0,5)    | 0,844 | 64,9 (0,6)        | 64,2 (0,6)    | 0,385 |
| Waist-to-hip ratio       | 0,86 (0,01)     | 0,86 (0,01)   | 0,519 | 0,87 (0,01)      | 0,85 (0,01)   | 0,032 | 0,86 (0,01)      | 0,87 (0,01)   | 0,172 | 0,87 (0,01)       | 0,86 (0,01)   | 0,299 |
| Glucose (mg/dL)          | 97,8 (0,9)      | 95,5 (0,9)    | 0,090 | 97,9 (0,8)       | 94,5 (1,0)    | 0,008 | 96,8 (0,9)       | 96,6 (0,8)    | 0,862 | 96,9 (0,9)        | 96,7 (0,9)    | 0,908 |
| Total-C (mg/dL)          | 202,1 (1,8)     | 201,1 (1,6)   | 0,652 | 201,3 (1,6)      | 201,4 (1,8)   | 0,962 | 201,1 (1,7)      | 202,3 (1,6)   | 0,645 | 200,8 (1,8)       | 202,3 (1,6)   | 0,650 |
| LDL-C (mg/dL)            | 127,9 (1,8)     | 128,6 (1,5)   | 0,761 | 127,1 (1,5)      | 129,9 (1,7)   | 0,222 | 127,8 (1,8)      | 128,9 (1,5)   | 0,653 | 127,7 (1,7)       | 129,2 (1,6)   | 0,442 |
| HDL-C (mg/dL)            | 54,3 (0,6)      | 54,8 (0,5)    | 0,498 | 54,2 (0,5)       | 55,0 (0,6)    | 0,361 | 54,1 (0,6)       | 54,9 (0,5)    | 0,245 | 53,8 (0,6)        | 55,1 (0,6)    | 0,120 |
| TAG (mg/dL)              | 99,5 (3,0)      | 95,1 (2,6)    | 0,099 | 102,5 (2,6)      | 89,4 (2,9)    | 0,005 | 102,0 (3,0)      | 95,4 (2,6)    | 0,207 | 100,1 (2,9)       | 95,3 (2,6)    | 0,314 |
| SBP (mmHg)               | 124,2 (0,9)     | 122,0 (0,8)   | 0,097 | 123,5 (0,8)      | 121,9 (0,9)   | 0,198 | 122,7 (0,9)      | 123,7 (0,8)   | 0,433 | 123,2 (0,9)       | 122,6 (0,8)   | 0,624 |
| DBP (mmHg)               | 75,5 (0,6)      | 74,1 (0,5)    | 0,105 | 74,8 (0,5)       | 74,6 (0,6)    | 0,841 | 74,4 (0,6)       | 75,0 (0,5)    | 0,410 | 74,4 (0,6)        | 75,9 (0,5)    | 0,562 |

SE: error estándar.

C-Totai: colesterol total. LDL-C: colesterol en lipoproteína de baja densidad. HDL-C: colesterol en lipoproteína de alta densidad. PAS: Presión arterial sistólica. PAD: Presión arterial diastólica.

Se ha ajustado el peso adicionalmente por la talla.

**Tabla 4.36. Prevalencia de los portadores del alelo 2 en PLIN1 y PLIN4 en sujetos obesos (casos) y no obesos (controles), y riesgo (OR y 95% CI) de obesidad en portadores del alelo 2 en comparación con los homocigotos del alelo común (11)**

| Gene variant               | Gender         | Obesos (casos)          |       | No obesos (controles)    |       | No ajustado <sup>1</sup> |        | Ajuste multivariado <sup>2</sup> |       |
|----------------------------|----------------|-------------------------|-------|--------------------------|-------|--------------------------|--------|----------------------------------|-------|
|                            |                | IMC<30kg/m <sup>2</sup> | n (%) | IMC>=30kg/m <sup>2</sup> | n (%) | OR (95%CI)               | p      | OR (95%CI)                       | p     |
| <b>PLIN1 (6209T&gt;C)</b>  | <b>HOMBRES</b> |                         |       |                          |       |                          |        |                                  |       |
| 11                         |                | 42                      |       | 266                      |       | 1                        |        | 1                                |       |
| 12+22                      |                | 71 (62,8)               |       | 378 (58,0)               |       | 1,18 (0.79-1.80)         | 0,410  | 1,01 (0.65-1.75)                 | 0,799 |
| <b>PLIN4 (11482G&gt;A)</b> |                |                         |       |                          |       |                          |        |                                  |       |
| 11                         |                | 62                      |       | 343                      |       | 1                        |        | 1                                |       |
| 12+22                      |                | 53 (46,1)               |       | 314 (47,0)               |       | 0,98 (0.70-1.48)         | 0,942  | 1,08 (0.66-1.76)                 | 0,756 |
| <b>PLIN1 (6209T&gt;C)</b>  | <b>MUJERES</b> |                         |       |                          |       |                          |        |                                  |       |
| 11                         |                | 84                      |       | 245                      |       | 1                        |        | 1                                |       |
| 12+22                      |                | 75 (47,2)               |       | 380 (60,0)               |       | 0,57 (0.40-0.81)         | 0,002  | 0,58 (0.38-0.93)                 | 0,023 |
| <b>PLIN4 (11482G&gt;A)</b> |                |                         |       |                          |       |                          |        |                                  |       |
| 11                         |                | 112                     |       | 337                      |       | 1                        |        | 1                                |       |
| 12+22                      |                | 48 (30,2)               |       | 283 (45,6)               |       | 0,52 (0.36-0.75)         | <0,001 | 0,56 (0.36-0.89)                 | 0,016 |

1: Modelo de regresión logística crudo

2: Modelo de regresión logística crudo ajustado por edad, origen, tabaco, alcohol, actividad física, diabetes y educación.

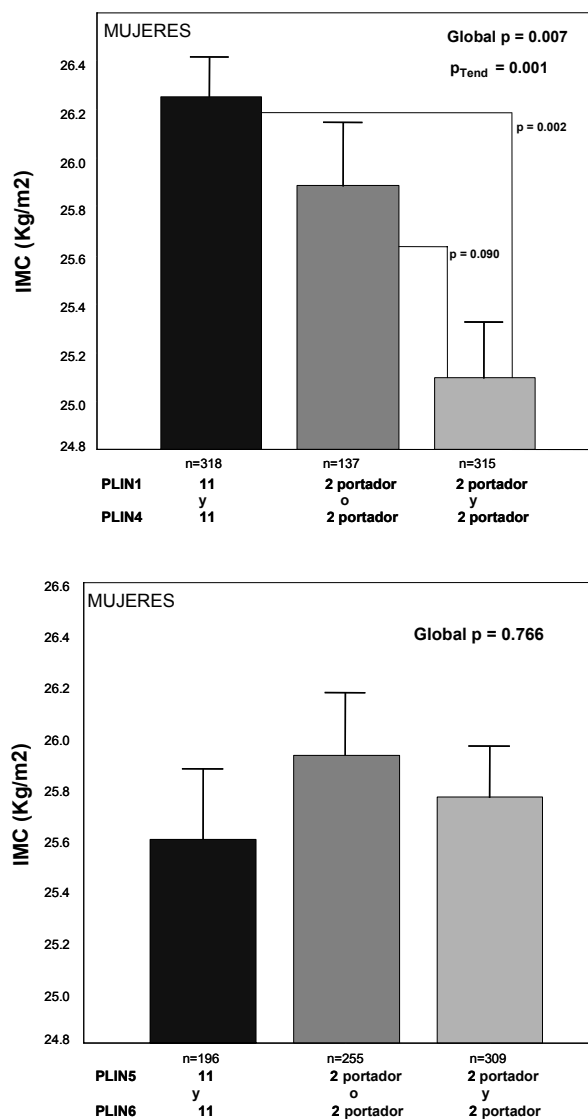


Figura 4.27. Medias de IMC dependientes de los genotipos combinados en el locus de PLIN en mujeres de la población general. Las medias ajustadas por edad de las combinaciones los SNP de PLIN1 y PLIN4 después de controlar por PLIN5 y PLIN6 (A), y combinaciones de lo SNP de PLIN5 y PLIN6 después de controlar por PLIN1 y PLIN4 (B). El valor global de P corresponde a la ANOVA ajustada multivariada para la comparación entre genotipos; P para tendencia es el valor de P en el test para la tendencia lineal entre genotipos combinados con el propósito de comprobar el efecto gen-dosis. Los valores de P en la comparación entre cada genotipo combinado se han obtenido después de la corrección de Bonferroni para comparaciones múltiples. Las barras de error: SE de las medias

En la Tabla 4.32 se presenta un resumen de las características demográficas, bioquímicas y de estilo de vida de los 1.589 sujetos caucásicos de la población general de la Comunidad Valenciana participantes en el estudio. En esta población, la prevalencia de obesidad mórbida era muy baja (sólo un 1% de los sujetos tenían un  $IMC \geq 40 \text{ Kg/m}^2$ ).

#### 4.3.6.1 Frecuencias alélicas y desequilibrio de ligamiento en sujetos de la población general

En la Tabla 4.33 se presentan los genotipos PLIN, las frecuencias alélicas y los coeficientes de desequilibrio de ligamiento para los sujetos de la población general. Las distribuciones de genotipos no se desviaban de lo esperado según el equilibrio de Hardy-Weinberg. Como las diferencias por sexo en las distribuciones genotípicas no eran significativas para ningún polimorfismo, se han analizado conjuntamente los datos para hombres y mujeres.

El alelo 2 (G) en el locus PLIN5 ha sido la variante genética más prevalente, mientras que el alelo 2 (A) en el locus PLIN4 ha sido la menos prevalente. El desequilibrio de ligamiento paritario más fuerte se han encontrado entre los polimorfismos PLIN1 y PLIN4 ( $D'$ : 0,958;  $p < 0,001$ ). A pesar de ser estadísticamente significativo, se ha observado un desequilibrio de ligamiento mucho menor entre los otros polimorfismos, con los coeficientes  $D'$  oscilando entre 0,453 a 0,149 (Tabla 4.33). Se han estimado todos de los posibles 16 haplotipos tetrapolimórficos que están presentes en esta población Mediterránea. El haplotipo compuesto por los alelos más frecuentes de cada polimorfismo (“6209T/11482G/13041A/14995A”; que se mencionará a partir de ahora como “1111”) ha sido el más prevalente, con una frecuencia relativa de 0,39; seguido por el haplotipo (“6209T/11482G/13041G/14995A”), con una frecuencia de 0,13. Del resto de los catorce haplotipos, sólo 3 tenían una frecuencia mayor de 0,08, incluyendo el haplotipo compuesto por los alelos menos frecuentes de cada polimorfismo (“6209C/11482A/13041G/14995T”), con una frecuencia de 0,081.

#### 4.3.6.2 Asociación entre los polimorfismos de PLIN y los fenotipos de obesidad en la población general. Análisis de genotipos de polimorfismos simples

Con el propósito de incrementar la potencia estadística y después de haber verificado la presencia de un efecto alélico compatible con un modelo dominante, o al menos, codominante, se han clasificado los individuos como homocigotos para el alelo más común (1/1) o como portadores del alelo menos común (1/2 + 2/2) para cada SNP. Además, se ha evaluado la homogeneidad del efecto genético por sexo y se han descubierto varias interacciones significativas. A continuación, se

han analizado los hombres y las mujeres por separado. La Tabla 4.34 muestra las medias ajustadas por edad de IMC y otras variables relacionadas con la obesidad para la población en general según el tipo de portador de la variante del alelo 2 con cada uno de los cuatro polimorfismos del PLIN. No se han encontrado diferencias significativas entre grupos de genotipos considerando IMC, peso, razón de cintura/cadera, glucosa, colesterol total, HDL-C, LDL-C, TG y presión sanguínea.

Sin embargo, se ha encontrado que en mujeres (Tabla 4.35) el IMC difería significativamente ( $p = 0,004$ ) entre los fenotipos para los polimorfismos PLIN1 (6209T>C) y PLIN4 (11482G>A), con el alelo 2 asociado con bajo IMC. De forma similar, los portadores del alelo 2 en el locus PLIN1 tenían un peso significativamente menor (-2,2 Kg;  $p = 0,007$ ) que las mujeres homocigotas para el genotipo natural. Lo mismo ha ocurrido con los portadores del alelo menos frecuente en PLIN4 ( $p = 0,01$ ). Además, las mujeres portadores del alelo 2 para el polimorfismo PLIN4 mostraban menores tasas cintura-cadera ( $p = 0,032$ ), glucosa en ayunas más baja ( $p = 0,008$ ) y menores concentraciones de TG ( $p = 0,005$ ) que las homocigotas 1/1. Se han encontrado diferencias similares para PLIN1, pero los valores de P umbrales con 0,090 para glucosa en ayunas y 0,099 para TG, revelaban un efecto mayor de PLIN4 en estos rasgos en comparación con PLIN1.

Cuando se ha comprobado específicamente en el modelo estadístico la diferencia en las asociaciones en hombres y mujeres, los dos SNP, PLIN1 y PLIN4, han hecho aflorar interacciones gen-gen en la determinación de IMC y peso corporal ( $p < 0,05$ ). Además, se han encontrado interacciones gen\*gen en la determinación de la tasa cintura-cadera ( $p = 0,023$ ) y TG ( $p = 0,009$ ) para PLIN4. Pero no se han detectado interacciones gen\*gen significativas para los polimorfismos PLIN5 y PLIN6.

Los portadores y no portadores del alelo 2 para cada polimorfismo no tenían diferencias significativas con respecto al consumo de tabaco y alcohol, educación, actividad física, medicación y diabetes en hombres y mujeres. Por lo tanto, las asociaciones encontradas para PLIN1 y PLIN4 en mujeres permanecían estadísticamente significativas después de haber ajustado por estas variables de confusión potenciales:  $p = 0,012$  para IMC y  $p = 0,020$  peso con PLIN1;  $p = 0,014$  para IMC,  $p = 0,029$  para peso,  $p = 0,046$  para tasa cintura-cadera,  $p = 0,003$  para glucosa y  $p = 0,042$  para TG, con PLIN4. Un ajuste adicional para IMC no ha modificado la significación de las asociaciones entre PLIN4 y glucosa en ayunas ( $116,4 \pm 1,3$  mg/dL en no portadores frente a  $113,7 \pm 1,7$  mg/dL en portadores del alelo 2 ( $p = 0,010$ ). Sin embargo, las diferencias entre las concentraciones de TG no conservaban su significación estadística ( $p = 0,327$ ).

#### 4.3.6.3 Asociación entre los polimorfismos de PLIN y los fenotipos de obesidad en la población general. Análisis de genotipos combinados

También se ha evaluado el efecto de la combinación de los genotipos de PLIN sobre IMC, TG y glucosa en ayunas. Como once de los dieciséis haplotipos posibles han aparecido con una frecuencia relativa muy baja (por debajo del 5%), se ha utilizado un enfoque de pseudohaplotipos comparando el efecto de la homocidad para el haplotipo más común con el efecto de una combinación de genotipos seleccionados.

En primer lugar, los resultados de las Tablas 4.34 y 4.35 se han ajustado para el efecto potencial de los otros polimorfismos mediante la inclusión de estas variables como factores de control en los modelos de regresión múltiple. Teniendo en cuenta la alta asociación entre PLIN1 y PLIN4, estas variables no se han ajustado simultáneamente uno por la otra para evitar el sesgo de multicolinealidad (de tolerancia baja en el modelo estadístico). Por tanto, las asociaciones de PLIN1 y PLIN4 se han ajustado complementariamente para PLIN5 y PLIN6; PLIN5 para PLIN4 y PLIN6; y PLIN6 para PLIN4 y PLIN5. La asociación entre PLIN1 e IMC en mujeres permanecía estadísticamente significativa después de estos ajustes ( $p = 0,002$ ).

Además, la asociación de PLIN1 con glucosa en ayunas en mujeres, bordeando la significación estadística, alcanzaba dicha significación después de haber ajustado por estos polimorfismos ( $p = 0,032$ ). De forma similar, para PLIN4 las asociaciones previamente descritas en la Tabla 4.35 permanecían estadísticamente significativas después del ajuste simultáneo por PLIN5 y PLIN6 ( $p = 0,023$  para IMC;  $p = 0,015$  para glucosa en ayunas; y  $p = 0,035$  para TG). En los hombres no se han detectado variaciones significativas cuando los resultados de la Tabla 4.35 se han ajustado adicionalmente por el resto de los polimorfismos.

Se han investigado también las asociaciones sinérgicas potenciales entre una combinación de parejas de SNP selectas y variables relevantes. En primer lugar, se ha examinado la combinación de PLIN1 y PLIN4. Los sujetos de la población general se han agrupado en tres categorías:

1. Homocigotos para el alelo 1 en PLIN1 y PLIN4.
2. Portadores del alelo 2, o en PLIN1 o en PLIN4.
3. Portadores del alelo 2 en PLIN1 y PLIN4.

No había asociación entre ninguna de las variantes combinadas y los fenotipos de obesidad en los hombres. La Figura 4.27 muestra las medias de IMC ajustadas por edad según los genotipos combinados de PLIN1 y PLIN4 en mujeres. El

modelo se ha ajustado además por PLIN5 y PLIN6. La variable de combinación de parejas de SNP estaba asociada significativamente con IMC ( $p = 0,007$ ), con mujeres homocigotas para el haplotipo más común "11". Estas mostraban mayores IMC ( $26,3 \pm 0,3 \text{ Kg/m}^2$ ;  $p = 0,002$ ) que las mujeres portadoras de al menos un alelo 2 en los SNP de PLIN1 y PLIN4 ( $25,1 \pm 0,3 \text{ Kg/m}^2$ ). Los portadores de al menos un alelo 2 en PLIN1 o en PLIN4 tenían IMC intermedio. Se han encontrado también asociaciones estadísticamente significativas entre la variable de SNP combinados y TG ( $p = 0,020$ ) y glucosa ( $p = 0,040$ ), presentando las mayores concentraciones los homocigotos del haplotipo más común. Cuando este análisis de genotipos combinados se ha realizado con los polimorfismos PLIN5 y PLIN6, después de haber controlado adicionalmente por PLIN1 y PLIN4, no se han detectado en las mujeres asociaciones con IMC (Figura 4.27B) o con ningún parámetro de obesidad.

#### **4.3.6.4 Estimación del riesgo de obesidad asociado con la variación del locus PLIN. Un enfoque de casos y controles en la población global**

Para estimar el riesgo de obesidad asociado con las variantes de PLIN, se han clasificado los sujetos como no obesos ( $\text{IMC} < 30 \text{ Kg/m}^2$ ) y obesos ( $\text{IMC} \geq 30 \text{ Kg/m}^2$ ). En los hombres no se han detectado diferencias en la prevalencia de cualquier polimorfismo de PLIN entre obesos y no obesos (Tabla 4.36). Sin embargo, las mujeres obesas tenían una prevalencia menor (50,2%) de portadoras del alelo 2 en PLIN1, en comparación con mujeres no obesas (60,4%;  $p = 0,004$ ).

Puesto que los obesos y no obesos diferían en edad, en el modelo de regresión logística se ha ajustado la OR por edad (Tabla 4.36). Después de este ajuste, las mujeres portadoras del alelo 2 en PLIN1 tenían un menor riesgo de obesidad en comparación con las no portadoras (OR: 0,65; 95%CI (0,48-0,88)). De forma similar, la prevalencia de las mujeres portadoras del alelo 2 en PLIN4 era menor en el grupo de obesos que en los no obesos (32,5% contra 45,2%;  $p < 0,001$ ). Después del ajuste por edad, el alelo 2 en el locus de PLIN4 estaba asociado consistentemente con un menor riesgo de obesidad en mujeres (OR: 0,60; 95%CI (0,44-0,83)). Además, estas estimaciones permanecían estadísticamente significativas después de posteriores ajustes por consumo de tabaco, de alcohol, actividad física, diabetes, educación y origen.

En el análisis de genotipos combinados de dos polimorfismos y después de haber ajustado por edad, las mujeres portadoras del alelo 2 en los SNP de PLIN1 y PLIN4 presentaban el menor riesgo de obesidad (OR: 0,56 (95% CI 0,39-0,79);  $p = 0,001$ ), en comparación con las homocigotas para los alelos más comunes. Estos resultados no cambiaban después de posteriores ajustes multivariados incluyendo controles por los polimorfismos PLIN 5 y PLIN6.



## 4.4 Auditoría Bioinformática

Estos resultados se han publicado en la revista *I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud* en el año 2003 con los títulos “Auditoría Bioinformática” (Coltell, 2003-AI) y “Sistemas de información en laboratorios de investigación en Biomedicina” (Coltell et al., 2003-SI). A continuación se presenta un resumen de los principales hallazgos.

### 4.4.1 Análisis de riesgos

El análisis preliminar de riesgos se inicia con un análisis de correspondencia entre las áreas de auditoría y auditorías tecnológicas y los elementos del Espacio del Problema de Auditoría (EEPA) identificados en la Tabla 1.4 (del Capítulo 3) . Así, se puede ver qué elementos son auditables por qué tipo de auditoría. La Tabla 4.37 establece la relación obtenida en el desarrollo del proyecto. Las áreas de auditoría se distinguen mediante la inicial “A” delante del término correspondiente. Las auditorías tecnológicas se distinguen mediante la inicial “T” delante del término correspondiente. La lista de tipos de auditorías está compuesta por 18 variantes, pero no es exhaustiva, de forma que se pueden incorporar o eliminar las variantes que no sean interesantes para el estudio.

Una vez efectuado el análisis para la identificación de riesgos, según los criterios de información e investigación para Bioinformática, el resultado es la asociación de riesgos a los elementos del Espacio del Problema de Auditoría (EEPA) y que se resume en la Tabla 4.38. Las intersecciones en la fila que ocupa cada uno de los EEPA indican los riesgos directos sobre dichos elementos. Por ejemplo, el elemento EAB01 tiene riesgos directos sobre la adecuación teórica y profesional.

**Tabla 4.37. Áreas de Auditoría y Auditorías Tecnológicas que se relacionan con los elementos del Espacio del Problema de Auditoría identificados**

| Áreas de Auditoría y Auditorías Tecnológicas | A. Organización y gestión | A. Producción y explotación | A. Seguridad física | A. Seguridad lógica | A. Calidad | A. Jurídica y Legal | T. Microordenadores | T. Ofimática | T. Bases de datos | T. Sistemas operativos | T. Proceso distribuido | T. Redes | T. Desarrollo proyectos | T. Aplicaciones | T. EIS/DSS | T. EDI | T. Entornos CASE | T. Sistemas expertos | Num. Elem./Audit. |
|--|---------------------------|-----------------------------|---------------------|---------------------|------------|---------------------|---------------------|--------------|-------------------|------------------------|------------------------|----------|-------------------------|-----------------|------------|--------|------------------|----------------------|-------------------|
| EAB01  |                           |                             |                     | *                   |            |                     |                     |              |                   |                        |                        |          |                         |                 |            |        |                  |                      | 1                 |
| EAB02  |                           |                             |                     | *                   |            |                     |                     |              |                   |                        |                        |          |                         |                 |            |        |                  |                      | 1                 |
| EAB03  |                           |                             |                     | *                   |            |                     |                     |              |                   |                        |                        |          | *                       |                 |            |        |                  |                      | 2                 |
| EAB04  |                           |                             |                     | *                   |            |                     |                     |              |                   |                        |                        |          |                         |                 |            |        |                  |                      | 1                 |
| EAB05  |                           |                             |                     |                     |            |                     |                     |              |                   |                        |                        |          |                         |                 |            |        |                  |                      | 0                 |
| EAB06  |                           |                             |                     | *                   | *          |                     |                     |              | *                 |                        |                        |          |                         | *               |            |        |                  |                      | 4                 |
| EAB07  |                           |                             |                     |                     |            |                     |                     |              |                   |                        |                        |          |                         |                 |            |        |                  |                      | 0                 |
| EAB08  |                           |                             | *                   | *                   |            | *                   |                     |              | *                 | *                      | *                      | *        | *                       | *               |            |        |                  |                      | 7                 |
| EAB09  |                           |                             | *                   | *                   |            | *                   |                     |              | *                 | *                      | *                      | *        | *                       | *               |            |        |                  |                      | 8                 |
| EAB10  |                           |                             |                     |                     |            | *                   |                     |              |                   |                        |                        |          | *                       |                 | *          |        |                  |                      | 3                 |
| EAB11  |                           |                             | *                   | *                   |            | *                   | *                   | *            | *                 | *                      | *                      | *        | *                       | *               | *          | *      | *                | *                    | 15                |
| EAB12  |                           |                             |                     |                     |            |                     | *                   | *            | *                 | *                      | *                      | *        | *                       | *               | *          | *      | *                | *                    | 12                |
| EAB13  |                           |                             |                     | *                   |            | *                   | *                   | *            | *                 | *                      | *                      | *        | *                       | *               | *          | *      | *                | *                    | 13                |
| EAB14  | *                         | *                           |                     | *                   | *          |                     | *                   | *            | *                 | *                      | *                      | *        | *                       | *               |            |        | ✓                | ✓                    | 14                |
| EAB15  | *                         | *                           | *                   | *                   | *          | *                   | *                   |              | *                 | *                      | *                      | *        | *                       | *               |            |        |                  |                      | 10                |
| EAB16  |                           | *                           | *                   | *                   |            |                     |                     |              |                   |                        |                        |          | *                       | *               |            |        |                  |                      | 5                 |
| EAB17  | *                         |                             |                     |                     |            |                     |                     |              |                   |                        |                        |          |                         |                 | *          |        |                  |                      | 2                 |
| EAB18  | *                         | *                           |                     | *                   |            |                     |                     |              |                   |                        |                        |          | *                       | *               |            |        |                  |                      | 5                 |
| EAB19  | *                         |                             |                     |                     |            |                     |                     |              |                   |                        |                        |          |                         |                 | *          |        |                  |                      | 2                 |
| EAB20  | *                         | *                           |                     | *                   |            | *                   | *                   | *            | *                 |                        |                        |          | *                       | *               |            |        |                  |                      | 8                 |
| EAB21  |                           | *                           |                     | *                   |            |                     |                     |              |                   |                        |                        |          | *                       | *               |            |        |                  |                      | 3                 |
| EAB22  | *                         | *                           |                     | *                   |            | *                   | *                   | *            | *                 | *                      | *                      | *        | *                       | *               |            | *      |                  |                      | 11                |
| EAB23  | *                         | *                           |                     |                     |            |                     |                     |              |                   |                        |                        |          | *                       |                 |            |        | *                |                      | 4                 |
| EAB24  | *                         | *                           |                     | *                   |            |                     |                     |              |                   |                        |                        |          |                         |                 |            |        | *                | *                    | 5                 |
| EAB25  | *                         | *                           |                     |                     |            |                     |                     |              |                   |                        |                        |          | *                       |                 |            |        |                  |                      | 3                 |
| Núm. Audit./Elem.                            | 10                        | 10                          | 5                   | 4                   | 14         | 5                   | 7                   | 6            | 8                 | 8                      | 8                      | 6        | 16                      | 11              | 6          | 4      | 6                | 5                    |                   |



Los criterios de información son los establecidos en el COBIT. Sin embargo, los criterios de investigación se han tenido que fijar específicamente para este contexto según las buenas prácticas de investigación científica. Estos criterios son los siguientes:

1. Adecuación teórica: Los enfoques teóricos científicos aplicados son los que realmente explican y predicen con mayor grado el problema.
2. Adecuación metodológica: Las metodologías científicas seleccionadas son las más adecuadas según el marco teórico y tecnológico (Sackman, 1997).
3. Adecuación técnica: Las técnicas aplicadas son las más adecuadas según el marco teórico y metodológico (Bernal et al., 1999) (Cornell et al., 2002) (Ermolaeva et al., 1998).
4. Adecuación profesional: El científico y el técnico, que intervienen en la investigación, tienen la formación científica, técnica y tecnológica adecuada.
5. Ética científica: El científico y el técnico, que intervienen en la investigación, procuran usar la información adquirida, los recursos de investigación y los resultados obtenidos conforme a las normas éticas vigentes (Sackman, 1997).

El siguiente paso consiste en la aplicación del proceso metodológico del COBIT para construir un modelo de desarrollo de auditoría para la Bioinformática. El modelo consiste en un subconjunto de objetivos de control y de guías de auditoría para auditar dichos objetivos. En primer lugar, se construye el modelo en función de los dominios y los procesos de cada dominio. En este modelo se busca la adscripción a un dominio y un proceso determinado de cada una de los EEPA. El proceso concreto es el siguiente:

1. Asignación de los elementos del Espacio del Problema de Auditoría a los Dominios de COBIT correspondientes (ISACAF-B, 2000), según los factores de riesgo identificados en la Tabla 4.38. Hay cuatro dominios distintos.
2. Asignación de los elementos del Espacio del Problema de Auditoría a los Procesos de TI de COBIT correspondientes (ISACAF-D, 2000), según los factores de riesgo identificados en la Tabla 4.38. Hay 34 procesos distintos.
3. Asignación de los elementos del Espacio del Problema de Auditoría a Objetivos de Control de Alto Nivel de COBIT correspondientes (ISACAF-D, 2000), según los factores de riesgo identificados en la Tabla 4.38. Hay 34 objetivos distintos, uno por cada proceso.

4. Asignación de los elementos del Espacio del Problema de Auditoría a Objetivos de Control Detallados de COBIT correspondientes (ISACAF-D, 2000), según los resultados del paso anterior. Hay 302 objetivos detallados distintos, con más de uno por cada objetivo de alto nivel.
5. Selección de las Guías de Auditoría de COBIT, según los Objetivos de Control de Bajo Nivel asignados, y adaptación de las mismas al contexto y alcance previamente fijados (ISACAF-E, 2000).
6. Elaboración de la guía de auditoría específica para la Bioinformática.
7. Aplicación de la guía de auditoría redactada en el paso anterior.

Dado que el proyecto en que se desarrolla esta auditoría está en curso, hasta la actualidad se han cubierto completamente los pasos 1 a 3. El paso 4 se ha realizado parcialmente porque solamente se han revisado los objetivos de control detallados con respecto al desarrollo y gestión de proyectos de investigación. También se dispone de un borrador que es la Guía de Auditoría para la parte estudiada en el paso 4.

El resultado de la aplicación de los Pasos 1 y 2 está presentado de forma resumida en la Tabla 4.39 (no se muestra la Tabla correspondiente a la asignación de objetivos de control de alto nivel). Se puede apreciar que los elementos que no son exclusivamente de TI, como las teorías, protocolos, bioética, etc., tienen poca intersección con los dominios (1 o 2 como máximo). Esto empieza a dar idea de las carencias del marco metodológico cuando se pretende realizar una auditoría científico-tecnológica.

**Tabla 4.39. Asignación de Dominios de COBIT a elementos del Espacio del Problema de Auditoría**

| Elemento del Espacio del Problema de Auditoría | Código | Dominios de COBIT               |                                 |                           |              | Número de Asignaciones |
|--|--------|---------------------------------|---------------------------------|---------------------------|--------------|------------------------|
|  |        | P. Planificación y Organización | A. Adquisición e Implementación | D. Distribución y Soporte | M. Monitoreo |                        |
| Teoría científica biológica                    | EAB01  | X                               |                                 |                           |              | 1                      |
| Teoría científica informática                  | EAB02  | X                               |                                 |                           |              | 1                      |
| Metodología informática                        | EAB03  |                                 | X                               |                           | X            | 2                      |
| Técnica experimental                           | EAB04  |                                 | X                               | X                         | X            | 3                      |
| Protocolo experimental                         | EAB05  |                                 |                                 |                           |              | 0                      |
| Formalización de datos                         | EAB06  | X                               |                                 |                           | X            | 2                      |
| Interpretación de datos                        | EAB07  |                                 |                                 |                           |              | 0                      |
| Datos sensibles                                | EAB08  |                                 | X                               | X                         | X            | 3                      |
| Uso de los datos                               | EAB09  |                                 | X                               |                           | X            | 2                      |
| Bioética                                       | EAB10  | X                               |                                 |                           |              | 1                      |
| Seguridad                                      | EAB11  | X                               | X                               | X                         | X            | 4                      |
| Eficiencia                                     | EAB12  | X                               | X                               | X                         | X            | 4                      |
| Calidad  | EAB13  | X                               | X                               | X                         | X            | 4                      |
| Sistemas SW                                    | EAB14  |                                 | X                               | X                         | X            | 3                      |
| Sistemas HW                                    | EAB15  |                                 | X                               | X                         |              | 2                      |
| Sistemas instrumentales                        | EAB16  |                                 | X                               |                           | X            | 2                      |
| Políticas de investigación                     | EAB17  | X                               |                                 |                           |              | 1                      |
| Sistemas de investigación                      | EAB18  | X                               | X                               |                           |              | 2                      |
| Políticas de gestión                           | EAB19  | X                               |                                 |                           |              | 1                      |
| Sistemas de gestión                            | EAB20  | X                               | X                               |                           | X            | 3                      |
| Tecnologías biológicas                         | EAB21  |                                 | X                               |                           |              | 1                      |
| Tecnologías de la información                  | EAB22  | X                               | X                               | X                         | X            | 4                      |
| Capacidad profesional                          | EAB23  | X                               |                                 |                           |              | 1                      |
| Formación específica                           | EAB24  | X                               |                                 |                           |              | 1                      |
| Capacidad de coordinación                      | EAB25  | X                               |                                 |                           |              | 1                      |

#### 4.4.2 El modelo de auditoría bioinformática

El modelo de desarrollo de auditoría para la Bioinformática basado en el COBIT se completa con la caracterización de las funciones principales de la Bioinformática (FPB) a partir de los EEPA, estableciendo una correspondencia en el modelo, M:  $EAB_i \rightarrow FPB_j$ . Estas funciones sirven para racionalizar los esfuerzos en la aplicación de la auditoría en los pasos de asignación de dominios, procesos y objetivos de control de COBIT y son las siguientes:

1. **FPB01.** Función de Investigación Teórica: Corresponde a las actividades de investigación básica, tanto biológica como informática, para la búsqueda de teorías y selección de las más adecuadas en función de la naturaleza del problema, y la elección de líneas de investigación que contribuyan positivamente a la comunidad. Comprende los elementos siguientes: EAB01, EAB02 y EAB10.
2. **FPB02.** Función de Gestión de la Investigación: Corresponde a las actividades relacionadas con la organización y gestión de la investigación en todos los aspectos. Comprende los elementos siguientes: EAB12, EAB13, EAB17, EAB19, EAB20 y EAB25.
3. **FPB03.** Función de Aplicación de la Investigación: Corresponde al desarrollo de la investigación en todos los aspectos a partir de las teorías seleccionadas. Comprende los elementos siguientes: EAB03, EAB04, EAB05, EAB06, EAB08, EAB12, EAB13, EAB14, EAB15, EAB16 y EAB18.
4. **FPB04.** Función de Desarrollo Tecnológico: Corresponde a las actividades de búsqueda, adaptación y desarrollo de las tecnologías necesarias para la disciplina. Comprende los elementos siguientes: EAB21 y EAB22.
5. **FPB05.** Función de Formación: Corresponde a las actividades de formación y preparación de los investigadores y técnicos que van a trabajar en la investigación y el desarrollo en esta disciplina. Comprende los elementos siguientes: EAB23 y EAB24.
6. **FPB06.** Función de Seguridad y Protección de la Información: Corresponde a las actividades de seguridad en general y de uso adecuado de la información, tanto la aportada externamente a la investigación, como la que generan los propios experimentos. Comprende los elementos siguientes: EAB07, EAB08, EAB09, EAB11 y EAB13.

Hay algunos EEPA que son comunes en más de una función, como por ejemplo, la Calidad y la Seguridad, ya que las funciones no son absolutamente excluyentes.

De la misma forma que se ha hecho en la subsección anterior, se procede a la adscripción a un dominio y un proceso determinado de cada una de las funciones identificadas. Para cada función se determina el grado de cumplimiento de los objetivos de gestión: P, Primario; y S, Secundario (ISACAF-B, 2000). En la Tabla 4.42 se puede ver el resultado respecto de los dominios y procesos de IT.

A continuación, para cada uno de los procesos que se relacionan con las funciones identificadas, con un grado de cumplimiento P o S, se obtienen los Objetivos de Control de Alto Nivel (ISACAF-D, 2000), uno por cada proceso, y los Objetivos de Control Detallados, que suelen ser más de uno por proceso.

Por ejemplo, para el proceso P01, Definir un plan estratégico de sistema, que consta de seis objetivos detallados, la función de Investigación Teórica (FPB01) debe auditarse con los siguientes objetivos de control detallados para cada uno de los aspectos de este proceso:

1. **P01.1.** Tecnología de Información como parte del Plan de la Organización a corto y largo plazo: La gerencia será la responsable de desarrollar e implementar planes a largo y corto plazo que satisfagan la misión y las metas de la organización.
2. ....
3. **P01.6.** Evaluación de Sistemas Existentes: La Gerencia de servicios informáticos debe evaluar los sistemas existentes en términos de nivel de automatización de negocio, funcionalidad, estabilidad, complejidad, costo y fortalezas y debilidades.

Dadas las características de la función FBP01, se ha apreciado que la misma tiene un grado de cumplimiento secundario con los objetivos de control de este proceso y con otros del mismo dominio. Y ninguna relación con el resto de dominios, según se muestra en la Tabla 4.40. Por lo tanto, las relaciones entre la FPB01 y los dominios, procesos y objetivos de COBIT son bastante indirectas o débiles. Este es un caso singular si se compara con el resto de funciones, donde sí hay relaciones más fuertes con la estructura de COBIT.

Seguidamente, el modelo desglosado por los Objetivos de Control Detallados se sigue extendiendo mediante la aplicación de las pautas y procedimientos de Auditoría que se detallan en el COBIT Audit Guidelines (ISACAF-E, 2000). Por supuesto, este modelo no es sencillo sino bastante complejo y debe estar apoyado por la utilización de distintas herramientas informáticas que proporciona el mismo COBIT (ISACAF-E, 2000), o que proporcionan otros proveedores. De forma alternativa, el modelo se puede aplicar por medio de un conjunto de hojas de



cálculo relacionadas entre sí para poder desglosar los Procesos, Objetivos Generales, Objetivos Concretos y Procedimientos de Auditoría.

#### 4.4.3 Propuesta metodológica para la Auditoría Bioinformática

En la realización de una auditoría sobre la Bioinformática, según se ha descrito en la sección anterior, se ha podido medir la capacidad de expresividad metodológica de COBIT para el contexto y el alcance especificados previamente. Esta medición ha producido buenos resultados en las FPB compuestas de EEPA relacionados directamente con las TI. Sin embargo, carece de potencia metodológica cuando se trata de aplicarlo a funciones compuestas de EEPA relacionados directamente con los enfoques y actividades científicas. Esto confirma la hipótesis que se había formulado antes de iniciar el proyecto de auditoría: el COBIT no es adecuado para la realización de auditorías científicas o científico-tecnológicas.

Por tanto, ha sido necesario formular un marco metodológico alternativo o una extensión de COBIT para el contexto científico tecnológico. La propuesta que se presenta en este trabajo es el COBSIT (*Control Objectives for Bioinformatics: Science, Information and related Technology* - Objetivos de Control para la Bioinformática: Ciencia, Información y Tecnologías Afines). Esta extensión consiste en la creación de dos nuevos dominios con sus respectivos procesos, objetivos de control y guías de auditoría (Tabla 4.41).

Estos nuevos dominios tienen la siguiente justificación:

- **Dominio RP. Investigación y Difusión (Research and Publishing):** Comprende los procesos y objetivos de control relacionados con la identificación estratégica y teórico de la investigación en Bioinformática, con la formulación de planes de investigación y su desarrollo, con la formulación y actividades de difusión de la información científica, y con las garantías éticas y de privacidad de la información y la práctica científica.
- **Dominio ST. Desarrollo y Formación Científicas (Scientific Development and Training):** Comprende los procesos y objetivos de control relacionados con el desarrollo de técnicas y procedimientos científicos y tecnológicos, con la formación científica y técnica del personal bioinformático, y con el apoyo al investigador y la garantía de ética científica del mismo.

Tabla 4.40. Dominios y Procesos de COBIT que intervienen sobre las funciones principales de la Bioinformática

| ESTRUCTURA DE COBIT          |      | BIOINFORMÁTICA                                    |                       |       |       |       |       |       |
|------------------------------|------|---|-----------------------|-------|-------|-------|-------|-------|
| DOMINIO                      | Nº   | PROCESOS  | FUNCIONES PRINCIPALES |       |       |       |       |       |
|                              |      | DENOMINACIÓN                                      | FPB01                 | FPB02 | FPB03 | FPB04 | FPB05 | FPB06 |
| Planeación y Organización    | PO1  | Definir un plan estratégico de sistema            | S                     | P     |       |       |       |       |
|                              | PO2  | Definir la arquitectura de información            |                       | P     | P     |       |       |       |
|                              | PO3  | Determinar la dirección tecnológica               | S                     | P     |       | P     |       |       |
|                              | PO4  | Definir la organización y sus relaciones          | S                     | S     |       |       |       |       |
|                              | PO5  | Administrar las inversiones (en TI)               |                       | P     |       | P     |       |       |
|                              | PO6  | Comunicar la dirección y objetivos de la gerencia |                       | S     |       | S     | S     | S     |
|                              | PO7  | Administrar los recursos humanos                  |                       | P     | P     | P     |       |       |
|                              | PO8  | Asegurar el apego a disposiciones externas        | S                     | P     | P     | S     |       | P     |
|                              | PO9  | Evaluar riesgo                                    |                       | P     | P     | P     | P     | P     |
|                              | PO10 | Administrar proyecto                              |                       |       |       | P     | P     |       |
|                              | PO11 | Administrar calidad                               |                       | P     | P     | P     | P     | P     |
| Adquisición e Implementación | AI1  | Identificar soluciones de automatización          |                       | P     |       | P     |       | P     |
|                              | AI2  | Adquirir y mantener software de aplicación        |                       | P     | S     | S     |       |       |
|                              | AI3  | Adquirir y mantener la arquitectura tecnológica   |                       | P     | S     | S     |       |       |
|                              | AI4  | Desarrollar y mantener procedimiento              |                       | S     | P     | S     |       | P     |
|                              | AI5  | Instalar y acreditar sistemas de información      |                       | P     | S     | S     |       | S     |
|                              | AI6  | Administrar cambios                               |                       | S     | P     | P     | P     | S     |

**ESTRUCTURA DE COBIT** **BIOINFORMÁTICA**

| <b>DOMINIO</b>                        |             | <b>PROCESOS</b>                         |                     |              |              |              | <b>FUNCIONES PRINCIPALES</b> |              |              |   |  |   |
|---------------------------------------|-------------|---|---------------------|--------------|--------------|--------------|------------------------------|--------------|--------------|---|--|---|
|                                       |             | <b>Nº</b>                               | <b>DENOMINACIÓN</b> | <b>FPB01</b> | <b>FPB02</b> | <b>FPB03</b> | <b>FPB04</b>                 | <b>FPB05</b> | <b>FPB06</b> |   |  |   |
| <b>Entrega de Servicios y Soporte</b> | <b>DS1</b>  | Definir niveles de servicio             |                     | P            |              |              | P                            |              |              | P |  |   |
|                                       | <b>DS2</b>  | Administrar servicios de terceros       |                     |              |              |              |                              |              |              |   |  |   |
|                                       | <b>DS3</b>  | Administrar desempeño y capacidad       |                     |              |              |              | S                            |              |              |   |  |   |
|                                       | <b>DS4</b>  | Asegurar continuidad de servicio        |                     | P            |              | P            | P                            |              |              | P |  |   |
|                                       | <b>DS5</b>  | Garantizar la seguridad de sistema      |                     | P            |              | P            | P                            |              |              | P |  |   |
|                                       | <b>DS6</b>  | Identificar y asignar costo             |                     | P            |              | S            | P                            |              |              |   |  |   |
|                                       | <b>DS7</b>  | Educar y capacitar a usuario            |                     | P            |              | P            | P                            |              |              | P |  |   |
|                                       | <b>DS8</b>  | Apoyar y orientar a clientes            |                     |              |              |              |                              |              |              |   |  |   |
|                                       | <b>DS9</b>  | Administrar la configuración            |                     |              |              |              | P                            |              | S            |   |  |   |
|                                       | <b>DS10</b> | Administrar problemas e incidente       |                     |              |              |              | P                            |              | S            |   |  | P |
|                                       | <b>DS11</b> | Administrar la información              |                     |              | P            |              | P                            |              | S            |   |  | P |
|                                       | <b>DS12</b> | Administrar las instalaciones           |                     |              | S            |              | P                            |              | S            |   |  | P |
|                                       | <b>DS13</b> | Administrar la operación                |                     |              |              |              | P                            |              | P            |   |  | P |
| <b>Monitorización</b>                 | <b>M1</b>   | Monitorear el proceso                   |                     |              |              |              |                              | P            |              | P |  | P |
|                                       | <b>M2</b>   | Evaluar lo adecuado del control interno |                     | S            |              | S            |                              | S            |              |   |  | P |
|                                       | <b>M3</b>   | Obtener aseguramiento independiente     |                     |              |              | S            |                              | S            |              | P |  | P |
|                                       | <b>M4</b>   | Proporcionar auditoría independiente    |                     | S            |              | P            |                              | P            |              | P |  | P |

**Tabla 4.41. Propuesta para la extensión de COBIT en nuevos Dominios y Procesos para el contexto científico: el COBSIT**

| <b>DOMINIO</b>  | <b>PROCESOS</b> |   |
|---|-----------------|---|
|   | <b>Nº</b>       | <b>DENOMINACIÓN</b>                               |
| <b>Investigación y Difusión<br/>(Research and Publishing)</b>                       | <b>RP01</b>     | Definir un plan estratégico de investigación      |
|   | <b>RP02</b>     | Determinar la dirección científica                |
|   | <b>RP03</b>     | Identificar teorías formales                      |
|   | <b>RP04</b>     | Identificar soluciones teóricas                   |
|   | <b>RP05</b>     | Asegurar la sumisión a disposiciones bioéticas    |
|   | <b>RP06</b>     | Evaluar riesgo científico                         |
|   | <b>RP07</b>     | Administrar la privacidad de la información       |
|   | <b>RP08</b>     | Administrar los recursos científicos              |
|   | <b>RP09</b>     | Instalar y acreditar sistemas de investigación    |
|   | <b>RP10</b>     | Identificar soluciones bioinformáticas            |
|   | <b>RP11</b>     | Adquirir y mantener instrumental científico       |
|   | <b>RP12</b>     | Adquirir y mantener software científico           |
|   | <b>RP13</b>     | Administrar la difusión de la información         |
|   | <b>RP14</b>     | Instalar y acreditar sistemas de difusión         |
|   | <b>RP15</b>     | Proporcionar evaluación independiente             |
| <b>Desarrollo y Formación Científicas<br/>(Scientific Development and Training)</b> | <b>ST01</b>     | Definir un plan estratégico de desarrollo         |
|   | <b>ST02</b>     | Identificar soluciones técnicas y procedimentales |
|   | <b>ST03</b>     | Desarrollar técnicas y protocolos                 |
|   | <b>ST04</b>     | Definir un plan estratégico de formación          |
|   | <b>ST05</b>     | Formar y capacitar al investigador                |
|   | <b>ST06</b>     | Formar y capacitar al técnico                     |
|   | <b>ST07</b>     | Apoyar y orientar al investigador                 |
|   | <b>ST08</b>     | Garantizar la ética científica                    |

**Tabla 4.42. Los nuevos Dominios y Procesos de COBSIT que intervienen sobre las funciones principales de la Bioinformática**

| ESTRUCTURA DE COBIT      |          | BIOINFORMÁTICA                                 |                       |       |       |       |       |       |
|--------------------------|----------|--|-----------------------|-------|-------|-------|-------|-------|
| DOMINIO                  | PROCESOS |  | FUNCIONES PRINCIPALES |       |       |       |       |       |
|                          | Nº       | DENOMINACIÓN                                   | FPB01                 | FPB02 | FPB03 | FPB04 | FPB05 | FPB06 |
| Investigación y Difusión | RP01     | Definir un plan estratégico de investigación   | P                     |       |       |       |       |       |
|                          | RP02     | Determinar la dirección científica             | P                     |       |       |       |       |       |
|                          | RP03     | Identificar teorías formales                   | P                     |       |       |       |       |       |
|                          | RP04     | Identificar soluciones teóricas                | P                     |       |       |       |       |       |
|                          | RP05     | Asegurar la sumisión a disposiciones bioéticas | P                     | P     | S     |       |       | P     |
|                          | RP06     | Evaluar riesgo científico                      |                       | P     | P     |       |       |       |
|                          | RP07     | Administrar la privacidad de la información    | S                     | P     | P     | S     |       | P     |
|                          | RP08     | Administrar los recursos científicos           | S                     | P     | P     | S     |       | S     |
|                          | RP09     | Instalar y acreditar sistemas de investigación |                       |       | P     |       |       |       |
|                          | RP10     | Identificar soluciones bioinformáticas         |                       |       | P     | P     |       |       |
|                          | RP11     | Adquirir y mantener instrumental científico    |                       |       | P     |       |       |       |
|                          | RP12     | Adquirir y mantener software científico        |                       |       | P     |       |       |       |
|                          | RP13     | Administrar la difusión de la información      | S                     | P     | S     |       |       | S     |
|                          | RP14     | Instalar y acreditar sistemas de difusión      |                       |       | P     |       |       |       |
|                          | RP15     | Proporcionar evaluación independiente          |                       | P     | S     | S     |       | S     |

| ESTRUCTURA DE COBIT                |          | BIOINFORMÁTICA                                    |                       |       |       |       |       |       |
|------------------------------------|----------|---|-----------------------|-------|-------|-------|-------|-------|
| DOMINIO                            | PROCESOS |   | FUNCIONES PRINCIPALES |       |       |       |       |       |
|                                    | Nº       | DENOMINACIÓN                                      | FPB01                 | FPB02 | FPB03 | FPB04 | FPB05 | FPB06 |
| Desarrollo y Formación Científicas | ST01     | Definir un plan estratégico de desarrollo         |                       | P     |       | P     |       |       |
|                                    | ST02     | Identificar soluciones técnicas y procedimentales |                       |       | S     | P     |       |       |
|                                    | ST03     | Desarrollar técnicas y protocolos                 |                       |       |       | P     |       |       |
|                                    | ST04     | Definir un plan estratégico de formación          |                       | P     |       |       | P     | P     |
|                                    | ST05     | Formar y capacitar al investigador                |                       |       | S     | P     | P     | S     |
|                                    | ST06     | Formar y capacitar al técnico                     |                       |       | S     |       | P     | S     |
|                                    | ST07     | Apoyar y orientar al investigador                 |                       | P     | P     | S     | P     | S     |
|                                    | ST08     | Garantizar la ética científica                    |                       | S     | S     |       | P     | S     |

Con esta alternativa definida, se ha procedido entonces a la adscripción de cada una de las FPB a los nuevos dominios y procesos planteados. Para cada función también se ha determinado el grado de cumplimiento de los objetivos de gestión: P, Primario; y S, Secundario (ISACAF-B, 2000). En la Tabla 4.42 se puede ver el resultado específico (donde no se muestran los dominios y procesos de IT del COBIT).

El resto de pasos puede aplicar como se ha descrito en las secciones anteriores para el COBIT. Pero, no se incluyen en esta memoria los resultados subsiguientes.

#### **4.4.4 El proyecto de auditoría: planteamiento**

El *contexto* del problema es la aplicación de una auditoría de sistemas de información a un laboratorio de investigación en Biomedicina que ha incorporado la Función Bioinformática como una función principal en sus actividades (Ermolaeva et al., 1998). El *alcance* del estudio consiste en la auditoría sobre la Función Bioinformática en las áreas de organización, metodologías y técnicas de ingeniería, tecnologías y políticas e instrumentos de investigación y gestión. Por lo tanto, el *objetivo de la auditoría* es la realización de un proyecto de auditoría preliminar sobre la Función Bioinformática como soporte de la investigación biomédica en un laboratorio de investigación en el mismo campo.

En la Figura 4.28 se muestran las fases de un proyecto completo de auditoría (Bernal et al., 1999) (ISACAF-F, 2000). Se ha hecho la distinción entre dos rutas, la ruta 1 y la ruta 2. La ruta 1 corresponde a una auditoría preliminar ya que no aplica dos fases importantes de pruebas y las pruebas realizadas no lo son en profundidad. En cambio, la ruta 2 corresponde a una auditoría ordinaria.

Dentro de la propuesta metodológica enunciada en el Capítulo 3, se pretende aplicar la Guía de Auditoría de la Función Bioinformática. La Guía de Auditoría de la Función Bioinformática tiene asociado un juego de cuestionarios que se diseñaron y elaboraron en un proyecto académico de la titulación de Ingeniería Informática dirigido por el doctorando.

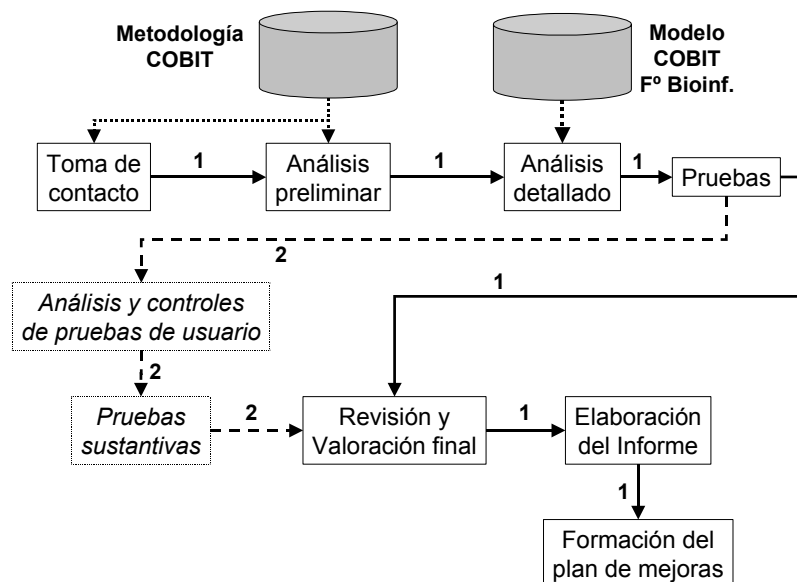


Figura 4.28. Fases de un proyecto de auditoría

#### 4.4.5 Resultados de la auditoría

Se ha realizado el proyecto de auditoría de sistemas de información con las siguientes características:

- Grado preliminar.
- Alcance restringido a la aplicación de sistemas informáticos como soporte a los procesos científicos y a la organización de los recursos.
- El ámbito es el laboratorio tomado como una unidad organizativa casi independiente donde se plantean problemas comunes a las empresas y administraciones, pero también se añaden otros relativos a los procedimientos científicos.

Se ha realizado una guía de auditoría específica para este ámbito adaptando la Guía de Auditoría de la Función Bioinformática. Solamente se describen los resultados más destacables.



#### **4.4.5.1 Resultados del análisis**

Una vez realizado el análisis se han identificado determinados riesgos relativos a la estructura orgánica del laboratorio actual, a la seguridad física y a la operatividad y mantenibilidad de los equipos. A continuación, se describen con más detalle estos resultados.

##### **Riesgos relativos a la estructura orgánica:**

La estructura bicéfala que existe actualmente puede dar lugar a colisión de flujos de órdenes y de información hacia los técnicos de laboratorio, y los técnicos de investigación. El director del laboratorio tiene una fuerte carga de trabajo científico y docente. Además, suele ausentarse con frecuencia para cumplir con diversos compromisos científicos en todo el mundo.

##### **Riesgos relativos a la seguridad física:**

El servicio de seguridad del centro no es completo ya que faltan o no funcionan elementos de vigilancia activa. Esta situación pone en riesgo físico determinados equipos que se encuentran en una salita de acceso libre que sirve además de distribuidor de los despachos. Actualmente, solamente se encuentra allí la impresora láser. Por otra parte, el acceso y el paso no es cómodo ya que todos los elementos mobiliarios han dejado un pasillo demasiado estrecho. Además, el propio funcionamiento de la impresora provoca un aumento de la contaminación acústica y térmica que puede llegar a interferir en el trabajo de quienes se encuentren en dicha salita.

##### **Riesgos relativos a operatividad de los equipos:**

Se ha observado que uno de los equipos que está en el despacho del director, el PC Gateway P-II Windows 98, prácticamente está fuera de uso, pero se mantiene porque en su almacenamiento secundario se guarda información importante para el director. En el otro equipo, se ha observado que la cuenta de usuario habitual es la de Administrador, lo que comporta altos riesgos en cuanto a la seguridad lógica y a la configuración del equipo. La coexistencia física de los dos equipos dificulta la comodidad de trabajo. Además, puesto que el director maneja también un ordenador portátil, existe el riesgo de que no se tenga una idea clara de dónde está la información importante o de que esté duplicada.

Los equipos de los despachos de staff y estudiantes utilizan cuentas de power user, a excepción del equipo del genetista, que utiliza la cuenta de Administrador. En este caso, existen los mismos riesgos que con el equipo del director.

Los equipos de los laboratorios, a excepción del que está asignado a uno de los técnicos de laboratorio, son todos equipos supeditados al instrumental descrito en la Figura 1. Dichos equipos no se utilizan para tareas personales, con lo cual los riesgos son menores, pero están conectados a la red local para que se pueda acceder desde el resto de equipos.

#### **4.4.5.2 Resultados de las pruebas**

Una vez realizadas algunas pruebas e inspecciones de los equipos, se han identificado determinados riesgos relativos a la seguridad lógica local y de red y a la fiabilidad de la información. A continuación, se describen con más detalle estos resultados.

##### **Riesgos relativos a la seguridad lógica local:**

Una vez inspeccionados cada uno de los equipos se han podido detectar algunas anomalías de almacenamiento que comportan graves riesgos. Todos los equipos que tienen instalado el sistema operativo MS Windows 2000 disponen de un disco duro con una sola partición. Además, la mayoría de usuarios utilizan la carpeta “My documents” para guardar sus datos. Esto tiene el riesgo de la pérdida irreparable de la información si se procede a la reinstalación del sistema operativo o al formateo de la partición, y no se ha hecho copia de seguridad o ésta esté desactualizada. Por otra parte, no existe ninguna política de copias de seguridad y, en consecuencia, no se aplica ningún procedimiento para hacer copias periódicas del contenido de los equipos.

##### **Riesgos relativos a la seguridad lógica de red:**

La utilización de cuentas de administrador en algunos equipos es un riesgo alto porque los privilegios de acceso que tienen permiten acceder a la red en general desde el exterior. Se ha observado que el Departamento de Computación ha instalado un programa de monitorización en cada uno de los equipos. Sin embargo, no se ha podido comprobar el grado de fiabilidad y no se ha recibido información sobre todas las funciones que realiza. Por otra parte, las cuentas de usuario de tipo power user dadas de alta en cada uno de los equipos tienen todas el mismo nombre, aunque distinta contraseña. Esto facilita el ataque por diccionario porque solamente se ha de obtener la contraseña.

Se ha comprobado también que los equipos no están asignados a un solo dominio de red, sino que algunos siguen con el dominio por defecto "WORKGROUP" y otros presentan diferencias de nombre. Esto hace que no sea fácil detectar cada uno de los equipos cuando se accede por medio de la red de Microsoft. Además, ninguno de los usuarios ha sido formado para utilizar con provecho las funciones de red a su alcance y así, por ejemplo, poder intercambiar información masiva. En tercer lugar, la ejecución de determinados programas en los equipos asignados al instrumental de laboratorio, está provocando riesgos de seguridad cuando se accede a dichos equipos desde los puestos de trabajo de los investigadores. Esto es porque dichos programas no incluyen mecanismos de seguridad o son demasiado antiguos.

#### **Riesgos relativos a la fiabilidad de la información:**

Una detallada inspección de la estructura de ficheros de cada uno de los equipos ha detectado que ninguno de los usuarios está aplicando un procedimiento estandarizado y claro cuando guarda la información en su partición de disco. Incluso se han detectado carpetas de información dentro de los directorios correspondientes a la instalación de aplicaciones y al propio sistema operativo. Además, se ha observado que la mayor parte de la información se guarda en la carpeta "My documents" y no existe ninguna carpeta espejo en la cuenta de usuario, ni fuera de ella. No se puede saber así con claridad la información que está duplicada, obsoleta o se debe guardar off-line para despejar el disco duro.

El proyecto de auditoría se ha concluido con la elaboración de un informe final que contiene las recomendaciones de modificación y mejora de las condiciones, que están resumidas en la sección siguiente. No se ha elaborado el plan de mejoras puesto que es necesario definir previamente las responsabilidades de su ejecución entre el Laboratorio y el Departamento de Computación del HNRC.

## **4.5 Bibliografía**

### **4.5.1 Artículos publicados**

Coltell O, Arregui M., Chalmeta R., Corella D., Parnell L., Ordovás JM. "SeqPacker: a nucleotide sequence manipulation utility". Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE2004) May 19-21, 2004, Taichung, Taiwan, ROC (<http://bibe2004.ece.uci.edu/>).

- Coltell O, Arregui M., Chalmeta R., Corella D., Parnell L., Ordovás JM. "A Biologist-Friendly User Interface to Manipulate Nucleotide Sequences in Genomic Epidemiology". 2nd International Workshop on Biological Data Management (BIDM'04). 1 - 3 September, 2004. Zaragoza, Spain.
- Coltell O. "Auditoría Bioinformática". I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 15-24. ISSN: 1579-8070. Editorial MIC.
- Coltell O. "El marco de auditoría básico para las funciones bioinformáticas". I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 25-28. ISSN: 1579-8070. Editorial MIC.
- Coltell O. "La representación, almacenamiento e interrelación de entidades genómicas y proteómicas en las bases de datos públicas: un caso de estudio". Curso de Informática Biomédica. Instituto de Salud Carlos III (ISCIII), Madrid, 13 de octubre de 2003.
- Coltell O. "Sistemas de información en laboratorios de investigación en Biomedicina". I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 9-14. ISSN: 1579-8070. Editorial MIC.
- Coltell O., Corella D., Sánchez JT., Chalmeta R., Ordovas JM. "Modelo Conceptual Bioinformático Aplicado al Análisis Genómico en las Enfermedades Cardiovasculares". Clínica e Investigación en Arteriosclerosis; 2004, 16(2): 43-52.
- Coltell O., Corella D., Tai E-S., Guillén M., Chalmeta R., Ordovas JM. "PLATEX: Una herramienta bioinformática para la conversión de datos en el estudio genético de la arteriosclerosis". Clínica e Investigación en Arteriosclerosis; 2004, 16(2): 53-60.
- Corella D, Tucker K, Lahoz C, Coltell O, Cupples LA, Wilson PWF, Schaefer EJ, Ordovas JM. "Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study". American Journal of Clinical Nutrition; 2001; 73(4):736-45. ISSN "0002-9165 (SCI: 2000, 5,012 p., 3/52 "NUTRITION AND DIETETICS").
- Doreen Osgood; Dolores Corella; Serkalem Demissie; L Adrienne Cupples; Peter WF Wilson; James B Meigs; Ernst J Schaefer; Oscar Coltell; Jose M Ordovas. "Genetic variation at the scavenger receptor class B type I (SR-BI) gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: The Framingham Study". J Clin Endocrinol Metab. 2003 Jun;88(6):2869-79. ISSN 0021-972X (SCI 2002, 5.199 p.; 13/88 "ENDOCRINOLOGY & METABOLISM").
- Lu Qi; Dolores Corella, José V. Sorlí, Olga Portolés, Haiqing Shen, Oscar Coltell, Diego Godoy, Andrew S. Greenberg, Jose M. Ordovas. "Genetic variation at the Perilipin (PLIN) locus is associated with obesity-related phenotypes in White women". Clinical Genetics 2004 Ref. CGE-00136-2004. ISSN 0009-9163 Online ISSN: 1339-0004 (SCI 2002, 2.237 p.; 57/115 "GENETICS & HEREDITY").

- Ordovás JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Polyunsaturated fatty acids modulate the effect of the APOA1-75(G/A) polymorphism on HDL-C levels in a specific fashion: The Framingham Study". *American Journal of Clinical Nutrition*, 2002; 75(1):38-46. ISSN "0002-9165 (SCI: 2000, 5,012 p., 3/52 "NUTRITION AND DIETETICS")".
- Ordovás JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on HDL metabolism: Evidence of a strong dose-effect in this gene-nutrient interaction in the Framingham Study". *Circulation*, 2002; 18: 2315-2321. ISSN 0009-7322 (SCI: (2000, 10,893 p.), (2001, 10,517 p.); 1/63 "CARDIAC & CARDIOVASCULAR SYSTEMS"; 1/60 "HEMATOLOGY"; 1/45 "PERIPHERAL VASCULAR DISEASE")".
- Ordovás JM, Cupples A, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PWF, Schaefer EJ. "Association of CETP TaqIB polymorphism with variation in lipoprotein subclasses and coronary heart disease risk: The Framingham Study". *Arteriosclerosis Thrombosis and Vascular Biology*; 2000; 20: 1323-1329. ISSN 1079-5642 (SCI: 1999, 5,406 p., 6/45 "PERIPHERAL VASCULAR DISEASE", 5/60 "HEMATOLOGY")".

## 4.5.2 Bibliografía de resultados de proyectos bioinformáticos

- Applied Biosystems home page. <http://www.appliedbiosystems.com/>. Accedido el 4 de junio de 2004.
- Applied Biosystems. ABI PRISM® 3100 Genetic Analyzer. User's Manual 2001. Applied Biosystems at <http://docs.appliedbiosystems.com/genindex.taf> (access data 10/31/2002).
- Chromas home page. <http://www.technelysium.com.au/chromas.html>. Accedido el 4 de junio de 2004.
- FASTA format Description web page. <http://ngfnblast.gbf.de/docs/fasta.html>. Accedido el 4 de junio de 2004.
- GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.
- OMG home page. <http://www.omg.org>. Accedido el 4 de junio de 2004.

### 4.5.2.1 Bibliografía de soluciones bioinformáticas: seqpacker

- Booch G., Rumbaugh J., Jacobson I. *The Unified Modeling Language User Guide*. Addison-Wesley, Reading, MA, 1999.

Fischer, C., Schweigert, S., Spreckelsen, C., Vogel, F. "Programs, databases, and expert systems for hu-man geneticists-a survey". Hum. Genet. (1997), 97:129-137.

GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.

<http://iubio.bio.indiana.edu/soft/molbio/mac/plasmid-maker.html>. Accedido el 4 de junio de 2004.

<http://iubio.bio.indiana.edu/soft/molbio/mac/plasmid-maker.html>. Accedido el 4 de junio de 2004.

<http://www.accessexcellence.org/AE/AEC/CC/restriction.html>. Accedido el 4 de junio de 2004.

<http://www.arches.uga.edu/~lace52/rplasmid.html>. Accedido el 4 de junio de 2004.

Rumbaugh J., Jacobson I., Booch G. El Lenguaje Unificado de Modelado. Manual de Referencia. Addison-Wesley, Madrid, 2000.

### 4.5.3 Bibliografía de LIPC

Couture P, Otvos JD, Cupples LA, et al. Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles. The Framingham Offspring Study. Arterioscler Thromb Vasc Biol 2000;20:815-822.

### 4.5.4 Bibliografía de PLIN

Lu X, Gruia-Gray J, Copeland NG, Gilbert DJ, Jenkins NA, Londos C, Kimmel AR 2001 The murine perilipin gene: The lipid droplet-associated perilipins derive from tissue-specific, mRNA splice variants and define a gene family of ancient origin. Mamm Genome 12:741-749.

### 4.5.5 Bibliografía auditoría de sistemas de información en laboratorios de investigación en Biomedicina

Benal R., Coltell O. Auditoría de los Sistemas de Información (reimpresión). Servicio de Publicaciones de la Universidad Politécnica de Valencia, Valencia, 1999.

Cornell M., Paton N.W., Wu S., Goble C.A., Miller C.J., Kirby P., Eilbeck K., Brass A., Hayes A., Oliver S.G. "GIMS-A Data Warehouse for Storage and Analysis of Genome Sequence and Functional Data". EBI, the European Bioinformatics Institute (EMBL Outstation, Hinxton, UK), <http://www.ebi.ac.uk/> (accedido 6/0272002).

Ermolaeva O., Rastogi M., Pret K.D., Schuler G.D., Bittner M.L., Chen Y., Simon R., Meltzer P., Trent J.M., Boguski M.S. "Data management and analysis for gene expression arrays". *Nature genetics*, 20; 1998: 19-23.

ISACAF-B. COBIT. Framework. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.

ISACAF-D. COBIT. Control Objectives. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.

ISACAF-E. COBIT. Audit Guidelines. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.

ISACAF-F. COBIT. Implementation Tool Set. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.

Sackman H. *Biomedical Information Technology. Global Social Responsibilities for the Democratic Age*. Academic Press, San Diego, CA (USA), 1997.

Dime y lo olvido, enséñame y lo recuerdo, involúcrame y lo aprendo.

Benjamín Franklin

## 5.1 Generalidades

En este capítulo se realiza, en primer lugar, la discusión de los resultados obtenidos para cada uno de los genes analizados y, en segundo lugar, una discusión general de los resultados bioinformáticos y de su integración en el ámbito de la investigación genómica cardiovascular aplicada.

## 5.2 Discusiones de cada uno de los genes

En esta sección, se presenta la discusión individualizada de cada uno de los genes estudiados de acuerdo con los artículos previamente publicados: CEPT, APOE, APOA1, LIPC, SR-BI y PLIN.

### 5.2.1 Discusión del gen CETP

En el estudio sobre el gen CETP en la población del *Framingham Offspring Study* se ha encontrado una asociación del polimorfismo TaqIB del CETP, con los niveles de HDL-C y la apoA-I en hombres, y con HDL-C en mujeres. No se han observado diferencias estadísticamente significativas en los niveles de otros lípidos según los genotipos de TaqIB después de controlar por relaciones familiares, edad, IMC, tabaco, ingesta de alcohol, uso de betabloqueantes, estado menopáusico y terapia de estrógenos en mujeres, y genotipo de APOE. Además, se ha encontrado que el alelo B2 estaba asociado significativamente con aproximadamente un riesgo de ECV 30% menor, que no se ha mantenido estadísticamente significativo después de haber ajustado por los niveles de HDL-C. Esto indica que una proporción



significativa de este efecto es debida a los niveles aumentados de HDL-C asociados con la presencia del alelo B2.

Esta asociación con HDL-C se ha informado anteriormente en varios estudios (Kondo et al., 1989) (Kuivenhoven et al., 1997) (Miechell et al., 1994) (Fumeron et al., 1995) (Kauma et al., 1996) (Kuivenhoven et al., 1998) (Bernard et al., 1998) (Freeman et al., 1990). Algunos de estos estudios también han encontrado asociaciones significativas con los niveles de LDL-C o triglicéridos (Kondo et al., 1989) (Kuivenhoven et al., 1998) (Gudnason et al., 1999). De acuerdo con algunos de estos estudios (Kuivenhoven et al., 1997) (Kauma et al., 1996) (Kuivenhoven et al., 1998) (Logan et al., 1997), en el que se presenta se ha encontrado una asociación significativa entre el polimorfismo TaqIB del CETP y la actividad de CETP, estando asociado el alelo B1 con el incremento de actividad en comparación con el alelo B2. Sin embargo, otros investigadores (Freeman et al., 1994) también han informado de la ausencia de asociación significativa entre la actividad de CETP y el polimorfismo TaqIB del CETP. No se conoce el mecanismo por el cual el polimorfismo TaqIB puede afectar la actividad de CETP o los niveles de HDL-C. Es poco probable que este polimorfismo localizado en un intrón represente una mutación funcional. Dadas las asociaciones publicadas del alelo B2 con el aumento de masa y/o actividad de CETP, la explicación más plausible es que este polimorfismo está en desequilibrio de ligamiento con una mutación funcional todavía desconocida en la región reguladora del gen CETP.

El papel de CETP en la aterogénesis está todavía a debate (Logan et al., 1997) (Stevenson et al., 1998). El CETP puede jugar un papel proaterogénico puesto que media una redistribución del colesterol plasmático de las lipoproteínas con una protección contra la aterosclerosis en la apoB proaterogénica conteniendo lipoproteínas. Este concepto también está apoyado por el hecho de que las especies animales que son resistentes a la aterosclerosis inducida por la dieta tienen menos actividad de CETP. Sin embargo, la CETP media solamente en uno de los pasos del transporte inverso del colesterol, que es un proceso antiaterogénico. Los resultados de este estudio refuerzan el supuesto de que el aumento en los niveles de HDL-C resultantes de la baja actividad de CETP parece estar asociada con un bajo riesgo de ECV en los hombres.

Los experimentos en humanos y en ratones transgénicos han mostrado que los factores ambientales juegan un papel importante en la modulación de la expresión del gen CETP (44). Hay varios estudios en poblaciones humanas que han analizado la posible interacción entre algunos factores ambientales y el polimorfismo TaqIB de CETP sobre los niveles de HDL-C. Kondo et al. (1989) proporcionan la evidencia de que prácticamente el efecto completo del gen CETP sobre HDL-C es debido al efecto observable en no fumadores. En otro estudio realizado en

Finlandia (Kauma et al., 1996), los hombres fumadores con el alelo B2 tendían a mantener niveles de HDL-C menores en un 10% que los hombres fumadores con el alelo B1. Pero este efecto no se había observado en las mujeres, y los autores también habían concluido que los efectos alélicos diferían según el sexo. Fumeron et al. (1995) no han encontrado una interacción con el consumo de tabaco, pero encontraron en cambio una interacción importante con el consumo de alcohol en el estudio ECTIM.

En el estudio que se presenta, al realizar la comprobación de los términos gen-ambiente no se han observado interacciones significativas de los genotipos de TaqIB con el consumo de alcohol y tabaco. Esta observación permite concluir que el efecto de este polimorfismo sobre el HDL-C plasmático en la cohorte del *Framingham Offspring Study* parece ser estadísticamente independiente y uniforme a lo largo de varios niveles de estos factores ambientales. El estudio de estas interacciones gen-ambiente puede proporcionar una base importante para refinar el valor predictivo de los factores de riesgo epidemiológicos tradicionales y para enfocar las actividades de intervención y prevención para individuos con alto riesgo.

Aparte de las asociaciones publicadas entre los factores de riesgo clásicos y el polimorfismo TaqIB, los resultados del estudio muestran que esta variante genética está asociada significativamente con las diferencias en la distribución de las subclases de lipoproteínas. El alelo B2 estaba asociado significativamente con mayores niveles de las subfracciones grandes de HDL más antiaterogénicas, en hombres y mujeres. Este efecto es consistente con el hecho de que el alelo B2 está asociado con menor actividad de CETP, y provoca un incremento del éster de colesterol enriquecido con HDL de partículas grandes tales como las que se encuentran en sujetos deficientes en CETP. Además, se han encontrado efectos significativos para la distribución de subfracciones de LDL solamente en hombres. El alelo B2 se ha asociado con un menor tamaño de partículas del LDL menos aterogénico, consistente en niveles decrecientes de la subfracción menor del LDL más aterogénico y niveles crecientes de la subfracción mayor del LDL menos aterogénico. Estos efectos no eran significativos en las mujeres. Por lo tanto, el efecto protector asociado al alelo B2 en hombres puede ser debido, no sólo a los cambios cuantitativos en los perfiles lipídicos, sino también a los cambios cualitativos en la composición de partículas, que se manifiesta como diferencias en la distribución de tamaños.

En resumen, este estudio, realizado en la cohorte de población caucásica de gran tamaño que maneja el *Framingham Offspring Study*, confirma la hipótesis de que el locus génico de CETP, examinado con el polimorfismo TaqIB, juega un papel importante en la determinación de la variabilidad de HDL-C en hombres y

mujeres, y de los niveles de apoA-I y tamaño de LDL en hombres, y aporta aproximadamente un 1% de la variabilidad de HDL-C. Estas asociaciones conducen a un perfil lipídico menos aterogénico en ambos sexos y a un menor riesgo de ECV en hombres. Son necesarios estudios posteriores para demostrar si el menor riesgo de ECV asociado con este polimorfismo se encuentra también en las mujeres.

### 5.2.2 Discusión del gen APOE

El estudio sobre el gen APOE en la población del *Framingham Offspring Study* destaca una asociación significativa entre la ingesta de alcohol y APOE en hombres, y contribuye al supuesto de que los efectos de la variabilidad genética de APOE sobre las concentraciones de lípidos plasmáticos son dependientes de los factores ambientales. Los efectos del polimorfismo del APOE sobre las concentraciones de LDL-C (Eichner et al., 1990) (Boerwinkle et al., 1988) (Wilson et al., 1994) no se habían observado en hombres no bebedores. En este grupo, el efecto de disminución esperado sobre el alelo  $\epsilon 2$  estaba ausente, y el efecto de incremento del alelo  $\epsilon 4$  era inverso.

Estudios previos han informado de interacciones significativas entre el genotipo de APOE y la respuesta de lípidos plasmáticos en las terapias dietética y farmacológica (Ordovas, 1999). Sin embargo, los estudios de dietas se han centrado sobre la grasa y el colesterol, y el trabajo previo no ha investigado las interacciones del consumo de alcohol y el genotipo de APOE sobre las concentraciones de lípidos. Aunque los resultados son todavía controvertidos (Lefevre et al., 1997), se ha lanzado la hipótesis que la contribución del alelo  $\epsilon 4$  a las elevadas concentraciones de LDL-C es mayor en presencia de dieta aterogénica (Ordovas, 1999). Además, varios autores (Aguilar et al., 1999) (Kamboh et al., 1991) (Deiana et al., 1998) (Sandholzer et al., 1995) sugieren que una dieta baja en grasas puede suprimir los efectos deletéreos del alelo  $\epsilon 4$  sobre los lípidos plasmáticos.

En este sentido, un estudio llevado a cabo en una población rural nativa americana que seguía su estilo de vida tradicional informaba de que no había diferencias en las concentraciones de LDL-C entre los sujetos  $\epsilon 3/\epsilon 3$  y  $\epsilon 3/\epsilon 4$  (Aguilar et al., 1999). Kamboh et al. (1991) han publicado similares resultados en los indios mayas de Méjico. Considerando que las poblaciones nativas americanas que viven su estilo de vida tradicional no consumen grandes cantidades de alcohol, el estudio que aquí se presenta informa de que la interacción con el consumo de alcohol, además de la hipótesis de la grasa en dieta, podría explicar potencialmente los resultados mostrados en los estudios anteriormente mencionados. Esta hipótesis

de la interacción gen-ambiente se podría aplicar también a los estudios publicados por Deiana et al. (1998) y Sandholzer et al. (1995). Sin embargo, estos estudios no presentaban datos sobre dieta o consumo de alcohol o análisis separados por sexo para verificar esta hipótesis.

Los resultados que se presentan en esta tesis refuerzan la importancia de considerar estas variables cuando se examina este tipo de asociaciones genotipo-fenotipo. En el caso presente, la identificación de APOE y alcohol viene permitido por la inclusión de un término de interacción de sexo en el modelo multivariado, la estratificación del consumo de alcohol y genotipo, y el control adicional para variables de confusión tales como grasa en dieta. En los estudios en que los análisis realizados no han tenido en cuenta estas consideraciones, los efectos de la interacción por alcohol pueden haber sido enmascarados por otros efectos y, en consecuencia, ignorados.

Otro elemento de interés en el estudio de las interacciones gen-ambiente es la influencia de la edad (Zerba et al., 1996). El efecto alélico de APOE puede diferir en los sujetos mayores (Jarvik et al., 1994) (Pablos-Mendez et al., 1997) (Zerba et al., 1996) (Davignon et al., 1988). Esto se ha comprobado en gemelos masculinos seguidos longitudinalmente durante 15 años (Jarvik et al., 1994). En el ciclo 1 del *Framingham Offspring Study* (media de edad de 48 años), los niveles de TC y LDL-C eran más bajos en el grupo  $\epsilon 3$  que en el grupo  $\epsilon 4$ . Pero en el ciclo 3 (media de edad de 63 años), los niveles de TC y LDL-C eran más altos en el grupo  $\epsilon 3$  que en el grupo  $\epsilon 4$ . El envejecimiento podría explicar también la ausencia de asociación entre  $\epsilon 4$  y las concentraciones de LDL-C publicada en una población multiétnica de edad avanzada (Pablos-Mendez et al., 1997). Las razones para estas diferencias relativas a la edad no están claras. Aparte del proceso de envejecimiento, se ha mostrado que la dieta grasa y los consumos de alcohol son menores en los mayores (Whichelow et al., 1996) (Posner et al., 1993) (Shaper et al., 1998) (Greenfield et al., 1999). Según estos datos, los efectos alélicos del APOE observados en la población general no deberían estar presentes en los grupos de población que consumen poco o nada de alcohol. Estos efectos se pueden aplicar a los niños. Aunque algunos estudios (Okada et al., 1998) (Lehtimäki et al., 1994) mostraban un efecto de APOE sobre LDL-C en niños, otros no conseguían hacerlo (Dixon et al., 1997) (Bercedo et al., 1998). Asumiendo que los niños no tomaban bebidas alcohólicas, estos resultados son consistentes con los observados en el estudio que se presenta para hombres no bebedores.

Aunque esta es la primera descripción de que una interacción entre APOE y alcohol afecta a las concentraciones de lípidos, Kauma et al. (1998) han publicado una interacción específica en hombres entre los alelos de APOE y el consumo de alcohol en la determinación de la presión sanguínea. Las diferencias por sexo en

los efectos del alcohol en los lípidos plasmáticos se han estudiado escasamente (Lin et al., 1995). Esto podría ser debido a la complejidad añadida introducida por los cambios en lípidos hormonalmente inducidos durante el ciclo menstrual (Jones et al., 1988), el estado menopáusico (Higuchi et al., 1990) y la terapia hormonal en mujeres (Garry et al., 1999). Según estos resultados, los estrógenos pueden ser cruciales en la determinación de los efectos del alcohol y la posible interacción con el genotipo de APOE. Una interacción con los estrógenos puede explicar las interacciones gen-sexo informadas, en que están implicados varios loci y factores ambientales (Ordovas, 1999) (Hegele et al., 1994). Sin embargo, la falta de información sobre niveles hormonales en esta población establece una prevención inicial a la hora de comprobar estas hipótesis.

Al interpretar los descubrimientos efectuados en este estudio se deben tener en cuenta varios aspectos. En primer lugar, se trata de un análisis transversal y por tanto no se ha determinado la secuencia temporal de los resultados observados. En segundo lugar, la ingesta de alcohol corresponde a lo que los propios participantes declaran, por lo que puede estar sometida a sesgo. El error de respuesta no diferencial en el *Framingham Offspring Study* se ha estimado que es muy bajo (Posner et al., 1992). Además, la ingesta de alcohol se ha utilizado principalmente como una variable categórica de dos valores (consumo y no consumo) que evita sesgos en la estimación de la cantidad consumida. Los factores que afectan esta clasificación dicotomizada pueden ser más importantes en el presente estudio. Sin embargo, la denegación de toma de alcohol por parte de los bebedores puede ser también baja considerando el diseño del *Framingham Offspring Study* y el tipo de participación (Feinleib et al., 1975). La posibilidad de que los no bebedores declaren algún consumo de alcohol es remota (Romelsjo et al., 1995). Finalmente, la limitación apuntada por Shaper (Shaper, 1990), donde argumenta que el grupo de no bebedores está típicamente contaminado por sujetos que ha reducido su consumo de alcohol debido a enfermedad preexistente, está minimizado en este estudio mediante la exclusión de sujetos con ECV o que toman medicación lípidepresora.

Otro asunto se refiere a la posible confusión de la grasa en el efecto del alcohol. Se ha informado que en las poblaciones cuyas ingestas de grasa y colesterol varían con el consumo de alcohol (Colditz et al., 1991), la dieta puede distorsionar la relación entre alcohol y ECV en general, o entre alcohol y LDL-C en particular. Debido a que los bebedores de alcohol pueden tener diferentes consumos de grasa (Mannisto et al., 1997), el control por dieta hace más específica la asociación entre alcohol y ECV. En este estudio, el efecto inicial descrito para el alcohol permanece en los modelos multivariados después de haber controlado por grasa y aporte de energía. Además, el IMC y el consumo de tabaco, que se han asociado distintamente con el consumo de alcohol (Wannamethee et al., 1992)

(Bottoni et al., 1997), también se han tenido en cuenta en los análisis. Existen todavía otros factores de confusión potencial, tales como la actividad física, que puede variar entre sujetos con diferentes niveles de consumo de alcohol. Por lo tanto, no se puede excluir la posibilidad de que este descubrimiento pueda representar una asociación espuria. Todos estos factores requieren exámenes más profundos en futuros estudios.

El mecanismo biomédico que subyace a la interacción informada es desconocido. Se pueden formular varias hipótesis teniendo en cuenta las numerosas rutas metabólicas potencialmente influidas por el etanol (Frohlich et al., 1996), incluyendo los efectos en la síntesis de colesterol hepático endógeno o intestinal, tasas incrementadas de absorción de colesterol intestinal, activación de la lipoproteína lipasa (LPL), inhibición de la proteína de transferencia del éster del colesterol (CETP), la modificación de LDL y consecuentemente un cambio en la afinidad de enlace del receptor de LDL (Liinamaa et al., 1997) (Simon et al., 1996) (Hirano et al., 1992) (Nishiwaki et al., 1994). Es conocido que apoE2 tiene una baja afinidad para unirse a los receptores de LDL (Davignon et al., 1988). Además, las partículas de LDL en bebedores de alcohol son más pequeñas y probablemente están acetiladas, lo que también reduce su afinidad hacia los receptores de LDL (Hirano et al., 1992) (Ayaori et al., 1997). En la ausencia de alcohol, los hombres con el alelo  $\epsilon 2$  tienen concentraciones de LDL-C cercanas a los individuos  $\epsilon 3$ . Pero, en la presencia de alcohol, la baja afinidad de apoE2 hacia los receptores se puede magnificar resultando en un incremento de los receptores de LDL y una disminución en las concentraciones de LDL-C. En los sujetos  $\epsilon 4$ , las lipoproteínas de muy baja densidad están enriquecidas con apoE y cuyas partículas las puede tomar el hígado con mayor facilidad, disminuyendo por tanto la expresión de receptores de LDL e incrementando las concentraciones de LDL-C plasmático. El alcohol induce un incremento en la lipemia postprandial y en los remanentes de TG, dando como resultado una gran aceptación de colesterol por los receptores hepáticos, que regulará a la baja incluso los receptores de LDL y mejorará el efecto de aumento de colesterol del apoE4 (Hendriks et al., 1998) (Van Tol et al., 1998).

Aunque la causa y el efecto no se pueden demostrar en este estudio, los descubrimientos realizados proveen la evidencia que confirma la hipótesis de que los efectos del APOE en lípidos plasmáticos y el riesgo de ECV se modula por factores ambientales. La interacción de alcohol con los alelos de APOE sobre las concentraciones de LDL-C observada en los hombres era consistente e independiente de la cantidad de grasa saturada consumida, aporte de energía, tabaco o IMC. Sin embargo, en las mujeres no se ha podido detectar un efecto similar sugiriendo una interacción sexo-sexo. Los datos manejados también sugieren que parte de la variabilidad informada en términos de los efectos de la

bebida de alcohol y las concentraciones de LDL-C puede ser debida a la distribución por sexo y a la estructura genética de la muestra estudiada.

### 5.2.3 Discusión del gen APOA1

El estudio sobre el gen APOA1 en la población del *Framingham Offspring Study* destaca una asociación significativa entre la ingesta dietética de PUFA y el polimorfismo  $-75$  A/G en el promotor del gen apoA-I sobre los niveles de HDL-C plasmático y apoA-I. Esta variación genética en la región promotora del gen APOA1 se asocia con las diferencias en los niveles de apoA-I y HDL-C en individuos sanos, pero este efecto está fuertemente modulado por el consumo de PUFA. La complejidad de esta interacción gen-ambiente difiere entre hombres y mujeres. En general, cuando se ha considerado en los modelos de regresión el genotipo de APOA1 por la interacción de PUFA, el alelo A se ha asociado con menores niveles de HDL-C y apoA-I, pero el efecto de aumento de la ingesta de PUFA en sujetos portadores del alelo A puede invertir el efecto genético. En las mujeres, esta interacción era altamente significativa para los niveles de HDL-C, y casi estadísticamente significativa para los niveles de apoA-I, e independiente de las interacciones de orden superior con tabaco o bebida. En los hombres, el efecto de PUFA era sólo estadísticamente significativo cuando las interacciones del alcohol y el consumo de tabaco también se incorporaban a los modelos de regresión.

Los descubrimientos de este estudio son particularmente relevantes para los investigadores que trabajan con el efecto de los polimorfismos en la región promotora del gen APOA1 sobre los niveles de lípidos plasmáticos. Debido a que la ingesta de PUFA puede modular significativamente el efecto de las variantes genéticas, esto puede explicar los resultados contradictorios ampliamente informados (Jeenah et al., 1990) (Paul-Hayase et al., 1992) (Meng et al., 1997) (Civeira et al., 1993) (Barre et al., 1994) (Akita et al., 1995). Además, la interacción adicional con el consumo de tabaco y alcohol en los hombres podría contribuir a las diferencias por sexo destacadas por varios autores (Kamboh et al., 15) (Talmud et al., 16).

El estudio más temprano en 1990 conducido por Jeenah et al. (1990) en Inglaterra documentaba que los hombres con el alelo A tenían significativamente mayores niveles de HDL y apoA-I que los portadores de G/G. En el mismo año, Pagani et al. (1990) en Italia, publicaron que en las mujeres la frecuencia del alelo A se incrementaba desde el menor a mayor decil de HDL-C, pero no se detectó ninguna evidencia en los hombres. Efectos similares del alelo A sobre HDL-C o apoA-I se han observado en hombres de un estudio en Bélgica (Paul-Hayase et al.,

1992) y otro en Finlandia (Meng et al., 1997). En este último estudio no se han detectado efectos en las mujeres. En contraste, no se informó de ninguna asociación del alelo A con HDL-C o apoA-I en cada uno de los sexos por parte de Civeira et al. (1993), Lopez-Miranda et al. (1994), Mata et al. (1998) y Carmena-Ramon et al. (1998) en España, así como Akita et al. (1995) en Japón, y Barre et al. (1994) en 22 familias nucleares caucásicas. Además, Matsunaga et al. (1995) observaron que los controles con el genotipo G/A tenían significativamente menores niveles plasmáticos de apoA-I. Estos hallazgos inconsistentes podrían indicar que los efectos del sitio polimórfico -75 bp podrían ser dependientes de los factores ambientales que diferían entre estudios poblacionales.

Solamente unos pocos estudios han abordado esta cuestión y se han centrado en el consumo de tabaco. Uno de estos estudios lo ha llevado a cabo Sigurdsson et al. (1992) en 149 hombres y 166 mujeres de Islandia, donde muestra una asociación significativa entre el genotipo de APOA1 y HDL-C y apoA-I solamente en hombres no fumadores, con los individuos G/A poseedores de mayores niveles que los portadores G/G. No se ha visto ninguna asociación en fumadores y mujeres. Saha et al. (1994) en China publicaron que los hombres no fumadores portadores del alelo A tenían significativamente mayores niveles de apoA-I que los G/G. No se encontró esta asociación en los niveles de HDL-C, y ningún efecto se ha detectado en fumadores o mujeres. Kamboh et al. (1996) han mostrado resultados similares en hombres no fumadores sobre los niveles de apoA-I. En un metaanálisis (Juo et al., 1999) sobre la asociación entre el polimorfismo A/G y los niveles de apoA-I, los autores concluyeron que los resultados provenientes de los individuos sanos sugieren que el raro alelo se asocia con niveles de apoA-I débilmente aumentados del orden de 5 mg/dL, y que esta asociación es más laxa entre mujeres que entre hombres. Aunque el estado tabáquico puede interactuar con los genotipos, solamente tres estudios han investigado esta interacción y no se pueden avanzar conclusiones al respecto.

Teniendo en cuenta el origen geográfico de las poblaciones mencionadas anteriormente y los resultados publicados en cada estudio, la interacción con PUFA podría explicar satisfactoriamente los resultados contradictorios. Los estudios que informan que el alelo A está asociado con mayores niveles de apoA-I y HDL-C se han realizado en países donde el consumo medio de PUFA es alto (Finlandia, Islandia, Inglaterra e Italia).

Según los resultados del estudio de esta tesis, cuando la ingesta de PUFA está por encima del 8% de energía/día, los individuos portadores del alelo A presentaban mayores niveles de HDL-C que los homocigotos G/G. Este efecto es estadísticamente significativo en mujeres y similar a lo largo de las distintas categorías de consumo de tabaco o alcohol. En los hombres, es necesaria una



interacción a cuatro bandas con tabaco y alcohol para alcanzar la significación estadística. Estos factores ambientales contribuyen a las diferencias previamente mencionadas en las asociaciones con HDL-C entre las categorías de consumo de tabaco y sexo. Para los niveles de apoA-I, la asociación era más alta en hombres que en mujeres, pero también debe considerarse una interacción a tres bandas con el consumo de alcohol. Las diferencias en la ingesta de PUFA y el descarte del consumo de alcohol puede explicar las discrepancias entre los resultados de los estudios previos. Las diferencias no significativas en los niveles de HDL-C o apoA-I ampliamente informadas en varios estudios son compatibles con la situación de ingesta intermedia de PUFA en la población. Una ingesta media de PUFA menor que 4% no es común, y el efecto aumentativo de G/G sobre los niveles de HDL-C no se ha mencionado en los estudios observacionales, y sólo uno de los estudios observó mayores niveles de apoA-I en los homocigotos G/G (Matsunaga et al., 1995).

Aunque no se han publicado estudios observacionales que comuniquen sobre la ingesta de PUFA en contraste con la hipótesis aquí planteada, los estudios de intervención han demostrado que el polimorfismo A/G tiene influencia sobre la respuesta lipídica a la grasa en dieta. En un trabajo de Mata et al. (1998) para investigar la influencia de este polimorfismo en la respuesta del LDL-C plasmático a la grasa saturada en la dieta, los autores mostraban que, en comparación con la dieta de SAFA, una dieta de PUFA inducía significativamente una gran respuesta plasmática. Es difícil distinguir las asociaciones de varios tipos de grasa con los niveles de lípidos en un estudio observacional debido a la multicolinealidad con la grasa. En este informe, se aplicaron en primer lugar modelos de regresión multivariada separadamente para cada tipo de grasa, siendo PUFA la más significativa. En segundo lugar, se incluyeron SAFA y MUFA en el modelo de regresión para PUFA como controles y el coeficiente de regresión para PUFA permanecía estadísticamente significativa. El papel de los ácidos grasos en la regulación de las concentraciones de lipoproteínas plasmáticas se ha documentado adecuadamente (Mensink et al., 1992). En comparación con SAFA, PUFA disminuye significativamente el LDL-C en los humanos. Altos niveles de PUFA pueden también disminuir las concentraciones de HDL-C plasmático (Mattson et al., 1985). Sin embargo, en estos estudios no se analizaron las variantes del promotor del gen APOA1.

Cuando se han considerado estas variantes en el estudio presentado aquí, el efecto depresor de PUFA sobre HDL-C se veía solamente en los homocigotos G/G. La respuesta de los individuos portadores del alelo A con respecto a una dieta rica en PUFA ha producido el efecto opuesto. Esta observación, que requiere posteriores investigaciones en estudios de intervención, puede ser de particular relevancia en los consejos dietéticos para modificar un perfil lipídico aterogénico.

Otra cuestión a tratar es la plausibilidad de que la ingesta de PUFA pueda modular los niveles de apoA-I o HDL-C. Se sabe que PUFA puede inducir la expresión de varios genes como se ha descubierto en modelos animales y en células cultivadas (Lopez-Miranda et al., 1994). Este sitio polimórfico, que se denota como -75 bp, -76 bp o -78 bp en distintos trabajos, está localizado en la región flanqueante 5' del gen de apoA-I del sitio del inicio transcripcional (start codon). Los estudios que utilizan el alelo más común G no ha identificado proteínas ligadas a este sitio específico (Danek et al., 1998). Sin embargo, la presencia de la sustitución por A crea una repetición de 6 bp que tiene homología a los sitios de ligamiento nuclear. Además, esta repetición puede permitir la formación de una estructura secundaria de ADN que podría interferir con la interacción proteínica del aparato transcripcional. Esto casa con los resultados de estudios en células cultivadas que muestran que el alelo A estaba asociado con baja actividad del promotor in vitro (Tuteja et al., 1992) (Smith et al., 1992) (Matsunaga et al., 1995). Se ha proporcionado evidencia adicional en apoyo de esta mutación y la transcripción del gen APOA1 por parte de los estudios metabólicos in vivo, que muestran que los sujetos con el alelo A tenían menores tasas de producción de apoA-I (Smith et al., 1992). Cuando la concentración de PUFA es alta estas moléculas pueden interactuar con el aparato transcriptor, y podrían incrementar la transcripción del gen APOA1. Se ha publicado que PUFA puede afectar la transcripción genética a través de múltiples mecanismos que pueden ser específicos con respecto a los promotores. Estos mecanismos incluyen principalmente una ruta dependiente del alfa receptor activado del proliferador de perioxosoma (en inglés *peroxisome-proliferator-activated receptor alpha*, abreviado como PPAR) y una ruta protanoide.

Aunque no se puede regular la posibilidad de introducir variables de confusión en la hipótesis de PUFA propia, es improbable que pueda explicar completamente la fuerte interacción observada. Este efecto depresor del alelo A sobre los niveles de HDL-C en mujeres también muestra una relación clara gen-dosis cuando se considera la interacción con PUFA. Los sujetos A/A tenían bajos niveles de HDL-C que los individuos heterocigotos A/G. Además, el efecto de elevación de PUFA sobre los niveles de HDL-C es mayor en los sujetos A/A que en los sujetos A/G. En conclusión, la ingesta de PUFA modula el efecto del polimorfismo G/A sobre el promotor del gen APOA1, elevando los niveles de HDL-C en los portadores del alelo A.

#### 5.2.4 Discusión del gen LIPC

El estudio sobre el gen LIPC en la población del *Framingham Offspring Study*, como una cohorte estadounidense correctamente caracterizada, destaca una

asociación altamente significativa entre genes y nutrientes. La asociación del polimorfismo -514C/T con las medidas del metabolismo de HDL-C depende de la cantidad y el tipo de grasa consumida. Los resultados del estudio realizado en el marco de la tesis confirman previos hallazgos, en esta población (Couture et al., 2000) y en otras poblaciones, que muestran que el alelo T está asociado con concentraciones más altas de HDL-C y HDL<sub>2</sub>-C y grandes tamaños de partículas (Guerra et al., 1997) (Vega et al., 1998) (Grundy et al., 1999) (Juo et al., 2001), cuando no se ha aplicado la estratificación por ingesta de grasa. Sin embargo, se ha encontrado que el alelo T estaba sólo asociado con mayores concentraciones y partículas de HDL-C en sujetos que consumían menos del 30% de energía de la grasa. Esta interacción gen-dieta podría ayudar a explicar algunos de los informes que muestran que no hay asociación entre el alelo T y el HDL-C (Tan et al., 2001) (Shohet et al., 1999) (Hegele et al., 1999).

Varios estudios han comunicado que el alelo T, que está en desequilibrio de ligamiento completo con otros tres polimorfismos de promotor (-250 G/A, -710 T/G y -763 A/C), está asociado con la actividad de HL plasmático marcadamente decreciente (Thuren et al., 2000) (Jansen et al., 1997), que, a su vez, puede conferir concentraciones y tamaño de HDL-C aumentados. Además, se ha mostrado suficientemente su funcionalidad (Deeb et al., 2000). HL es responsable de la lipólisis de las partículas remanentes de VLDL y gran LDL, así como de la conversión de HDL<sub>2</sub> a partículas de HDL<sub>3</sub>. En algunos estudios (Dreon et al., 1998) (Zambon et al., 1998), que no en otros (Juo et al., 2001) (Tan et al., 2001), un incremento en la actividad de HL se ha relacionado con el incremento de pequeñas partículas de LDL.

En el estudio que se presenta aquí, el tamaño de partícula de HDL aumentado relativo al alelo T no estaba asociado significativamente con grandes partículas de LDL (Couture et al., 2000), sugiriendo que otros factores genéticos o ambientales pueden modular esta asociación. Además, hay datos experimentales y epidemiológicos contradictorios sobre la asociación entre la actividad de HL y ECV (Thuren et al., 2000). Dugi et al. (2000) informaba que una baja actividad de HL es un factor de riesgo para ECV, consistente con los resultados de Jansen et al. (1997), y mostraba que los pacientes con deficiencia genérica de HL desarrollan ECV prematura, añadiendo la observación de que los ratones que sobreexpresan HL reducen la aterosclerosis a pesar de que se disminuya HDL-C (Thuren et al., 2000) (Santamarina et al., 1998). Estas asociaciones son paradójicas porque se ha descrito que la baja actividad de HL conduce a mayor HDL-C y a un perfil más antiaterogénico. Esta contradicción aparente se puede resolver mediante otras consecuencias de la actividad de HL. Por tanto, se ha descrito también un papel antiaterogénico para la HL porque la actividad de HL se correlaciona con las

concentraciones de lípidos postprandiales y la deficiencia de HL está asociada con aclaramiento disminuido de remanentes de lipoproteínas (Dugi et al., 2000).

Se posee un conocimiento limitado sobre la regulación genética de la HL en humanos. Se ha descrito que incluso entre los homocigotos -514C/T, la actividad de HL es bastante variable, indicando que otros factores modulan los efectos alélicos sobre la expresión de LIPC (Zambon et al., 1998). La actividad de HL es mayor en hombres fumadores que padecen diabetes, y aumenta con la grasa intraabdominal y el IMC (Tan et al., 2001) (Kong et al., 2001). Sin embargo, la información sobre los efectos de la grasa en dieta sobre la activada de HL es limitada y contradictoria. Campos et al. (1995) han descrito incrementos en las actividades de lipasa lipoproteínica (LPL) y HL en sujetos que consumían una dieta rica en grasa (46% de energía) comparados con los de dieta baja en grasa (24% de energía). Por el contrario, Kasim et al. (1993) han observado que la LPL, pero no la actividad de HL, decrecían durante la restricción de grasa en dieta en mujeres. Dreon et al. (1998) han informado de que un incremento en SAFA en dieta está asociada con actividad de HL decreciente, sugiriendo que los cambios en HL inducidos por dieta pueden contribuir a la regulación de grandes partículas de LDL.

Tomando la base de los resultados obtenidos en el trabajo de la tesis, se plantea la hipótesis de que el efecto de la grasa en dieta sobre la actividad de HL depende del polimorfismo -514C/T. Un descubrimiento importante en el estudio es la consistencia de esta interacción en la determinación de varias variables de salida relacionadas con el metabolismo de HDL y con la actividad de HL, después de haber controlado por posibles variables de confusión. Además, esta fuerte interacción ha mostrado un claro efecto dosis-respuesta. Se ha encontrado un efecto gen-dosis cuando se ha considerado la variable de grasa total como categórica y también como continua. Se han asociado los aumentos de grasa total con incrementos en las concentraciones de HDL-C en individuos CC y TT pero con una pendiente distinta. En contraste, se ha asociado en los individuos TT un incremento similar en la ingesta de grasa total con una disminución en HDL-C. Estos resultados están en consonancia con la observación general de que las concentraciones de HDL-C aumentan con las dietas ricas en grasa (Berglund et al., 1999) (Kasim et al., 1993) debido a la alta prevalencia de individuos CC y TT en la población general. Además, se ha identificado un subgrupo específico (sujetos TT) cuyo HDL-C puede responder distintamente a los cambios en la ingesta de grasa en dieta que los sujetos de la población general (sujetos CC).

En cambio, se dispone de menor información sobre los cambios en las subclases de HDL asociados con la cantidad de grasa consumida. Se ha encontrado que la ingesta más alta de grasa total estaba asociada con un aumento del tamaño de partículas de HDL solamente en los sujetos CC, mientras que lo contrario es

cierto para los individuos CT y TT. Berglund et al. (1999) han investigado el efecto de la reducción de grasa total y SAFA sobre la concentración y tamaño de HDL en un estudio aleatorio, que indica que las subpoblaciones de gran tamaño de HDL disminuye con la reducción de grasa en dieta. Considerando que los sujetos CC componen la mayoría de poblaciones blancas, los resultados globales están de acuerdo con la observación en estos individuos efectuada en el estudio de la tesis.

Cuando se han analizado los tipos específicos de FA, se han encontrado términos de interacción estadísticamente significativos con SAFA y MUFA, pero no con PUFA. Esto es consistente con el planteamiento de que la ingesta de SAFA y MUFA está asociada con el mayor HDL-C (Mensink et al., 1992) (Katan et al., 1997). También es consistente con las observaciones de que MUFA y SAFA están altamente correlacionados como resultado de los recursos compartidos de estos tipos de grasas (por ejemplo, carne y productos lácteos) en la población que se ha estudiado, se han encontrado solamente interacciones estadísticamente significativas para la grasa animal. El efecto de este origen de grasa es análogo a lo que se ha descrito para la grasa total, y el efecto de la ingesta de grasa animal sobre los tamaños mayores y menores de las subpoblaciones de HDL dependiente del genotipo de LIPC era más significativo. Por lo tanto, en los sujetos CC, los grandes tamaños de HDL disminuían con la reducción en grasa animal en dieta, y se ha visto un aumento relativo correspondiente para los tamaños menores de HDL. Lo contrario ocurría para los individuos CT y TT. Sin embargo, no se puede concluir que estos resultados son exclusivos porque los componentes grasos de la dieta, así como otros componentes de los alimentos animales, podrían jugar un papel determinado, especialmente los que son solubles en grasa. Otro tema que precisa mayor consideración y que no se ha tratado en los trabajos actuales, se refiere al papel que juegan los ácidos grasos trans en las interacciones anteriormente mencionadas.

El diseño del estudio de la tesis no puede dirigir el mecanismo por el cual la grasa en dieta interactúa con el polimorfismo -514C/T. Se ha informado que altas concentraciones de glucosa incrementa los niveles de ARN mensajero de LIPC (Tu et al., 2001), produciendo un fundamento molecular para el papel de la hiperglicemia en el metabolismo de HDL alterado. Estudios anteriores han comunicado que la insulina suprarregula la actividad de HL a través de los elementos de respuesta a la insulina en el promotor de LIPC, lo que sugiere que las variantes de este promotor pueden afectar la habilidad de la insulina para estimular la actividad de HL (Thuren et al., 2000). Jansen et al. (1997) han publicado que existe una asociación entre las variantes del promotor de LIPC y la resistencia a la insulina, lo que sugiere la presencia de algunos mecanismos potenciales compatible con los resultados del estudio que aquí se presenta.

En resumen, esta interacción gen-nutriente altamente significativa puede ayudar en la explicación de resultados conflictivos con respecto a la actividad de HDL, el perfil lipídico antiaterogénico y el riesgo cardiovascular, así como las diferencias intraindividuales en la respuesta de los lípidos plasmáticos a la grasa en dieta. Mas interesante es el hecho de que los efectos observados se han encontrado exclusivamente para la grasa animal pero no para la grasa vegetal. Este punto requiere posteriores investigaciones en estudios experimentales correctamente controlados.

### 5.2.5 Discusión del gen SR-BI

El estudio sobre el gen SR-BI en la población del *Framingham Offspring Study* destaca una asociación significativa entre las variantes genéticas de SR-BI y las concentraciones y tamaños de partículas que podrían contribuir a conocer mejor el papel de este gen en el metabolismo lipídico humano y en la diabetes de tipo 2.

La evidencia de que el SR-BI juega un papel fisiológico en la eliminación de HDL ha sido aportada por los modelos de ratones donde la sobreexpresión hepática de SR-BI, aplicando transferencia de genes mediada por adenovirus o enfoques transgénicos, provocaba la merma de HDL-C plasmático, el aumento del consumo de HDL-C selectivo en el hígado, y colesterol biliárico elevado (Ueda et al., 2000) (Krieger et al., 1999). En contraste, la reducción de la expresión de SR-BI, por mutaciones que rompen la región codificadora del gen, se producía en el ratón con aumento de HDL-C plasmático y disminución del consumo selectivo de HDL-C, todo ello con una dependencia gen-dosis (heterocigotos y homocigotos) (Rigotti et al., 1997) (Trigatti et al., 2000). La contribución específica del locus SR-BI a la homeostasis del metabolismo lipídico en humanos permanece prácticamente indefinida. Entonces, en la población que se ha estudiado, que está adscrita a esta cohorte norteamericana perfectamente caracterizada, se han encontrado asociaciones significativas entre los variantes del gen de SR-BI y las concentraciones y tamaños de partículas de lípidos y lipoproteínas.

Hasta la fecha, no se ha publicado ningún gran estudio de población que haya examinado la asociación entre los polimorfismos genéticos del gen SR-BI y las concentraciones de lípidos y lipoproteínas. Estudios anteriores incluían muestras de pequeño tamaño (Acton et al., 1999) (Durst et al., 2001) (Plat et al., 2002) y sus descubrimientos no han sido totalmente consistentes induciendo a pensar que se produjeron errores en la determinación de interacciones gen-gen, gen-ambiente, o en diabetes de tipo 2. Por tanto, Durst et al. (2001), en un estudio que incluía a 56 individuos judíos Ashkenazi con hipercolesterolemia familiar, no encontró ninguna asociación significativamente estadística entre el polimorfismo del exón 8 (la única

variante analizada) en el locus del gen SR-BI y el colesterol total plasmático, el HDL-C plasmático, el LDL-C plasmático o los triglicéridos. Sin embargo, el equipo de Durst observaron altas concentraciones de HDL-C en portadores del alelo T en comparación con pacientes que son homocigotos para el alelo común C. De forma similar, Plat et al. (2002) en Holanda, examinaron a 41 hombres y 71 mujeres que eran jóvenes (media de edad de 33 años), saludables y no hipercolesterolémicos. Aunque el principal objetivo de estos autores fue el estudio de la respuesta a la dieta, también examinaron la asociación entre el polimorfismo del exón 8 del SR-BI y los lípidos y lipoproteínas plasmáticos al inicio del estudio. Encontraron una diferencia ligeramente significativa ( $p = 0,069$ ) entre las concentraciones de HDL-C plasmático en portadores del alelo T ( $1,56 \pm 0,36$  mmol/L) y homocigotos para el alelo C ( $1,72 \pm 0,42$  mmol/L). En contraposición a Plat et al. (2002), y de acuerdo con Durst et al. (2001), en el presente estudio se han encontrado menores concentraciones de HDL-C plasmático en los homocigotos CC en comparación con los individuos portadores del alelo T, que tiene una tamaño de partícula específico.

Es estudio que aquí se presenta es el primero en examinar la asociación entre las variantes del gen SR-BI y el tamaño de partículas de lipoproteínas. Se ha encontrado un incremento estadísticamente significativo en el diámetro de partícula de HDL en relación con el alelo T en hombres y mujeres. Esto es consistente con los descubrimientos en los modelos animales (Rigotti et al., 1997) (Varban et al., 1998) (Ueda et al., 2000) (Ueda et al., 1999). El incremento en el tamaño de partícula de HDL-C estaba consistentemente asociado con las más altas concentraciones de gran HDL-C observado en los sujetos CT y TT, con un claro efecto codominante en mujeres.

Estos descubrimientos, magnificados en el análisis de haplotipos entre en exón 8 y el intrón 5, son consistentes con la noción de que la expresión o la actividad de SR-BI disminuye en portadores del alelo T en el polimorfismo del exón 8. Si el SR-BI hepático juega un papel crucial en la remoción selectiva del colesterol en el HDL circulante en humanos, se esperaba que los bajos niveles de SR-BI hepático aumentasen el HDL-C plasmático pero no alterasen directamente las concentraciones de apoA-I como ocurre en el ratón (Rigotti et al., 1997) (Krieger, 1999) (Trigatti et al., 2000). Los datos del exón 8 obtenidos en el presente estudio confirman fehacientemente este supuesto porque no se han encontrado diferencias en las concentraciones de apoA-I entre hombres y mujeres en los genotipos del exón 8.

Los portadores del alelo T en el polimorfismo del exón 8 del SR-BI también tenían concentraciones reducidas de LDL-C en comparación con los homocigotos CC. Este efecto, estadísticamente significativo en mujeres en el análisis de

genotipos, y significativo en hombres en el análisis de haplotipos, era consistente con un estudio previo realizado con población europea del sur (Acton et al., 1999).

Estudios experimentales tempranos han mostrado que el receptor de SR-BI podía ligarse con las lipoproteínas HDL y LDL en una competición cruzada no recíproca (Acton et al., 1996) (Krieger, 1999). Subsiguientes estudios *in vitro* y en animales suponían que el SR-BI juega un papel en el metabolismo de LDL en humanos. Los resultados del presente estudio para el exón 8 difieren de lo se habría esperado según el modelo del ratón transgénico deficiente en receptores con expresión atenuada del SR-BI (Huszar et al., 2000) que mostraban concentraciones de LDL-C aumentadas. La disminución en humanos de las concentraciones de LDL-C, que se han observado en portadores del alelo T, son comparables con una hiperexpresión, más que con una hipoexpresión, del SR-BI como se había formulado hipotéticamente para esta variante genética. Estas observaciones, y el hecho de que el cambio C→T en el exón 8 no provocaba ningún cambio en la secuencia de aminoácidos, sugiere que el SNP del exón 8 podría estar en desequilibrio de ligamiento con una mutación funcional en el gen SR-BI, o alternativamente con otra variante funcional en un locus vecino, donde se han localizado otros genes candidatos implicados en el metabolismo lipídico como, por ejemplo, ACACB, PLA2, CLTA, MVK, ACADS y TCF1.

Ocurre que en un trabajo previo sobre una población del sur de Europa se han detectado asociaciones estadísticamente significativas entre la variante genética del intrón 5 y los triglicéridos plasmáticos en hombres, y con el IMC en mujeres. Eso sugiere que haya un posible enlace entre esta variante y el síndrome metabólico (Acton et al., 1999), pero en la población americana no se ha observado ninguna asociación significativa. Las diferencias en edad, IMC, la composición ambiental y genética de las poblaciones estudiadas, pueden producir distintas interacciones gen-ambiente, así como diferentes desequilibrios de ligamiento entre las variantes genéticas, que podrían ser las bases para las diferencias en los resultados observados.

Además, cuando se ha comprobado directamente la hipótesis sobre la asociación posible entre la variación del gen SR-BI y la diabetes de tipo 2 en el *Framingham Offspring Study*, no se han encontrado diferencias en la prevalencia para cualquiera de las variantes genéticas del SR-BI entre diabéticos de tipo 2 y no diabéticos. Todavía más, se han examinado los efectos de las interacciones entre el estado diabético y las variantes del gen SR-BI en la determinación de los lípidos y lipoproteínas plasmáticos y de los tamaños de partículas. Para ello se ha considerado que la diabetes de tipo 2 está caracterizada por bajas concentraciones de HDL-C plasmático, altas concentraciones de triglicéridos, una preponderancia de pequeño y denso LDL, un incremento en lipoproteínas oxidadas, así como por



otros parámetros relativos a la resistencia a la insulina, que pueden modular la expresión del gen SR-BI (Taskinen et al., 2002) (Goff et al., 2000). A este respecto, Milliat et al. (2000) informaron de que la diabetes inducida por estreptozotina en ratas Rico producía la modificación de la expresión de los receptores de lipoproteínas del hígado, que pueden contribuir en las alteraciones del perfil de lipoproteínas. En particular, estos investigadores mostraron que en ratas diabéticas apareció una temprana y fuerte mejora de la expresión hepática de SR-BI en el sexto día y persistió hasta el día 28, pero con menor extensión.

En el estudio que se presenta, no se han encontrado interacciones significativas con diabetes y SR-BI para las variantes del exón 8 y el intrón 5 en la determinación del perfil lipídico. Sin embargo, para el SNP del exón 1 se ha observado una modificación consistente de la asociación. Aunque en la población global el polimorfismo del exón 1 no estaba asociado con ningún lípido ni lipoproteína, esta variación genética influye particularmente el perfil lipídico en los sujetos diabéticos. Los sujetos diabéticos portadores del alelo menos común A en el exón 1, en comparación con los diabéticos homocigotos GG, habían disminuido las concentraciones de HDL-C plasmático total. Este decremento se encontraba principalmente en la subfracción grande de HDL-C y esto se notaba por la reducción media del diámetro de partícula de HDL-C, alcanzando la significación estadística en las concentraciones de HDL<sub>2</sub>-C. Estas observaciones sugieren que el alelo A en sujetos diabéticos está relacionado con una expresión hepática del gen SR-BI aumentada, que ha sido bien caracterizada en estudios experimentales por reducción en HDL-C, motivado por el incremento de la eliminación selectiva del colesterol en el plasma (Kozarsky et al., 1997) (Krieger et al., 2001). Además, se ha mostrado que el tamaño de HDL, la composición y estructura de lipoproteínas influye en el ligamiento de HDL al SR-BI (de Beer et al., 2000) (de Beer et al., 2001). Por tanto, el SR-BI expresado en células ováricas de hámster transfectadas se une al HDL<sub>2</sub> humano con mayor afinidad que el HDL<sub>3</sub>, y esta afinidad de ligamiento era significativamente alta (50 veces) para partículas grandes frente a las partículas pequeñas. Esto lleva a la conclusión de que el ligamiento preferencial de grandes partículas de HDL promocionarían el consumo eficiente de éster de colesterol a partir del éster de colesterolina grande enriquecido en HDL frente a lípidos pobres en HDL (de Beer et al., 2001).

Los hallazgos en el estudio que se presenta para el exón 1 son consistentes con estos resultados experimentales. Además, se ha observado también un decremento fuerte y estadísticamente significativo de las concentraciones de LDL-C plasmático total en los sujetos diabéticos portadores del alelo A del exón 1. Esto está de acuerdo con la observación de disminución de concentraciones de LDL-C encontradas en estudios animales donde se hiperexpresaba el SR-BI (Wang et al., 1998) (Ueda et al., 1999). En contraste, con el SNP en el exón 8 y el SNP en el

exón 1 (G → A) se codifica un cambio de glicina a serina en el aminoácido de la segunda posición que modifica uno de los sitios potenciales de acilación en la molécula de SR-BI, confirmando esta función. Sin embargo, Gu et al (1998) no encontraron diferencia en la acilación de un mutante in vitro (Gly → Ala) transitoriamente expresado en la células COS (*Carcinoma Ovary*).

Si bien no hay estudios en humanos que hayan examinado la asociación entre la expresión de SR-BI y la aterosclerosis, hay varios estudios de modelos animales que son consistentes en sus resultados y muestran una aterosclerosis decreciente asociada con la mayor expresión de SR-BI (Krieger, 2001). Por tanto, en los modelos de aterosclerosis en roedores (el ratón knockout de apoE y el ratón knockout de LDLR con dieta rica en grasa o colesterol) la ausencia de SR-BI acelera drásticamente la aparición de aterosclerosis (Braun et al., 2002). Mientras que la aterosclerosis se suprime por la hiperexpresión hepática de este gen (Arai et al., 1999) (Kozarsky et al., 2000). Los hallazgos en los modelos animales son aparentemente paradójicos. Sin embargo, debe destacarse que la hiperexpresión hepática de SR-BI, a pesar de la disminución en las concentraciones de HDL-C plasmático total, aumenta el transporte inverso de colesterol y por ende la remoción del colesterol del cuerpo. Además, la hiperexpresión de SR-BI podría disminuir también las lipoproteínas aterogénicas que contienen apoB, como se ha observado claramente en los sujetos diabéticos portadores de la variante del exón 1. Esto sugiere que este efecto protector puede jugar un papel significativo contra la aterosclerosis.

Es necesario investigar más si los portadores del raro alelo A en el exón 1, sean o no diabéticos, están protegidos contra la aterosclerosis. La diabetes induce también algunos cambios en los lípidos y lipoproteínas que pueden ser moduladores importantes. Al respecto, se ha descrito que el contenido elevado de triglicéridos, que ocurre frecuentemente en la diabetes de tipo 2, puede afectar a los efectos beneficiosos de la hiperexpresión de SR-BI (Greene et al., 2001). Además, a pesar de la reducción significativa en LDL-C total que se ha visto en los sujetos diabéticos portadores del alelo A, esta reducción ocurre inicialmente en las partículas grandes de LDL. Esta disminución en el tamaño de partículas de LDL se ha documentado también en estudios animales y se ha explicado como consecuencia del consumo selectivo, donde el receptor de SR-BI tiene menos afinidad hacia las partículas pequeñas de LDL (Trigatti et al., 2000). Por otra parte, existe una evidencia cada vez mayor de que el SR-BI puede jugar otros papeles de especial importancia en la diabetes de tipo 2. Por tanto, el SR-BI, como receptor de recolector de residuos, puede también ligarse a LDL oxidado, que confiere apoyo a sus propiedades antiaterogénicas (Gillotte et al., 2001, Iwashima et al., 2000).

En resumen, se ha mostrado que el locus de SR-BI está asociado significativamente con las concentraciones de lípidos plasmáticos y tamaños de partículas de lipoproteínas. Además, parece que existe una interacción significativa entre la variación en este locus y la diabetes de tipo 2 en la determinación del perfil lipoproteico. Los datos del estudio presentado sugieren que, a través de su asociación con los factores de riesgo lipídicos, el SR-BI puede modular el riesgo de aterosclerosis en la población general y especialmente en los sujetos diabéticos.

### 5.2.6 Discusión del gen PLIN

El estudio sobre el gen PLIN en la población de la Comunidad Valenciana destaca una asociación significativa entre algunas de las variantes genéticas de PLIN y la obesidad. En este estudio se ha investigado la asociación de cuatro nuevos polimorfismos de PLIN con medidas de obesidad, metabolismo lipídico y sensibilidad a la insulina en una muestra de individuos caucásicos. Por primera vez se ha descubierto que las variaciones en el locus del PLIN humano están asociadas con las variables de obesidad, lo que sugiere que las perilipinas pueden jugar un papel relevante en la obesidad humana y posiblemente en el desarrollo del síndrome metabólico.

Aunque los datos de los modelos animales han mostrado con coherencia que la disrupción dirigida del gen de perilipina provoca en los ratones sanos que sean más delgados y musculosos que los controles y que son resistentes a la obesidad inducida por la dieta (Martinez-Botas et al., 2000) (Tansey et al., 2001), hasta la fecha no hay estudios epidemiológicos en los humanos en los que la variación en el gen PLIN se haya relacionada significativamente a los fenotipos de obesidad. En el transcurso del estudio que aquí se presenta, Mottagui-Tabar et al. (2003) publicaron un artículo en el que examinaban la variación del gen PLIN en 117 mujeres obesas de Suecia. Después de la selección de siete SNP potenciales los autores solamente describían como polimórficas dos de las siete variantes (rs894160 y rs1052700) en esta pequeña muestra. La primera variante corresponde con el polimorfismo PLIN4, localizado en el intrón 6, y la segunda al polimorfismo PLIN6 localizado en el exón 9. Los autores informaron de que la frecuencia alélica para el alelo menos común en el locus PLIN4 en mujeres obesas era de 0,24, en completa consonancia con las estimaciones del estudio que se presenta. Sin embargo, la frecuencia para el alelo menos común en el locus PLIN6 era más alta que en el presente estudio, ilustrando las diferencias potenciales entre poblaciones europeas del norte y del sur. En contraste con los resultados del presente estudio, Mottagui-Tabar et al. (2003) no encontraron asociaciones estadísticamente significativas entre las variantes genéticas e IMC, glucosa plasmática, o TG en las mujeres obesas examinadas. Desgraciadamente, no estudiaron hombres obesos o sujetos de

la población general para efectuar la comparación de sus resultados con los del presente estudio.

Una de las principales ventajas del presente estudio es que realiza un estudio epidemiológico cuidadosamente controlado en un muestra grande combinando dos enfoques distintos, estudio transversal y de casos y controles, en dos grupos de sujetos: uno extraído de la población general y otro de pacientes del hospital. Este diseño mixto ha permitido estimar propiamente la prevalencia de las variantes de PLIN en la población general, para poder estudiar un amplio rango de fenotipos de obesidad, superar la limitación de que las muestras hospitalarias de pacientes obesos no se pueden considerar representativas de la población general, y para proveer una muestra aleatoria de controles representativa de la población de base con el propósito de prevenir el sesgo con el uso de los controles del hospital en el caso de diseño de casos y controles (Little et al., 2002) (Wacholder et al., 1992).

Como resultado de este enfoque mixto, se ha encontrado consistentemente que, en las mujeres, el alelo 2 en los polimorfismos de PLIN1 y PLIN4 (ambos en fuerte desequilibrio de ligamiento) se puede considerar como una variante genética protectora en términos de obesidad e IMC. En el estudio transversal en mujeres de la población general, que son portadoras del alelo 2 en PLIN1 o en PLIN4, tenían estadísticamente menor IMC que las mujeres homocigotas con el alelo común. En términos de peso, estos alelos se han asociado con una disminución media de 2,2 Kg (sobre un 3,5% del peso corporal), que es una figura clínicamente relevante teniendo en cuenta que el control adicional para las covariantes. En los pacientes obesos, estos alelos se han asociado con menor IMC y peso (una disminución de – 5,8 Kg (5%) en el caso de PLIN4). Sin embargo, los resultados no han alcanzado la significación estadística debido al pequeño número de mujeres severamente obesas. Por consiguiente, cuando la heterogeneidad de estos efectos alélicos debido al origen de la muestra se ha comprobado estadísticamente, se ha obtenido una homogeneidad, y en la muestra global de mujeres, se han encontrado asociaciones estadísticamente significativas con el menor IMC. Además, en el enfoque de casos y controles, las mujeres portadoras del alelo 2 en los polimorfismos de PLIN1 y PLIN4 tenían dos veces menos riesgo de obesidad que las mujeres homocigotas para el alelo más común.

El menor riesgo de obesidad se relacionaba con los alelos menos comunes de los polimorfismos PLIN1 y PLIN4, que son hallazgos análogos en el ratón sin el gen de la perilipina ligado a la ablación de perilipina con un fenotipo pobre (Martinez-Botas et al., 2000) (Tansey et al., 2001). Esto apoya la hipótesis de que el alelo 2 en estos loci se pueda asociar con la menor expresión del gen PLIN o con la reducida actividad de perilipina. Además, se ha publicado que en los modelos animales este tipo de desactivaciones del gen PLIN también protegía del desarrollo

de obesidad al ratón *Lepr(db/db)*, un modelo genético de obesidad causada por la resistencia a la leptina (Tansey et al., 2001). El ratón sin el gen de la perilipina tenía un 62% menos de adipositos blancos (Martinez-Botas et al., 2000) y mostraba una lipólisis basal elevada debido a la pérdida de función protectora de la perilipina (Martinez-Botas et al., 2000) (Tansey et al., 2001). En los humanos, un contenido bajo en perilipina probablemente está directamente ligado a la tasa de lipólisis (et al., 27). Sin embargo, hay controversias sobre si la obesidad está asociada con un incremento o decremento de la tasa lipolítica (Frayn et al., 2003) (Nielsen et al., 2003) (Arner, 2001). Se ha argumentado que algunas inconsistencias vienen de las diferentes unidades y situaciones en que esto se ha medido (Nielsen et al., 2003). Según el ratón sin el gen de la perilipina, se debería encontrar una lipólisis perjudicial en los pacientes obesos, y hay algunos estudios en humanos que confirman esta situación (Hellstrom et al., 1996) (Schiffelers et al., 2003).

Aunque ninguno de los polimorfismos examinados en el estudio que se presenta aparece como tradicionalmente funcional (los PLIN1 y PLIN4 son mutaciones intrónicas, el PLIN5 es una mutación silente en el exón 8, y el PLIN6 está en la región no trasladada del exón 9), el trabajo de Mottagui-Tabar et al. (2003) confirma la funcionalidad del polimorfismo PLIN4. Estos autores han demostrado que el polimorfismo PLIN4 estaba asociado estadísticamente con las diferencias en la lipólisis basal e inducida por diaminasa adenosina en células de grasa aisladas de mujeres obesas. Las mujeres homocigotas para el alelo 2 tenían tasas medias de lipólisis doblemente mayores que las homocigotas para el alelo 1, mientras que las heterocigotas tenían tasas medias intermedias ( $p < 0,05$ ). Además, las portadoras del alelo 2 en el locus PLIN4, con lipólisis inducida por noradrenalina, tenían una lipólisis un 50% más alta que las homocigotas para el alelo común. También demostraron que el contenido de proteína perilipina estaba destacadamente disminuido en los portadores del alelo 2 (Mottagui-Tabar et al., 2003).

Como apoyo de estos vínculos, hay alguna evidencia que sugiere que los polimorfismos intrónicos también deberían regular la expresión genética actuando sobre el ligamiento de los factores nucleares (Horikawa et al., 2000). Una hipótesis alternativa es que estos polimorfismos están directamente implicados en dicha regulación, o en el LD con mutaciones que alteran el splicing del ARN mensajero. Los PLIN4, PLIN5 y PLIN6 están todos próximos a las regiones sujetas al splicing alternativo (Figura 3.3 –Capítulo 3). Todas las perilipinas comparten un término amino 22-kDa idéntico con distintas secuencias de carboxilos terminales de longitud variable (Lu et al., 2001). Las dos principales variantes de splicing del gen PLIN, perilipina A y perilipina B, mostraban distinta respuesta a la activación de PKA y deberían ejercer una protección diferente contra la lipólisis. Las diferencias estructurales entre estas variantes de splicing pueden determinar sus funciones,

especialmente la longitud de la cola terminal C que afecta a la envoltura de la superficie de las gotas (Tansey et al., 2003) (Lu et al., 2001).

Otro descubrimiento interesante en los resultados del estudio que se presenta es la ausencia de asociaciones significativas entre la variación de PLIN y las variables de obesidad en hombres de la población general, resaltando la importancia de los factores hormonales sexuales en la regulación del peso corporal y la distribución de grasa en humanos. Los efectos de los genotipos de PLIN específicos por sexo son consistentes con las bien conocidas diferencias entre hombres y mujeres en el desarrollo y distribución del tejido adiposo, así como con los riesgos de enfermedades relacionadas con la obesidad (Gasteyger et al., 2002). Se ha publicado que la lipólisis en el tejido adiposo es altamente dependiente del sexo (Nielsen et al., 2003) (Lofgren et al., 2002) (Kolehmainen et al., 2002), dando a entender que son necesarios nuevos enfoques, que consideren estas diferencias por sexo, para evaluar cómo las variaciones en la actividad lipolítica deberían causar impacto en los fenotipos de obesidad.

Según los resultados del estudio que se presenta, los estudios epidemiológicos que se centran en la variación del gen HSL, la enzima limitante en la lipólisis de adipositos intracelulares que actúa en una reacción concertada con la perilipina A (Sztalryd et al., 2002), también han mostrado asociaciones específicas por sexo relevantes. Por tanto, Lavebratt et al. (2002), en un estudio de casos y controles en Suecia, informaron de que la homocigosidad para HSLi6 A5 era un factor de obesidad en mujeres (OR: 1,9; 95% CI 1,6-2,2). Es también interesante que los estudios funcionales (Hoffstedt et al., 2001) han mostrado que el alelo 5 del polimorfismo HSLi6 está asociado con una destacada disminución en la tasa lipolítica de las células de grasa abdominales, que a su vez puede contribuir al desarrollo de la obesidad como se ha formulado en la hipótesis para los homocigotos 1/1 en el locus del PLIN4.

Por otra parte, se ha encontrado que el alelo 2 en el locus de PLIN4 está asociado con menor peso en paciente obesos. La razón de esto es todavía desconocida, aunque es probable que la mayor adiposidad observada en hombres severamente obesos, en comparación con los hombres de la población general, contribuya ampliamente a esta asociación.

Aparte del efecto de la variación de PLIN sobre las variables antropométricas, se han encontrado asociaciones destacables con glucosa en ayunas y TG que dependen del estado de obesidad. En mujeres de la población general, el alelo 2 en el locus de PLIN4 estaba asociado con menores concentraciones de glucosa en ayunas y TG. Sin embargo, en mujeres severamente obesas, este alelo estaba relacionado con las concentraciones más altas. En hombres severamente obesos sí

se ha encontrado este incremento en glucosa en ayunas y TG. Estos resultados están de acuerdo con las observaciones de Tansey et al. (2001) en el ratón knockout de perilipina y se reconcilian con los descubrimientos de Martínez-Botas et al (2000). La liberación de ácido graso del tejido adiposo está implicada en el desarrollo de la diabetes de tipo 2 (Nielsen et al., 2003) (Arner, 2002), por ella cabría esperar que el ratón sin el gen de la perilipina sea más susceptible a desarrollar resistencia a la insulina. Martínez-Botas et al. (2000) no pudieron detectar intolerancia a la glucosa en sus animales nulos de perilipina, lo que sugiere que los ácidos grasos libres producidos en las células grasas pueden ser reesterificados in situ, en el ciclo fútil de la lipogénesis y la lipólisis. Tansey et al. (2001), reprodujeron los hallazgos de Martínez-Botas et al. (2000) en animales de menos de 30 g de peso. Sin embargo, a medida que los animales excedían de 30 g se desarrollaba una intolerancia a la glucosa significativa en los ratones nulos de perilipina frente a los ratones naturales. Reunidos todos estos resultados se ve que son consistentes con la noción de que la perilipina, que protege contra la obesidad, puede producir un fenotipo más perjudicial una vez el individuo se convierte en obeso.

En conclusión, el estudio que aquí se presenta es el primero en informar sobre asociaciones entre los genotipos del PLIN y las medidas de obesidad en humanos. Aunque la variante genética representada por el alelo 2 del PLIN4 parece estar consistentemente asociada con menores IMC en mujeres y en hombres obesos, sus efectos concomitantes sobre la glucosa en ayunas y TG dependen del grado de obesidad, produciendo un perfil peor en pacientes severamente obesos. Esto es consistente con estudios previos en adipositos cultivados así como con datos provenientes de modelos animales. Como colofón, de la misma forma que en cualquier trabajo inicial, el presente trabajo debe considerarse dentro de los hallazgos preliminares que deben ser reproducidos en otras poblaciones.

### **5.3 Discusiones de la parte bioinformática**

En esta sección se presenta la discusión individualizada sobre cada una de las aportaciones en esta tesis en el ámbito de la Bioinformática y disciplinas asociadas, tales como las herramientas de tratamiento de los datos de instrumentos de laboratorio, de tratamiento y visualización de secuencias de ADN, de compresión y búsqueda de secuencias de ADN, y de los enfoques de control y auditoría propuestos.

### 5.3.1 Discusión sobre PLATEX

Se ha presentado PLATEX, un pequeño sistema software para la gestión y la conversión de datos entre diferentes estructuras de datos para la utilización con el *ABI PRISM® 3100 Genetic Analyzer*. Este sistema es un conjunto de macros de Microsoft® Excel™ diseñadas en forma de arquitectura de componentes cuya interfaz de usuario es una de las hojas de cálculo que componen el libro Excel™ de trabajo.

La distribución de PLATEX en forma de libro de trabajo es para agrupar interfaces, hojas de trabajo y hojas de parámetros en un fichero de objetos (hojas, celdas, etc.) y componentes. Cada componente es una macro que ejecuta una tarea simple. Esto facilita la modificación o sustitución de cada componente debido al cambio de las condiciones externas, o la adición de nuevas macros. Además, ya que las macros están escritas en VBasic™ (desde la versión 97), son bastante simples pero potentes a la vez. Microsoft® VBasic™ es un lenguaje de programación orientado a objetos y que presenta un entorno orientado a componentes. Cada macro tiene las líneas de instrucciones suficientemente comentadas de forma que es fácil modificarlas o crear nuevas para los investigadores que no sean profesionales informáticos. El entorno proporciona un editor de macros y las mismas se pueden exportar e importar como ficheros de texto<sup>3</sup>.

El sistema propuesto en este trabajo puede ayudar a los investigadores en Genómica Funcional, o a cualquiera que trabaje con instrumental de análisis de ADN, en la gestión de la gran cantidad de datos producidos en los procesos de fragmentación de ADN. De hecho, desde el punto de vista del rendimiento temporal, este sistema es unas 300 veces más rápido que si tuviera que hacer manualmente, completando el proceso en sólo 30 segundos para placas de 96 pocillos. A pesar de que la solución descrita se ha basado en un instrumento de laboratorio concreto, su aplicación puede generalizarse a otros instrumentos similares con la modificación adecuada de los parámetros del libro de trabajo.

### 5.3.2 Discusión sobre la visualización y manejo de secuencias

Como una de las aportaciones en el marco de la integración de soluciones bioinformáticas para el soporte de la investigación en los laboratorios de Genómica Cardiovascular y Biología Molecular, se ha elaborado la herramienta SeqPacker, que permite la lectura, manipulación, visualización y almacenamiento de secuencias de ADN y ARN. Dado que se ha construido con Java permite que pueda ejecutarse en varios tipos de plataformas distintas como MS Windows, Linux y



MacOS X. Además, esta herramienta se ofrece libremente y se puede bajar desde la URL <http://www.iris.uji.es/ocoltell/seqpacker.html>.

SeqPacker pretende ser una herramienta simple, aunque potente, para que la utilicen quienes no están muy familiarizados con otras herramientas de ADN más complejas y necesitan de soluciones capaces de ahorrarles tiempo en las tareas diarias de investigación en el laboratorio. Por lo tanto, se ha diseñado su funcionamiento e interfaz de forma que sea lo mas amigable posible.

Por otra parte, dado su diseño modular, no se puede afirmar que esta herramienta ya está completa, sino que todavía puede incorporar nuevas funcionalidades. El Dr. Larry Parnell, bioinformático del Nutrigenomics Laboratory del HNRCA, a las órdenes del Dr. Ordovás, y con el cual el doctorando ha tenido el placer de colaborar, aplicó un test sobre los ficheros que él maneja e hizo una revisión funcional y ergonómica de la versión actual sugiriendo algunas ampliaciones y mejoras para poder aplicarla en otras áreas de investigación. Las sugerencias más importantes son las siguientes:

1. Que se pueda mezclar la cadena principal con la cadena patrón pero evitando que se intente casar la combinación de cadenas directas e inversas, ya que esto no tiene ningún sentido biológico o práctico.
2. Que se pueda insertar una secuencia entera entre dos nucleótidos de otra secuencia en forma de “selección de sitio de inserción”. Esto podría ser útil para simular la inserción de genes en un plásmido para replicar ADN (Figura 5.1) (PLASMID-1, 2004) (PLASMID-2, 2004).
3. Que se pueda trabajar con enzimas de restricción. Es decir, que haya un mecanismo trabajando por debajo que, en función de las características de cada enzima en particular, haga búsquedas automáticas en la secuencia principal (RESTRICTION, 2004).
4. Ampliación de la interfaz y las funciones asociadas para poder trabajar con secuencias de aminoácidos.

SeqPacker no pretende competir con otras herramientas de manipulación de ADN que permiten incluso la búsqueda no exacta. Sin embargo, estas herramientas tienen de entrada mayor dificultad en la comprensión de su interfaz y no llegan a resultar útiles para tareas muy sencillas, como, por ejemplo, la búsqueda de primers en una secuencia dada.

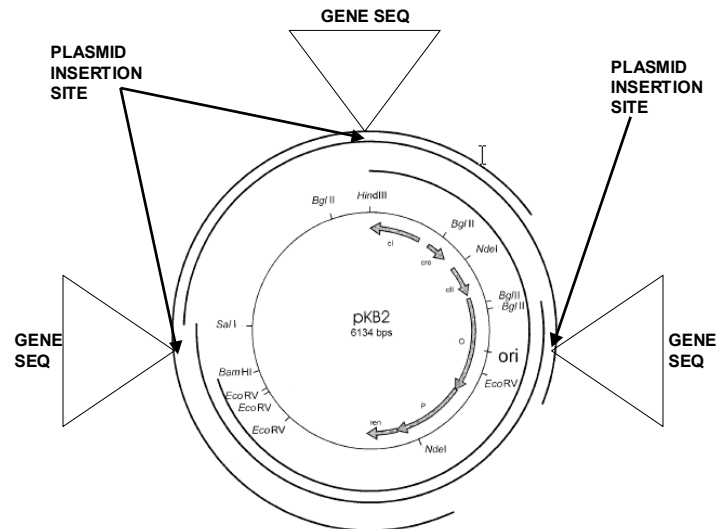


Figura 5.1. Replicación de AND con secuencias de plásmidos

Con estas consideraciones, no se ha dado por cerrado el capítulo del SeqPacker con esta tesis, sino que se continua trabajando en ella para obtener, en un futuro muy próximo, la siguiente versión operativa que pueda incorporar, si no todas, algunas de las sugerencias del Dr. Parnell.

### 5.3.3 Discusión sobre la búsqueda y detección de polimorfismos

Se presenta una experiencia en el desarrollo de un proyecto bioinformático en el que se necesita trabajar con bases de datos genómicas públicas. El proyecto ha consistido en la confección de un Modelo Cardiovascular Genómico para una utilización posterior en otros proyectos.

Por medio de la realización de un proceso con cinco actividades sucesivas, se ha obtenido una base de datos con información sobre genes asociados a las enfermedades cardiovasculares, sus polimorfismos asociados y los SNP correspondientes.

En el proceso descrito en la sección anterior se han producido una serie de dificultades e incidencias científicas y técnicas que no estaban previstas en el diseño del proyecto. La experiencia obtenida en el proyecto piloto se empleará en la modificación del diseño del proyecto para acometer la identificación y documentación de polimorfismos de la lista completa de genes. No obstante,

también resulta interesante informar de estas dificultades e incidencias para poner sobre aviso a los colaboradores de INBIOMED para cuando aborden trabajos de índole similar. Las incidencias se han agrupado en las siguientes categorías:

1. Identificación de los SNP de los polimorfismos.
2. Identificación de los polimorfismos en PubMed.
3. Desarrollo del proyecto.

A lo largo de esta sección se dan los detalles pertinentes.

### 5.3.3.1 Identificación de los SNP de los polimorfismos

Una de las primeras dificultades encontradas es que no hay una nomenclatura fija para identificar los polimorfismos. En la literatura publicada se utilizan paralelamente los tres tipos de nomenclaturas siguientes:

- Por nombre del enzima de restricción (sistema clásico): PstI, EcoRI, HintIII.
- Variación de nucleótido y posición en el gen: -75 A>G, 1567 A>G, 52 G>A.
- Variación de aminoácido y posición en la proteína: D19H, glut/arg 3500, pro2712-to-leu, Q604E.

En segundo lugar, no hay unicidad de estructura ni siquiera en las variantes de nomenclatura. Por ejemplo:

- Coexisten varias formas: D19H, glut/arg 3500, pro2712-to-leu, Q604E.
- Se usan dos nomenclaturas paralelas en aminoácidos: con una letra y con tres letras (Figura 5.2).

En tercer lugar, en el tratamiento de la subnomenclatura de variación de nucleótido y posición en el gen para identificar los SNP se producen las siguientes situaciones:

- Hay errores de cuenta en algunos polimorfismos informados: por ejemplo, para el mismo polimorfismo, unos lo identifican como “-1750G/T”, otros como “210G/T”.

|   | Aminoacid     | Amino3 | Amino1 | AminoList |
|---|---------------|--------|--------|-----------|
| ▶ | alanine       | Ala    | A      | Ala [A]   |
|   | arginine      | Arg    | R      | Arg [R]   |
|   | asparagine    | Asn    | N      | Asn [N]   |
|   | aspartate     | Asp    | D      | Asp [D]   |
|   | cysteine      | Cys    | C      | Cys [C]   |
|   | glutamate     | Glu    | E      | Glu [E]   |
|   | glutamine     | Gln    | Q      | Gln [Q]   |
|   | glycine       | Gly    | G      | Gly [G]   |
|   | histidine     | His    | H      | His [H]   |
|   | isoleucine    | Ile    | I      | Ile [I]   |
|   | leucine       | Leu    | L      | Leu [L]   |
|   | lysine        | Lys    | K      | Lys [K]   |
|   | methionine    | Met    | M      | Met [M]   |
|   | phenylalanine | Phe    | F      | Phe [F]   |
|   | proline       | Pro    | P      | Pro [P]   |
|   | serine        | Ser    | S      | Ser [S]   |
|   | threonine     | Thr    | T      | Thr [T]   |
|   | tryptophan    | Trp    | W      | Trp [W]   |
|   | tyrosine      | Tyr    | Y      | Tyr [Y]   |
|   | valine        | Val    | V      | Val [V]   |
| * |               |        |        |           |

Figura 5.2. Nomenclaturas paralelas en aminoácidos

- Se han tomado distintos orígenes para empezar a contar los nucleótidos:
  - Start Codon: es el que se toma habitualmente ya que corresponde al principio del gen, después de la región promotora.
  - Otros. Hay investigadores que toman el origen con respecto al origen de la región promotora del gen.
- Hay polimorfismos complejos: por ejemplo, en el gen APOE, cada uno de los polimorfismos E2, E3 y E4, tiene asociados dos SNP en regiones distintas.
- Hay genes complejos: por ejemplo, los genes PPAR-Alfa y PPAR-Gamma presentan una estructura en que las regiones promotoras y los start codon de ambos están intercalados en la misma región y están casi contiguos.

En cuarto lugar, se han descubierto deficiencias técnicas de diseño o de contenido en el acceso a la base de datos NCBI SNP. Las más importantes son las siguientes:

- Para cada gen hay gran número de SNP incluidos.

- No hay garantía de que todos los SNP correspondan a polimorfismos (errores de lectura, errores de transcripción entre documentos, mutaciones puntuales, etc.) (Figura 5.3).

**SNP's linked from LocusLink**

SNP's are linked from Locus [APOA1](#) via the following methods:

[Contig Annotation](#) [GenBank\(mrna\) Mapping](#)

Send  the list of rs# to Batch Query. Download  the list of rs# to file.

---

**Gene Model (mRNA alignment) information from genome sequence** ↑

Total gene model (contig mRNA transcript): **1**

| Contig                    | mrna                      | protein                   | mrna orientation | snp graph                                  |
|---------------------------|---------------------------|---------------------------|------------------|--|
| <a href="#">NT_033899</a> | <a href="#">NM_000039</a> | <a href="#">NP_000030</a> | reverse          | <a href="#">transcript on minus strand</a> |

view rs  in gene region  cSNP  has frequency  double hit  haplotype tagged

| gene model (contig mRNA transcript): | Contig                    | mrna                      | protein                   | mrna orientation | snp graph                  |
|--------------------------------------|---------------------------|---------------------------|---------------------------|------------------|----------------------------|
|                                      | <a href="#">NT_033899</a> | <a href="#">NM_000039</a> | <a href="#">NP_000030</a> | reverse          | transcript on minus strand |

| Contig position | dbSNP rs# cluster id)     | Heterozygosity | Validation | 3D  | OMIM | Function            | dbSNP allele | Protein residue | Codon position | Amino acid position |
|-----------------|---------------------------|----------------|------------|-----|------|---------------------|--------------|-----------------|----------------|---------------------|
| 20250256        | <a href="#">rs3178226</a> | N.D.           |            |     |      | untranslated region |              |                 |                |                     |
| 20250383        | <a href="#">rs5080</a>    | 0.095          |            | Yes |      | synonymous          | G            | Pro [P]         | 3              | 244                 |
|                 |                           | 0.095          |            | Yes |      | contig reference    | C            | Pro [P]         | 3              | 244                 |
| 20250445        | <a href="#">rs14081</a>   | N.D.           |            | Yes |      | synonymous          | T            | Leu [L]         | 1              | 224                 |

Figura 5.3. Interfaz de la base de datos NCBI SNP

- A veces no se encuentran los SNP correspondientes a los polimorfismos reportados. Hay que acudir a la NCBI Protein y ver la variación de aminoácidos que se asocia al polimorfismo. Se ha creado la tabla Gene\_Protein en la base de datos del proyecto para manejar esta información.
- Para cada SNP, el grado de Heterogocidad puede estar presente, con valores entre 0,000 y 0,999, o no (valor “N/A”). Algunas veces, el polimorfismo identificado presenta el valor “N/A” en la página del gen correspondiente.

En quinto lugar, se han detectado algunos límites de diseño en la base de datos NCBI SNP que no facilitan las búsquedas. Son los siguientes:

- Cada registro de la base de datos dbSNP corresponde a un gen, es decir, a la región genómica o conjunto de contigs que forman un gen.
- Por la razón anterior, resulta que el registro no incluye los SNP que estén fuera de la región del gen (por ejemplo, las regiones promotoras). Por ejemplo, el polimorfismo -75 A>G de APOA1 no está en el registro dbSNP correspondiente (Figura 5.3).
- No es posible encontrar un SNP si el identificador del polimorfismo se da por medio del encima de restricción (polimorfismos reportados en artículos antiguos).

### 5.3.3.2 Identificación de los polimorfismos en PubMed

En primer lugar, se ha producido un punto de inflexión o fecha umbral en la información genética contenida en las publicaciones científicas. Resulta que los artículos publicados antes de generalizarse los experimentos de genotipado y expresión génica describen los polimorfismos en la forma clásica de la Biología Molecular (enzima de restricción). Aproximadamente, se puede hablar de artículos de antes de 1998. Entonces, se puede decir que el “periodo genómico” en las referencias de resultados en la literatura publicada empieza más o menos en 1998.

En segundo lugar, existe una gran heterogeneidad en la información genética. En los artículos del “periodo genómico” no hay unidad de criterio en la descripción de los polimorfismos. La mayor parte no incluyen datos como el número de contig o, al menos, un valor relativo de offset respecto del inicio del gen. Esto hace bastante difícil la localización precisa de los polimorfismos en la dbSNP o la extracción de cadenas con el NCBI Viewer.

### 5.3.3.3 Desarrollo del proyecto

En el desarrollo del proyecto se habían formulado previamente algunas previsiones como las siguientes:

- Se había previsto un tiempo de tres meses para la realización del proyecto con la lista completa, o casi completa.
- Se había previsto la utilización de herramientas sencillas para automatizar el acceso a las BD genómicas.

En cambio, el estado actual, dadas las dificultades encontradas:

- Se ha reducido el proyecto a un proyecto piloto con 20 genes.
- Se ha avanzado en la obtención de SNP “fáciles”.
- Se está trabajando todavía en el “descubrimiento” de los polimorfismos “difíciles”.
- El término previsto ahora es en diciembre de 2004.

De todas formas, las previsiones temporales no son precisas puesto que hay toda una serie de variables (proyectos, oportunidad de investigación, disponibilidad de los investigadores, etc.) que están fuera de control del doctorando.

En resumen, en la ejecución de algunas de las actividades se han encontrado dificultades no previstas por causas científicas y técnicas. Por una parte, no hay un sistema homogéneo y universal para la identificación de polimorfismos. Por la otra parte, la utilización de las bases de datos genómicas públicas presenta algunas dificultades a causa de su diseño y de su contenido. Entonces, cuando se aborda el diseño de proyectos bioinformáticos se ha de prever las dificultades provocadas por el uso de recursos públicos.

### **5.3.4 Discusión sobre la propuesta de Auditoría Bioinformática**

La Bioinformática es una disciplina científico-tecnológica multidisciplinar donde concurren principalmente la Biología Molecular y Genética y la Ciencia de la Computación y la Ingeniería Informática. Dado la espectacular evolución que ha sufrido en poco tiempo y su cariz multidisciplinar, en la actualidad no existen enfoques y sistemas de evaluación e inspección global para dicha disciplina.

Como alternativa, se ha pensado en aplicar los planteamientos propios de la Auditoría de Sistemas de Información. Uno de los principales es el COBIT, un marco metodológico formal ampliamente aceptado por la comunidad de auditores, y que permite elaborar modelos de auditoría específicos para los contextos y alcances previamente establecidos. La hipótesis de partida ha sido que COBIT no aportaba la potencia y expresividad metodológica necesaria para realizar auditorías de tipo científico-tecnológico.

Por tanto, se ha presentado el diseño general de un Modelo de Auditoría, basado en COBIT, para el contexto de la Bioinformática aplicada a la investigación sobre individuos humanos, denominada Bioinformática Humana. Este modelo ha dado la oportunidad de descubrir las carencias de COBIT para el contexto mencionado. Aunque no se dispone de todos los datos, ya que el proyecto está en

curso, los resultados parciales indican que, para las funciones que no sean específicamente de TI, los procesos y objetivos de control tienen mucha utilidad.

Por lo tanto, se ha formulado una propuesta que significa una extensión al COBIT, denominado COBSIT (Control Objectives for Bioinformatics: Science, Information and related Technology - Objetivos de Control para la Bioinformática: Ciencia, Información y Tecnologías Afines), en el que incluyen dos nuevos dominios con sus procesos y objetivos de control asociados: el Dominio de Investigación y Difusión y el Dominio de Desarrollo y Formación Científicas. Esta propuesta no está completa ya que es necesario desarrollar todavía los Objetivos de Control Detallados y las Guías de Auditoría correspondientes. Y además, constituye una primera versión sujeta a crítica y revisión, tanto en el seno del proyecto en curso, como por la comunidad científica.

### **5.3.5 Discusión sobre aplicación de la auditoría bioinformática**

Una de las conclusiones generales del estudio efectuado como proyecto de auditoría es que se pueden mejorar varios de los aspectos en que se han detectado altos riesgos, a corto y a medio plazo. A continuación se detallan dichas mejoras.

#### **5.3.5.1 Mejoras en la organización**

Es necesario definir con claridad todas las funciones realizadas por el Laboratorio y agruparlas a continuación en cinco grandes categorías: científicas, técnicas, administrativas, docentes y dirección y supervisión general. Seguidamente, se deben asignar las responsabilidades sobre dichas categorías. El director debe asumir directamente las funciones de dirección y supervisión general, delegando algunas de ellas en el genetista, en previsión de las ausencias por motivos profesionales.

Las funciones docentes deben estar asumidas por el director, aunque debe delegar algunas en los técnicos de investigación, que son estudiantes de doctorado. Las funciones científicas deben estar asumidas por el genetista y el director, pero asignando las no estratégicas al genetista de forma que, si el director no está disponible, no se produzcan retrasos o paros de la actividad del Laboratorio por falta de decisiones científicas.

Las funciones técnicas y administrativas deben asignarse al genetista para que ejerza su gestión y supervisión. Sin embargo, la ejecución de las mismas necesita de la incorporación de nuevo personal: un bioinformático y un administrativo. El bioinformático (puesto en proceso de provisión) debe asumir el control,



supervisión y mantenimiento de los elementos que constituyen la Función Bioinformática. El administrativo debe llevar la administración general y económica del Laboratorio, descargando así al director y a los técnicos de laboratorio. El administrativo también puede hacerse cargo de las tareas de asistencia del director y del genetista: atención telefónica, gestión de la agenda, gestión de reservas para viajes, gestión de reservas de recursos del centro, filtro de visitantes puntuales, asistencia básica a los científicos visitantes, etc.

#### **5.3.5.1 Mejoras en las instalaciones y su gestión**

Con respecto a la arquitectura de sistemas informáticos, la mejor solución es la introducción de un servidor de red que controle directamente los equipos asociados al laboratorio y que actúe de firewall frente al acceso desde los equipos personales de los miembros del Laboratorio. Así, cualquier acceso a dichos equipos sería filtrado por el servidor que, por otra parte, daría servicio homogéneo y común a todos los investigadores. Este servidor podría tener la capa intermedia de servidor de aplicaciones y así centralizar las aplicaciones comunes de todo el Laboratorio. Además, dicho servidor podría servir como repositorio de la información común y particular de los miembros del Laboratorio y facilitar entonces la aplicación sistemática de un procedimiento de copias de seguridad y otro de mantenimiento remoto. La administración del servidor en una función asociada al bioinformático.

Los equipos deben ser revisados y sus discos particionados, para dejar una de las particiones dedicada a guardar solamente la información. Los usuarios deben tener una formación básica en el manejo de las estructuras de carpetas y ficheros y en el acceso a los recursos de red para que apliquen procedimientos de almacenamiento y salvaguarda estandarizados. Se debe revisar la configuración de red de cada equipo, para asignarles un único nombre de dominio. También se debería aplicar un procedimiento de asignación de nombres lógicos distinto del actual para facilitar el cambio de equipos y su identificación en el acceso remoto. Se debe revisar también el estado del sistema operativo y del hardware de cada equipo para optimizar su funcionamiento. Se debe acometer también la administración de cuentas, revisando las características de cada cuenta de usuario, diferenciando los nombres, y creando cuentas de usuario de tipo power user para el director y el genetista, y aplicando una política homogénea a las cuentas de administrador.

## 5.4 Bibliografía

### 5.4.1 Bibliografía sobre resultados de proyectos bioinformáticos

Applied Biosystems home page. <http://www.appliedbiosystems.com/>. Accedido el 4 de junio de 2004.

Applied Biosystems. ABI PRISM® 3100 Genetic Analyzer. User's Manual 2001. Applied Biosystems at <http://docs.appliedbiosystems.com/genindex.taf> (access data 10/31/2002).

Chromas home page. <http://www.technelysium.com.au/chromas.html>. Accedido el 4 de junio de 2004.

FASTA format Description web page. <http://ngfnblast.gbf.de/docs/fasta.html>. Accedido el 4 de junio de 2004.

GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.

GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.

GenBank-ftp. [ftp://ftp.ncbi.nih.gov/genomes/H\\_sapiens/](ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/). Accedido el 4 de junio de 2004.

OMG home page. <http://www.omg.org>. Accedido el 4 de junio de 2004.

PLASMID-1. <http://iubio.bio.indiana.edu/soft/molbio/mac/plasmid-maker.html>. Accedido el 4 de junio de 2004.

PLASMID-1. <http://www.arches.uga.edu/~lace52/rplasmid.html>. Accedido el 4 de junio de 2004.

RESTRICTION. <http://www.accessexcellence.org/AE/AEC/CC/restriction.html>. Accedido el 4 de junio de 2004.

### 5.4.2 Bibliografía CETP

Bernard S, Moulin P, Lagrost L, Picard S, Elchebly M, Ponsin G, et al. Association between plasma HDL-cholesterol concentration and TaqIB CETP gene polymorphism in non-insulin-dependent diabetes mellitus. *J Lipid Res* 1998; 39:59-65.

Freeman D, Packard CJ, Shepherd J, Gaffney D. Polymorphisms in the gene coding for cholesteryl ester transfer protein are related to plasma high density lipoprotein cholesterol and transfer ester activity. *Clin Sci* 1990; 79:575-581.

- Freeman DJ, Griffin BA, Holmes AP, Lindsay GM, Gaffney D, Packard CJ, et al. Regulation of plasma HDL cholesterol and subfraction distribution by genetic and environmental factors: Associations between the TaqI B RFLP in the CETP gene and smoking and obesity. *Arterioscler Thromb* 1994; 14:336-344.
- Fumeron F, Betoulle D, Luc G, Behague I, Ricard B, Poirier O, et al. Alcohol intake modulates the effect of a polymorphism of the cholesteryl ester transfer protein gene on plasma high density lipoprotein and the risk of myocardial infarction. *J Clin Invest* 1995; 96:1664-1671.
- Gudnason V, Kakko S, Nicaud V, Savolainen MJ, Kesaniemi YA, Thavanainen E, et al. Cholesteryl ester transfer protein gene effect on CETP activity and plasma high density lipoprotein in European populations. *Eur J Clin Invest* 1999; 29:116-128.
- Jiang XC, Agellon LB, Walsh A, Breslow JL, Tall A. Dietary cholesterol increases transcription of the human cholesteryl ester transfer protein gene in transgenic mice. Dependence on natural flanking sequences. *J Clin Invest* 1992; 90:1290-1295.
- Kauma H, Savolainen MJ, Heikkilä R, Rantala AO, Lilja M, Reunanen A, et al. Sex difference in the regulation of plasma high density lipoprotein cholesterol by genetic and environmental factors. *Hum Genet* 1996; 97:156-162.
- Kondo I, Berg K, Drayna DT, Lawn RM. DNA polymorphism at the locus for human cholesteryl ester transfer protein (CETP) is associated with high density lipoprotein cholesterol and apolipoprotein levels. *Clin Genet* 1989; 35:49-56.
- Kuivenhoven JA, de Knijff P, Boer JMA, Smalheer HA, Botma GJ, Seidell JC, et al. Heterogeneity at the CETP gene locus - Influence on plasma CETP concentrations and HDL cholesterol levels. *Arterioscler Thromb Vasc Biol* 1997; 17:560-568.
- Kuivenhoven JA, Jukema JW, Zwinderman AH, de Knijff P, McPherson R, Bruschke AV, et al. The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. The Regression Growth Evaluation Statin Study Group. *N Engl J Med* 1998; 338:86-93.
- Logan AG, Greenwood CM. Salt, blood pressure, and cointervention [letter; comment]. *Archives of Internal Medicine* 1997; 157:2666-2667.
- Mitchell RJ, Earl L, Williams J, Bisucci T, Gasiamis H. Polymorphisms of the gene coding for the cholesteryl ester transfer protein and plasma lipid levels in Italian and Greek migrants to Australia. *Human Biology* 1994; 66:13-25.
- Stevenson CG. Cholesterol ester transfer protein: a molecule with three faces? *Crit Rev Clin Lab Sci* 1998; 35:517-546.

### 5.4.3 Bibliografía APOE

- Aguilar CA, Talavera G, Ordovas JM, et al. The apolipoprotein E4 allele is not associated with an abnormal lipid profile in a Native American population following its traditional lifestyle. *Atherosclerosis* 1999;142:409-14.
- Ayaori M. Beneficial effects of alcohol withdrawal on LDL particle size distribution and oxidative susceptibility in subjects with alcohol-induced hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 1997;17:2540-7.
- Bercedo A, Gonzalez-Lamuno D, Muñoz O, et al. Association between lipid profile and Apo E genotype in Spanish children (8-15 years old). *Ann Esp Pediatr* 1998;42:120-4.
- Boerwinkle E, Utermann G. Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am J Hum Genet* 1988;42:104-12.
- Bottoni A, Cannella C, Del Balzo V. Lifestyle and dietary differences in smokers and non-smokers from an Italian employee population. *Public Health* 1997;111:161-4.
- Braeckman L, De Bacquer D, Rosseneu M, De Backer G. Apolipoprotein E polymorphism in middle-aged Belgian men: Phenotype distribution and relation to serum lipids and lipoproteins. *Atherosclerosis* 1996;120:67-73.
- Colditz GA, Giovannucci E, Rimm EB, et al. Alcohol intake in relation to diet and obesity in women and men. *Am J Clin Nutr* 1991;54:49-55.
- Davignon J, Gregg RE, Sing CF. Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 1988;8:1-21.
- Deiana L, Pes GM, Carru C, et al. Lack of influence of apolipoprotein E4 on lipoprotein levels in the island population of Sardinia. *Eur J Clin Invest* 1998;28:290-4.
- Dixon LB, Shannon BM, Tershakovec AM, Bennett MJ, Coates PM, Cortner JA. Effects of family history of heart disease, apolipoprotein E phenotype, and lipoprotein(a) on the response of children's plasma lipids to change in dietary lipids. *Am J Clin Nutr* 1997;66:1207-17.
- Eichner JE, Kuller LH, Ferrell RE, Meilahn EN, Kamboh MI. Phenotypic effects of apolipoprotein structural variation on lipid profiles.III. Contribution of apolipoprotein E phenotype to prediction of total cholesterol, apolipoprotein B, and low density lipoprotein cholesterol in the healthy women study. *Arteriosclerosis* 1990; 10: 379-85.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975;4:518-25.
- Frohlich JJ. Effects of alcohol on plasma lipoprotein metabolism. *Clin Chim Acta* 1996;246:39-49.

- Garry PJ, Baumgartner RN, Brodie SG, et al. Estrogen replacement therapy, serum lipids, and polymorphism of the apolipoprotein E gene. *Clin Chem* 1999;45:1214-23.
- Greenfield TK, Rogers JD. Who drinks most of the alcohol in the US? The policy implications. *J Stu Alcohol* 1999;60:78-89.
- Hegele RA, Evans AJ, Tu L, Ip G, Brunt JH, Connelly PW. A gene-gender interaction affecting plasma lipoproteins in a genetic isolate. *Arterioscler Thromb* 1994;14:671-8.
- Hendriks HFJ, Veenstra J, Van Tol A, Groener JE, Schaafsma G. Moderate doses of alcoholic beverages with dinner and postprandial high density lipoprotein composition. *Alcohol* 1998;33:403-10.
- Higuchi M, Iwaoka K, Ishii K, et al. Plasma lipid and lipoprotein profiles in pre- and post-menopausal middle-aged runners. *Clin Physiol* 1990;10:69-76.
- Hirano K, Yamashita S, Sakai N, et al. Low-density lipoproteins in hyperalphalipoproteinemic heavy alcohol drinkers have reduced affinity for the low-density lipoprotein receptor. *Clin Biochem* 1992;25:357-62.
- Jarvik GP, Austin MA, Fabsitz RR, et al. Genetic influences on age-related change in total cholesterol, low density lipoprotein-cholesterol, and triglyceride levels: Longitudinal apolipoprotein E genotype effects. *Genet Epidemiol* 1994;11:375-84.
- Jones DY, Judd JT, Taylor PR, Campbell WS, Nair PP. Menstrual cycle effect on plasma lipids. *Metabolism* 1988;37:1-2.
- Kamboh MI, Weiss KM, Ferrell RE. Genetic studies of human apolipoproteins. XVI. APOE polymorphism and cholesterol levels in the Mayans of the Yucatan Peninsula, Mexico. *Clin Genet* 1991;39:26-32.
- Kauma H, Savolainen MJ, Rantala AO, et al. Apolipoprotein E phenotype determines the effect of alcohol on blood pressure in middle-aged men. *Am J Hypertens* 1998;11:1334-43.
- Lefevre M, Ginsberg HN, Kris-Etherton PM, et al. ApoE genotype does not predict lipid response to changes in dietary saturated fatty acids in a heterogeneous normolipidemic population. The DELTA Research Group. Dietary Effects on Lipoproteins and Thrombogenic Activity. *Arterioscler Thromb Vasc Biol* 1997;17:2914-23.
- Lehtimäki T, Porkka K, Viikari J, Ehnholm C, Åkerblom HK, Nikkari T. Apolipoprotein E phenotypes and serum lipids in newborns and 3-year-old children: The cardiovascular risk in young Finns study. *Pediatrics* 1994;94:489-93.
- Liinamaa MJ, Hannuksela ML, Kesaniemi YA, Savolainen MJ. Altered transfer of cholesteryl esters and phospholipids in plasma from alcohol abusers. *Arterioscler Thromb Vasc Biol* 1997;17:2940-7.
- Lin RC, Miller BA, Kelly TJ. Concentrations of apolipoprotein AI, AII, and E in plasma and lipoprotein fractions of alcoholic patients: Gender differences in the effects of alcohol. *Hepatology* 1995;21:942-9.

- Mannisto S, Uusitalo K, Roos E, Fogelholm M, Pietinen P. Alcohol beverage drinking, diet and body mass index in a cross-sectional survey. *Eur J Clin Nutr* 1997;51:326-32.
- Nishiwaki M, Ishikawa T, Ito T, et al. Effects of alcohol on lipoprotein lipase, hepatic lipase, cholesteryl ester transfer protein, and lecithin:cholesterol acyltransferase in high-density lipoprotein cholesterol elevation. *Atherosclerosis* 1994;111:99-109.
- Okada T, Sato Y, Iwata F, Hara M, Kim H, Harada K. Relationship of apolipoprotein E phenotypes to serum lipid and lipoprotein levels in Japanese schoolchildren. *Acta Paediatr* 1998;87:460-1.
- Ordovas JM. The genetics of serum lipid responsiveness to dietary interventions. *Proc Nutr Soc* 1999;58:171-87.
- Pablos-Mendez A, Mayeux R, Ngai C, Shea S, Berglund L. Association of apo E polymorphism with plasma lipid levels in a multiethnic elderly population. *Arterioscler Thromb Vasc Biol* 1997;17:3534-41.
- Posner BM, Cupples LA, Miller DR, Cobb JL, Lutz KJ, D'Agostino RB. Diet, menopause, and serum cholesterol levels in women: The Framingham Study. *Am Heart J* 1993;125:483-89.
- Posner BM, Martin-Munley SS, Smigelski C, et al. Comparison of the techniques for estimating nutrient intake: The Framingham Study. *Epidemiology* 1992;3:171-7.
- Romelsjo A, Leifman H, Nystrom SA. A comparative study of two methods for the measurement of alcohol consumption in the general population. *Int J Epidemiol* 1995;24:929-36.
- Sandholzer C, Delport R, Vermaak H, Utermann G. High frequency of the apo epsilon 4 allele in Khoi San from South Africa. *Hum Genet* 1995;95:46-8.
- Shaper AG, Wannamethee SG. The J-shaped curve and changes in drinking habit. *Novartis Found Symp* 1998;216:173-88.
- Shaper AG. Alcohol and mortality: a review of prospective studies. *Br J Addict* 1990;85:837-47.
- Simon JA, Fong J, Bernert JT, Jr., Browner NS. Relation of smoking and alcohol consumption to serum fatty acids. *Am J Epidemiol* 1996;144:325-34.
- Van Tol A, Van der Gaag MS, Scheek LM, Van Gent T, Hendriks HFJ. Changes in postprandial lipoproteins of low and high density caused by moderate alcohol consumption with dinner. *Atherosclerosis* 1998;141:S101-S3.
- Wannamethee G, Shaper AG. Blood lipids: the relationship with alcohol intake, smoking, and body weight. *J Epidemiol Commun Health* 1992;46:197-202.
- Whiclow MJ, Prevost AT. Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. *Br J Nutr* 1996;76:17-30.

- Wilson PWF, Myers RH, Larson MG, Ordovas JM, Wolf PA, Schaefer EJ. Apolipoprotein E alleles, dyslipidemia, and coronary heart disease: The Framingham Offspring Study. *JAMA* 1994;272:1666-71.
- Zerba KE, Ferrell RE, Sing CF. Genotype-environment interaction: Apolipoprotein E (ApoE) gene effects and age as an index of time and spatial context in the human . *Genetics* 1996;143:463-78.

#### 5.4.4 Bibliografía APOA1

- Akita H, Chiba H, Tsuji M, Hui SP, Takahashi Y, Matsuno K, Kobayashi K. Evaluation of G-to-A substitution in the apolipoprotein A-I gene promoter as a determinant of high-density lipoprotein cholesterol level in subjects with and without cholesteryl ester transfer protein deficiency. *Hum Genet* 1995;96:521-526.
- Barre DE, Guerra R, Verstraete R, Wang Z, Grundy SM, Cohen JC. Genetic analysis of a polymorphism in the human apolipoprotein A-I gene promoter: Effect on plasma HDL-cholesterol levels. *J Lipid Res* 1994;35:1292-1296.
- Carmena-Ramon RF, Ordovas JM, Ascaso JF, Real J, Priego MA, Carmena R. Influence of genetic variation at the apoA-I gene locus on lipid levels and response to diet in familial hypercholesterolemia. *Athero* 1998;139:107-113.
- Civeira F, Pocovi M, Cenarro A, Garces C, Ordovas JM. Adenine for guanine substitution - 78 base pairs to the apolipoprotein(APO) A-I gene: relation with high density lipoprotein cholesterol and apoA-I concentrations. *Clin Genet* 1993;44:307-312.
- Danek GM, Valenti M, Baralle FE, Romano M. The A/G polymorphism in the 278 position of the apolipoprotein A-I promoter does not have a direct effect on transcriptional efficiency. *Biochim Biophys Acta* 1998;1398:67-74.
- Jeenah M, Kessling A, Miller N, Humphries SE. G to A substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI and high density lipoprotein cholesterol concentrations. *Mol Biol Med* 1990;7:233-241.
- Juo SHH, Wyszynski DF, Beaty TH, Huang HY, Bailey-Wilson JE. Mild association between the A/G polymorphism in the promoter of the apolipoprotein A-I gene and apolipoprotein A-I levels: A meta-analysis. *Am J Med Genet* 1999;82:235-241.
- Kamboh MI, Aston CE, Nestlerode CM, McAllister AE, Hamman RF. Haplotype analysis of two APOA1/MspI polymorphisms in relation to plasma levels of apo A-I and HDL-cholesterol. *Athero* 1996;127:255-262.
- Lopez-Miranda J, Ordovas JM, Espino A, Marin C, Salas J, Lopez-Segura F, Jimenez-Perez J, Perez-Jimenez F. Influence of mutation in human apolipoprotein A-I gene promoter on plasma LDL cholesterol response to dietary fat. *Lancet*, 1994;343:1246-1249.

- Mata P, Lopez-Miranda J, Pocovi M, Alonso R, Lahoz C, Marin C, Garces C, Cenarro A, Perez-Jimenez F, De Oya M, Ordovas JM. Human apolipoprotein A-I gene promoter mutation influences plasma low density lipoprotein cholesterol response to dietary fat saturation. *Athero* 1998;137:367-376.
- Matsunaga A, Sasaki J, Mori T, Moriyama K, Nishi K, Hidaka K, Arakawa K. Apolipoprotein A-I gene promoter polymorphism in patients with coronary heart disease and healthy controls. *Nutr Metab Cardiovasc Dis* 1995;5:275.
- Mattson FH, Grundy SM. Comparison of the effects of dietary saturated, monounsaturated, and polyunsaturated fatty acids on plasma lipids and lipoproteins in man. *J Lipid Res* 1985;26:194-202.
- Meng QH, Pajukanta P, Valsta L, Aro A, Pietinen P, Tikkanen MJ. Influence of apolipoprotein A-I promoter polymorphism on lipid levels and responses to dietary change in Finnish adults. *J Intern Med* 1997;241:373-378.
- Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins: A meta-analysis of 27 trials. *Arterioscler Thromb* 1992;12:911-919.
- Pagani F, Sidoli A, Giudici GA, Barengi L, Vergani C, Baralle FE. Human apolipoprotein A-I gene promoter polymorphism: Association with hyperalphalipoproteinemia. *J Lipid Res* 1990;31:1371-1377.
- Paul-Hayase H, Rosseneu M, Robinson D, Van Bervliet JP, Deslypere JP, Humphries SE. Polymorphisms in the apolipoprotein (apo) AI-CIII-AIV gene cluster: Detection of genetic variation determining plasma apo AI, apo CIII and apo AIV concentrations. *Hum Genet* 1992;88:439-446.
- Saha N, Tay JSH, Low PS, Humphries SE. Guanidine to adenine (G/A) substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI levels in Chinese non-smokers. *Genet Epidemiol* 1994;11:255-264.
- Sigurdsson G, Jr., Gudnason V, Sigurdsson G, Humphries SE. Interaction between a polymorphism of the Apo A-I promoter region and smoking determines plasma levels of HDL and Apo A-I. *Arterioscler Thromb* 1992;12:1017-1022.
- Smith JD, Brinton EA, Breslow JL. Polymorphism in the human apolipoprotein A-I gene promoter region. Association of the minor allele with decreased production rate in vivo and promoter activity in vitro. *J Clin Invest* 1992;89:1796-800.
- Talmud P, Ye S, Humphries SE. Polymorphism in the promoter region of the apolipoprotein AI gene associated with differences in apolipoprotein AI levels: the European Atherosclerosis Research Study. *Genet Epidemiol* 1994;11:265-280.
- Tuteja R, Tuteja N, Melo C, Casari G, Baralle FE. Transcription efficiency of human apolipoprotein A-I promoter varies with naturally occurring A to G transition. *FEBS Lett* 1992;304:98-101.



### 5.4.5 Bibliografía LIPC

- Berglund L, Oliver EH, Fontanez N, et al. HDL-subpopulation patterns in response to reduction in dietary total and saturated fat intakes in healthy subjects. *Am J Clin Nutr* 1999;70:992-1000.
- Campos H, Dreon DM, Krauss RM. Associations of hepatic and lipoprotein lipase activities with changes in dietary composition and low density lipoprotein subclasses. *J Lipid Res* 1995;36:462-472.
- Couture P, Otvos JD, Cupples LA, et al. Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles. The Framingham Offspring Study. *Arterioscler Thromb Vasc Biol* 2000;20:815-822.
- Deeb S, Peng R. The C-514T polymorphism in the human hepatic lipase gene promoter diminishes its activity. *J Lipid Res* 2000;41:155-158.
- Dreon DM, Fernstrom HA, Campos H, et al. Change in dietary saturated fat intake is correlated with change in mass of large-low-density-lipoprotein particles in men. *Am J Clin Nutr* 1998;67:828-836.
- Dugi KA, Amar MJA, Haudenschild CC, et al. In vivo evidence for both lipolytic and nonlipolytic function of hepatic lipase in the metabolism of HDL. *Arterioscler Thromb Vasc Biol* 2000;20:793-800.
- Grundy SM, Vega GL, Otvos JD, et al. Hepatic lipase activity influences high density lipoprotein subclass distribution in normotriglyceridemic men: genetic and pharmacological evidence. *J Lipid Res* 1999;40:229-234.
- Guerra R, Wang J, Grundy SM, et al. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *PNAS (USA)* 1997;94:4532-4537.
- Hegele RA, Harris SB, Brunt JH, et al. Absence of association between genetic variation in the LIPC gene promoter and plasma lipoproteins in three Canadian populations. *Atherosclerosis* 1999;146:153-160.
- Jansen H, Verhoeven AJ, Weeks L, et al. A common C-to-T substitution at position -480 of the hepatic lipase promoter associated with a lowered lipase activity in coronary artery disease participants. *Arterioscler Thromb Vasc Biol* 1997;17:2837-2842.
- Juo SH, Han Z, Smith JD, et al. Promoter polymorphisms of hepatic lipase gene influence HDL(2) but not HDL (3) in African American men: CARDIA study. *J Lipid Res* 2001;42:258-264.
- Kasim SE, Martino S, Kim P-N, et al. Dietary and anthropometric determinants of plasma lipoproteins during a long-term low-fat diet in healthy women. *Am J Clin Nutr* 1993;57:146-153.

- Katan MB, Grundy SM, Willett WC. Should a low-fat, high-carbohydrate diet be recommended for everyone? Beyond low-fat diets. *N Engl J Med* 1997;337:563-566.
- Kong C, Nimmo L, Elatrozy T, et al. Smoking is associated with increased hepatic lipase activity, insulin resistance, dyslipaemia and early atherosclerosis in Type 2 diabetes. *Atherosclerosis* 2001;156:373-378.
- Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins: A meta-analysis of 27 trials. *Arterioscler Thromb* 1992;12:911-919.
- Otvos JD, Jeyarajah EJ, Bennett DW, et al. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. *Clin Chem* 1992;38:1632-1638.
- Santamarina-Fojo S, Haudenschild CC, Amar M. The role of hepatic lipase in lipoprotein metabolism and atherosclerosis. *Curr Opin Lipidol* 1998;9:211-219.
- Shohet RV, Vega GL, Anwar A et al. Hepatic lipase (LIPC) promoter polymorphism in men with coronary artery disease. Allele frequency and effects on hepatic lipase activity and plasma HDL-C concentrations. *Arterioscler Thromb Vasc Biol* 1999;19:1975-1978.
- Tan K-C, Shiu SW, Chu BY. Effects of gender, hepatic lipase gene polymorphism and type 2 diabetes mellitus on hepatic lipase activity in Chinese. *Atherosclerosis* 2001;157:233-239.
- Thuren T. Hepatic lipase and HDL metabolism. *Curr Opin Lipidol* 2000;11:277-283.
- Tu A-Y, Albers JJ. Glucose regulates the transcription of human genes relevant in HDL metabolism: responsive elements for peroxisomes proliferator-activated receptor are involved in the regulation of phospholipid transfer protein. *Diabetes* 2001;50:1851-1856.
- Vega GL, Clark LT, Tang A, et al. Hepatic lipase activity is lower in African American than in white American men: effects of 5' flanking polymorphism in the hepatic lipase gene. *J Lipid Res* 1998;39, 228-232.
- Zambon A, Deeb S, Hokanson JE, et al. Common variants in the promoter of the hepatic lipase gene are associated with lower levels of hepatic lipase activity, buoyant LDL, and higher HDL2 cholesterol. *Arterioscler Thromb Vasc Biol* 1998;18:1723-1729.

#### 5.4.6 Bibliografía SR-BI

- Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler Thromb Vasc Biol* 1999; 19:1734-1743.

- Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, Krieger M. Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* 1996; 271:518-520.
- Arai T, Wang N, Bezouevski M, Welch C, Tall AR. Decreased atherosclerosis in heterozygous low density lipoprotein receptor-deficient mice expressing the scavenger receptor BI transgene. *J Biol Chem* 1999; 274:2366-2371.
- Braun A, Trigatti BL, Post MJ, Sato K, Simons M, Edelberg JM, Rosenberg RD, Schrenzel M, Krieger M. Loss of SR-BI expression leads to the early onset of occlusive atherosclerotic coronary artery disease, spontaneous myocardial infarctions, severe cardiac dysfunction, and premature death in apolipoprotein E-deficient mice. *Circ Res* 2002; 90:270-276.
- de Beer FC, Connell PM, Yu J, de Beer MC, Webb NR, van der Westhuyzen DR. HDL modification by secretory phospholipase A(2) promotes scavenger receptor class B type I interaction and accelerates HDL catabolism. *J Lipid Res* 2000; 41:1849-57.
- de Beer MC, Durbin DM, Cai L, Jonas A, de Beer FC, van der Westhuyzen DR. Apolipoprotein A-I conformation markedly influences HDL interaction with scavenger receptor BI. *J Lipid Res* 2001; 42:309-313.
- Durst R, Colombo R, Shpitzen S, Avi LB, Friedlander Y, Wexler R, Raal FJ, Marais DA, Defesche JC, Mandelshtam MY, Kotze MJ, Leitersdorf E, Meiner V. Recent origin and spread of a common Lithuanian mutation, G197del LDLR, causing familial hypercholesterolemia: positive selection is not always necessary to account for disease incidence among Ashkenazi Jews. *Am J Hum Genet* 2001; 68:1172-1188.
- Gillotte-Taylor K, Boullier A, Witztum JL, Steinberg D, Quehenberger O. Scavenger receptor class B type I as a receptor for oxidized low density lipoprotein. *J Lipid Res* 2001; 42:1474-1482.
- Goff DC Jr, D'Agostino RB Jr, Haffner SM, Saad MF, Wagenknecht LE. Lipoprotein concentrations and carotid atherosclerosis by diabetes status: results from the Insulin Resistance Atherosclerosis Study. *Diabetes Care* 2000; 23:1006-1011.
- Greene DJ, Skeggs JW, Morton RE. Elevated triglyceride content diminishes the capacity of high density lipoprotein to deliver cholesteryl esters via the scavenger receptor class B type I (SR-BI). *J Biol Chem* 2001; 276:4804-4811.
- Gu X, Trigatti B, Xu S, Acton S, Babitt J, Krieger M. The efficient cellular uptake of high density lipoprotein lipids via scavenger receptor class B type I requires not only receptor-mediated surface binding but also receptor-specific lipid transfer mediated by its extracellular domain. *J Biol Chem* 1998; 273:26338-26348.
- Huszar D, Varban ML, Rinninger F, Feeley R, Arai T, Fairchild-Huntress V, Donovan MJ, Tall AR. Increased LDL cholesterol and atherosclerosis in LDL receptor-deficient mice with attenuated expression of scavenger receptor BI. *Arterioscler Thromb Vasc Biol* 2000; 20:1068-1073.

- Iwashima Y, Eto M, Hata A, Kaku K, Horiuchi S, Ushikubi F, Sano H. Advanced glycation end products-induced gene expression of scavenger receptors in cultured human monocyte-derived macrophages. *Biochem Biophys Res Commun* 2000; 277:368-80.
- Kozarsky KF, Donahee MH, Glick JM, Krieger M, Rader DJ. Gene transfer and hepatic overexpression of the HDL receptor SR-BI reduces atherosclerosis in the cholesterol-fed LDL receptor-deficient mouse. *Arterioscler. Thromb. Vasc. Biol* 2000; 20:721-727.
- Kozarsky KF, Donahee MH, Rigotti A, Iqbal SN, Edelman ER, Krieger M. Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels. *Nature* 1997; 387:414-417.
- Krieger M. Charting the fate of the "good cholesterol": Identification and characterization of the high-density lipoprotein receptor SR-BI. *Annu Rev Biochem* 1999; 68: 523-558.
- Krieger M. Scavenger receptor class B type I is a multiligand HDL receptor that influences diverse physiologic systems. *J Clin Invest* 2001; 108:793-797.
- Milliat F, Gripois D, Blouquit M-E, Ferezou J, Serougne C, Fidge NH, Lutton C. Short and long-term effects of streptozotocin on dietary cholesterol absorption, plasma lipoproteins and liver lipoprotein receptors in RICO rats. *Exp Clin Endocrinol Diabetes* 2000; 108:436-46.
- Plat J, Mensink RP. Relationship of genetic variation in genes encoding apolipoprotein A-IV, scavenger receptor BI, HMG-CoA reductase, CETP and apolipoprotein E with cholesterol metabolism and the response to plant stanol ester consumption. *Eur J Clin Invest* 2002; 32:242-250.
- Rigotti A, Trigatti BL, Penman M, Rayburn H, Herz J, Krieger M. A targeted mutation in the murine gene encoding the high density lipoprotein (HDL) receptor scavenger receptor class B type I reveals its key role in HDL metabolism. *Proc Natl Acad Sci U S A* 1997; 94:12610-12615.
- Taskinen MR. Diabetic dyslipidemia. *Atherosclerosis S* 2002; 3:47-51.
- Trigatti BL, Rigotti A, Braun A. Cellular and physiological roles of SR-BI, a lipoprotein receptor which mediates selective lipid uptake. *Biochim Biophys Acta* 2000; 1529:276-86.
- Ueda Y, Gong E, Royer L, Cooper PN, Francone OL, Rubin EM. Relationship between expression levels and atherogenesis in scavenger receptor class B, type I transgenics. *J Biol Chem* 2000; 275:20368-20373.
- Ueda Y, Royer L, Gong E, Zhang J, Cooper PN, Francone O, Rubin EM. Lower plasma levels and accelerated clearance of high density lipoprotein (HDL) and non-HDL cholesterol in scavenger receptor class B type I transgenic mice. *J Biol Chem* 1999; 274:7165-7171.

- Varban ML, Rinninger F, Wang N, Fairchild-Huntress V, Dunmore JH, Fang Q, Gosselin ML, Dixon KL, Deeds JD, Acton SL, Tall AR, Huszar D. Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A* 1998; 95:4619-4624.
- Wang N, Arai T, Ji Y, Rinninger F, Tall AR. Liver-specific overexpression of scavenger receptor BI decreases levels of very low density lipoprotein ApoB, low density lipoprotein ApoB, and high density lipoprotein in transgenic mice. *Biol Chem* 1998; 273:32920-32926.

### 5.4.7 Bibliografía PLIN

- Arner P. Genetic variance and lipolysis regulation: implications for obesity 2001 *Ann Med* 2001; 33:542-546.
- Arner P. Insulin resistance in type 2 diabetes: role of fatty acids. *Diabetes Metab Res Rev* 2002; 18:S5-S9.
- Frayn KN, Karpe F, Fielding BA, Macdonald IA, Coppack SW. Integrative physiology of human adipose tissue. *Int J Obes Relat Metab Disord* 2003; 27:875-888.
- Gasteyger C, Tremblay A. Metabolic impact of body fat distribution. *J Endocrinol Invest* 2002; 25:876-883.
- Hellstrom L, Langin D, Reynisdottir S, Dauzats M, Arner P. Adipocyte lipolysis in normal weight subjects with obesity among first-degree relatives. *Diabetologia* 1996; 39:921-928.
- Hoffstedt J, Arner P, Schalling M, Pedersen NL, Sengul S, Ahlberg S, Iliadou A, Lavebratt C. A common hormone-sensitive lipase *lpl* gene polymorphism is associated with decreased human adipocyte lipolytic function. *Diabetes* 2001; 50:2410-2413.
- Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000; 26:163-175.
- Kolehmainen M, Vidal H, Ohisalo JJ, Pirinen E, Alhava E, Uusitupa MI. Hormone sensitive lipase expression and adipose tissue metabolism show gender difference in obese subjects after weight loss. *Int J Obes Relat Metab Disord* 2002; 26:6-16.
- Lavebratt C, Ryden M, Schalling M, Sengul S, Ahlberg S, Hoffstedt J. The hormone-sensitive lipase *lpl* gene polymorphism and body fat accumulation. *Eur J Clin Invest* 2002; 32:938-942.

- Lewis GF, Carpentier A, Adeli K, Giacca A. Disordered Fat Storage And Mobilization in the pathogenesis of insulin resistance and type 2 diabetes. *Endocr Rev* 2002; 23:201-229.
- Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, Hanson J, Khoury M, Lau J, O'Brien TR, Rothman N, Stroup D, Taioli E, Thomas D, Vainio H, Wacholder S, Weinberg C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002; 156:300-310.
- Lofgren P, Hoffstedt J, Ryden M, Thorne A, Holm C, Wahrenberg H, Arner P. Major gender differences in the lipolytic capacity of abdominal subcutaneous fat cells in obesity observed before and after long-term weight reduction. *J Clin Endocrinol Metab* 2002; 87:764-771.
- Lu X, Gruia-Gray J, Copeland NG, Gilbert DJ, Jenkins NA, Londos C, Kimmel AR. The murine perilipin gene: The lipid droplet-associated perilipins derive from tissue-specific, mRNA splice variants and define a gene family of ancient origin. *Mamm Genome* 2001; 12:741-749.
- Martinez-Botas J, Anderson JB, Tessier D, Lapillonne A, Chang BH, Quast MJ, Gorenstein D, Chen KH, Chan L. Absence of perilipin results in leanness and reverses obesity in *lepr(db/db)* mice. *Nat Genet* 2000; 26:474-479.
- Mottagui-Tabar S, Ryden M, Lofgren P, Faulds G, Hoffstedt J, Brookes AJ, Andersson I, Arner P. Evidence for an important role of perilipin in the regulation of human adipocyte lipolysis. *Diabetologia* 2003; 46:789-797.
- Nielsen S, Guo Z, Albu JB, Klein S, O'Brien PC, Jensen MD. Energy Expenditure, sex, and endogenous fuel availability in humans. *J Clin Invest* 2003; 111:981-988.
- Schiffelers SL, Akkermans JA, Saris WH, Blaak EE. Lipolytic and nutritive blood flow response to beta-adrenoceptor stimulation in situ in subcutaneous abdominal adipose tissue in obese men. *Int J Obes Relat Metab Disord* 2003; 27:227-231.
- Sztalryd C, Xu G, Dorward H, Tansey JT, Contreras JA, Kimmel AR, Londos C. Perilipin A is essential for the translocation of hormone-sensitive lipase during lipolytic activation. *J Cell Biol* 2002; 161:1093-1103.
- Tansey JT, Huml AM, Vogt R, Davis KE, Jones JM, Fraser KA, Brasaemle DL, Kimmel AR, Londos C. Functional studies on native and mutated forms of perilipins: A role in protein kinase A-mediated lipolysis of triacylglycerols in CHO cells. *J Biol Chem* 2003; 278:8401-8406.
- Tansey JT, Sztalryd C, Gruia-Gray J, Roush DL, Zee JV, Gavrilova O, Reitman ML, Deng CX, Li C, Kimmel AR, Londos C. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proc Natl Acad Sci USA* 2001; 98:6494-6499.
- Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992; 135:1019-1028.



En este capítulo se describen las principales aportaciones realizadas en esta tesis y, además, se recogen las conclusiones generales obtenidas al final de todo el proceso.

### 6.1 Aportaciones

En esta tesis, las aportaciones principales se pueden ver desde dos puntos de vista complementarios: el punto de vista de la Ciencia de la Computación y la Ingeniería Informática y el punto de vista de la Epidemiología Genómica de las Enfermedades Cardiovasculares.

Desde el punto de vista de la Ciencia de la Computación y la Ingeniería Informática, la aportación principal es la formulación de un marco de integración de las aproximaciones conceptuales de la Ingeniería del Software y la Auditoría de Sistemas de Información con la Bioinformática en el ámbito de la Epidemiología Genómica de las Enfermedades Cardiovasculares. En este sentido, las aportaciones concretas han sido las siguientes:

- Un modelo de proceso para desarrollar proyectos bioinformáticos basado en los principios y normas de la Ingeniería del Software.
- Un conjunto de soluciones bioinformáticas operativas para mejorar los sistemas LIMS (herramienta PLATEX); para facilitar la visualización y manejo de secuencias de nucleótidos (herramienta SeqPacker); para obtener variantes polimórficas en genes candidatos, asociados a la ocurrencia de ECV, y el diseño de los cebadores correspondientes para la amplificación mediante PCR (proyecto PREDICARD); y para compactar eficientemente ficheros de secuencias de nucleótidos con la disminución subsiguiente de los costes computacionales en memoria y en tiempo (herramientas Squeezer2B, Squeezer3B, DNA@Random y BioViewer).



- Un Modelo de Auditoría Bioinformática basado en el marco conceptual de Auditoría de Sistemas de Información COBIT con un conjunto de Guías e Auditoría para poder realizar auditorías específicas sobre las funciones bioinformáticas.

Desde el punto de vista de la Epidemiología Genómica de las Enfermedades Cardiovasculares, la aportación principal es la identificación y caracterización de un conjunto de factores genéticos y ambientales que permiten mejorar y ampliar la estimación del riesgo de padecer ECV. Con este enfoque, las aportaciones concretas han sido las siguientes:

- Un modelo conceptual para el cálculo del riesgo cardiovascular a escala individual, teniendo en cuenta factores de riesgo y de protección de tipo genético (polimorfismos de genes candidatos) y ambiental (dieta, ejercicio físico, consumo de alcohol y tabaco, etc.), así como la interacción multidimensional entre dichos factores.
- En el marco del Estudio Framingham, se ha obtenido la identificación de polimorfismos específicos de los genes candidatos (CETP, APOE, APOA1, LIPC, SR-BI y PLIN) y se han caracterizado las interacciones significativas con fenotipos patológicos: (1) marcada asociación del polimorfismo TaqIB en el gen CETP con las concentraciones de HDL-C; (2) interacción gen\*ambiente muy significativa entre el polimorfismo de la apoE y el consumo de alcohol; (3) interacción gen\*sexo entre el polimorfismo -75G/A en el promotor del gen APOA1 y la grasa poliinsaturada de la dieta; (4) interacción gen\*dieta entre el polimorfismo -514C/T en el promotor de gen LIPC y la grasa total de la dieta; (5) asociaciones significativas de tres variantes genéticas (exón 1, intrón 5 y exón 8) del gen SR-BI con las concentraciones de LDL-C, HDL-C y tamaños de partículas; (y (6) asociación de cuatro polimorfismos (PLIN1, PLIN4, PLIN5 y PLIN6) del gen PLIN a distintos fenotipos relacionados con la obesidad. En la sección de Conclusiones se amplían los detalles de estas aportaciones particulares.

La publicación en revistas y congresos internacionales de los resultados obtenidos, avalan la consistencia y calidad de las aportaciones realizadas en el curso de las investigaciones conducentes a esta tesis.

## 6.2 Conclusiones generales

Los resultados de esta tesis han permitido llegar a las siguientes conclusiones:

1. La Bioinformática, entendida como una disciplina científico-técnica, es indispensable como instrumento de integración en la investigación genómica cardiovascular en los distintos niveles de adquisición, tratamiento, análisis, almacenamiento y salvaguarda de datos. Sin embargo, debido a su reciente desarrollo, tropieza todavía con las dificultades derivadas de la ausencia de un cuerpo teórico común y consistente, que permita dar respuesta rápida a las grandes demandas de proceso de información ómica que se están detectando en la actualidad.
2. Las aportaciones bioinformáticas concretas desarrolladas en esta tesis como respuesta a las necesidades surgidas a lo largo de la investigación genómica cardiovascular en el Estudio Framingham, han dado resultados satisfactorios en su ámbito de aplicación. Así, PLATEX ha conseguido facilitar y disminuir el tiempo de proceso de los datos de las placas de genotipado en un orden de 300 veces sobre el proceso manual. SeqPacker ha servido para la visualización y gestión de secuencias de ADN y ARN, en diversos formatos, para su aplicación en el análisis genómico posterior. PREDICARD ha facilitado la búsqueda de variantes polimórficas en genes candidatos y el diseño de los cebadores correspondientes para la amplificación mediante PCR de la mejor zona de estudio. Squeezer ha conseguido unas tasas de compresión de secuencias de ADN y ARN que mejoran las de la mayoría de algoritmos conocidos, permitiendo así el intercambio vía Internet de grandes secuencias de ADN, incluso de cromosomas enteros, y acelerando los procedimientos de comparación exacta. Finalmente, la aplicación del Modelo de Auditoría Bioinformática ha permitido la realización de una auditoría práctica y la mejora de los procesos de investigación en el laboratorio genómico de referencia.
3. En cuanto a las conclusiones que se derivan de las investigaciones específicas de los genes analizados, como se puede afirmar que el polimorfismo TaqIB en el gen CETP, se asocia muy marcadamente con las concentraciones de HDL-C en hombres y en mujeres del Estudio Framingham, siendo los portadores del alelo B2 los que presentan las concentraciones más elevadas. Además, los hombres portadores del alelo B2 presentaron una reducción del 30% del riesgo de ECV, reducción que no se observó en las mujeres. En la investigación de las interacciones gen\*ambiente para este polimorfismo, no se encontraron factores ambientales que interaccionaran con sus efectos genéticos, resultando independiente del consumo de alcohol, tabaco y grasa de la dieta.

4. Con respecto a las interacciones gen\*ambiente con el polimorfismo de la apoE, se encontró una interacción muy significativa con el consumo de alcohol modulando las concentraciones de LDL-C en hombres, pero no en mujeres, poniendo de manifiesto una interacción gen\*sexo.
5. Otra importante interacción gen\*sexo ha sido encontrada entre el polimorfismo -75G/A en el promotor del gen APOA1 y la grasa poliinsaturada de la dieta, determinando las concentraciones de HDL-C en mujeres, pero no en hombres. Así, un consumo alto de ácidos grasos poliinsaturados en la dieta disminuye el HDL-C en mujeres homocigotas GG, pero ejerce un efecto contrario en las portadoras del alelo A.
6. Otra interacción gen\*dieta presentando una gran consistencia y efectos dosis-dependientes ha sido caracterizada, tanto en hombres como en mujeres de la cohorte de Framingham. Esta interacción implica al polimorfismo -514C/T en el promotor de gen LIPC y la grasa total de la dieta, determinando, tanto las concentraciones de HDL-C como los diámetros de las partículas.
7. Al realizar un análisis de haplotipos incluyendo simultáneamente tres variantes genéticas (exón 1, intrón 5 y exón 8) del gen SR-BI, se han hallado asociaciones significativas entre estos haplotipos y las concentraciones de LDL-C, HDL-C y tamaños de partículas, que a su vez podían ser posteriormente moduladas por el fenotipo de diabetes tipo 2 que presentaran los participantes.
8. El análisis de las variaciones genéticas en el gen PLIN en población general de la Comunidad Valenciana, ha permitido identificar cuatro polimorfismos (PLIN1, PLIN4, PLIN5 y PLIN6), algunos de ellos en elevado desequilibrio de ligamiento, que se han asociado a distintos fenotipos relacionados con la obesidad. El resultado más consistente ha sido el menor riesgo de obesidad encontrado en los portadores del alelo 2 en PLIN4 en comparación con los homocigotos normales.
9. Estas conclusiones específicas derivadas del estudio de cada uno de los genes y del los comportamientos ambientales objeto de esta tesis, ofrecen datos concretos de asociaciones reales observadas entre variaciones genéticas y fenotipos cardiovasculares, así como de interacciones gen\*ambiente. Teniendo en cuenta la diversidad de genes (mas de 3.000) y de variantes genéticas funcionales en los mismos, que pueden estar implicadas en la expresión de fenotipos cardiovasculares, junto con su interacción con el ambiente, es necesaria la participación masiva de herramientas computacionales para integrar en la práctica clínica la información que se deriva de las investigaciones específicas.

10. Por todo ello, la Bioinformática y el conjunto de disciplinas surgidas al amparo de la misma, van a experimentar un espectacular desarrollo en un futuro muy cercano.



### 7.1 Limitaciones de los resultados

La Genómica Cardiovascular estudia el problema de las enfermedades cardiovasculares desde el punto de vista de las interacciones gen\*gen y gen\*ambiente que se asocian con manifestaciones disfuncionales o patologías en el sistema cardiovascular y lipídico de los animales mamíferos superiores. Esta tesis se ha adscrito solamente a la parte correspondiente al Homo Sapiens.

Los descubrimientos y resultados vertidos en esta tesis se enmarcan en un conjunto de estrategias científicas donde confluyen varias disciplinas y estrategias con el objetivo común de investigar las ECV en humanos y encontrar las causas genéticas específicas y genéticas combinadas con factores ambientales. Sin embargo, dichas estrategias se hallan en constante evolución, puesto que no ha sido suficiente con haber completado el genoma humano, sino que es necesario encontrar las relaciones entre las regiones genómicas (fundamentalmente genes) y todas las funciones fisiológicas, morfológicas y metabólicas del cuerpo humano (esto lo estudia la Genómica Funcional) en general, y en el campo de la Medicina Cardiovascular en particular. También hay que seguir trabajando en la identificación de nuevas proteínas (se cree que hay mas de un millón de proteínas sin descubrir) que son los componentes estructurales y funcionales de cualquier ser vivo. Si además, se incorporan las investigaciones sobre los polimorfismos de los genes y su asociación con anomalías o patologías cardiovasculares, la situación se complica hasta extremos inimaginables. Se calcula que existen casi 3.000 genes, de los 30.000 que organizan las 300.000 Mbases del genoma humano, relacionados directamente con las patologías cardiovasculares

En este contexto, hay que entender que los resultados sobre los genes estudiados no son en absoluto definitivos y, como se expresa en las discusiones correspondientes del Capítulo 5, se necesitan posteriores estudios de otros tipos, como estudios experimentales en humanos (dentro de los límites éticos) y en

modelos animales, estudios de intervención y estudios a bajo nivel (moleculares, bioquímico-estructurales, etc.) que vayan corroborando dichos resultados.

Otras limitaciones que tienen los resultados se relacionan con los valores de prevalencia de las ECV y los estudios epidemiológicos que han servido dichos resultados. Las ECV se manifiestan, salvo patologías estructurales muy específicas, como deformaciones congénitas en elementos del aparato cardiovascular, etc.) normalmente en la madurez de los individuos. Y además, son débilmente hereditarias (es decir, puede haber portadores y enfermos en función de la combinación alélica de los ascendientes). Por ello aportan mayor evidencia los estudios epidemiológicos, denominados genéricamente de cohorte, que toman una muestra representativa de la población y la siguen estudiando a lo largo de su vida e, incluso, van incorporando a sus descendientes. El Estudio Framingham es uno de este tipo. Por lo tanto, los resultados de esta tesis deben de ir comparándose con los que se vayan obteniendo en los sucesivos estudios sobre las generaciones de la cohorte.

Una limitación adicional, que se puede considerar como principio metodológico en la Epidemiología Genómica, es que los resultados de esta tesis no son aplicables directamente a otras poblaciones que no sean la estudiada, debido a la gran variabilidad, sobre todo, de los factores ambientales de otras poblaciones, aunque también de las características genéticas próximas de los individuos localizados en una misma región geográfica. Por ello en esta tesis se ha incorporado el estudio de un gen en otra población distinta, la población general de la Comunidad Valenciana, para comprobar si se pueden trasladar las hipótesis y protocolos y métodos de trabajo del Estudio Framingham a este último.

Los resultados de esta tesis también se supeditan a las limitaciones de los instrumentos de medida aplicados en los estudios realizados que, sobre todo en la medición de factores ambientales como estilo de vida, dieta, ejercicio, tabaco y alcohol, no son todo lo precisos que sería deseable (es decir, no son mediciones con la precisión que se aplica, por ejemplo, en la Física de Altas Energías). Aunque dichas limitaciones ya se consideran a la hora de la realización de los análisis estadísticos.

Y finalmente, existen unas limitaciones éticas, legales y organizativas que impiden el acceso a la información completa de cada uno de los individuos participantes del estudio. Esto hace, la mayoría de las veces, más complicado el proceso de verificación de resultados.

En cuanto a los resultados en forma de soluciones bioinformáticas, una de las grandes limitaciones está en el manejo de la complejidad computacional con

respecto al tratamiento de las secuencias de ADN y los mecanismos de búsqueda y alineamiento. La singular heterogeneidad y deficiente diseño de interfaz de las bases de datos genómicas públicas (sobre todo para accesos directos desde aplicaciones locales) limita la potencia de las soluciones aportadas. Por ello, una de las conclusiones expresadas en la sección correspondiente del Capítulo 5 indica que es preciso mejorar e intentar homogenizar, tanto el diseño de dichas bases como sus interfaces. Una de las vías posibles de esta mejora está en la incorporación de tecnología XML.

Otra limitación viene impuesta por la precisión del instrumental utilizado para el genotipado. Si bien las soluciones bioinformáticas pueden manipular secuencias de nucleótidos sin errores, al menos, los intrínsecos a dicho código, no ocurre lo mismo con los instrumentos que analizan muestras biológicas, es decir, hebras o fragmentos de hebras, y producen una salida en forma de secuencia. Este problema se traslada a las bases de datos que no comparan y verifican las secuencias que van aportando los investigadores, como es el caso de GenBank, frente a las que sí lo hacen, como SwissProt. Las soluciones bioinformáticas aportadas deben manejar con precaución las secuencias obtenidas de, por ejemplo, GenBank, puesto que no hay garantía absoluta que correspondan exactamente a las secuencias totalmente naturales de ADN o ARN.

En cuanto a la metodología de auditoría bioinformática aportada como resultado en forma de un modelo de auditoría y unas guías de auditoría para funciones bioinformáticas, las limitaciones provienen de la propia naturaleza fuertemente evolutiva de esta disciplina y las crecientes interrelaciones con otras disciplinas englobadas en el concepto de “ómicas”. El modelo propuesto es válido para la Bioinformática tal como estaba cuando se aplicó. Sin embargo, no se puede afirmar que sea igualmente válida a corto plazo (un semestre, o un año) y, por tanto, será necesaria una revisión en función de la situación científico-tecnológica de ese momento.

## 7.2 Líneas de investigación abiertas

Esta tesis se puede ver también, desde el punto de vista de la trayectoria de investigación, como una de las primeras entregas de informe de resultados que se dan en un proyecto, donde se plasman los resultados parciales conseguidos hasta la fecha, pero que no constituyen los resultados definitivos, puesto que el proyecto continua. En este sentido, y dado que esta tesis tiene un cariz multidisciplinar, han quedado no una, sino varias líneas de investigación abiertas que pueden ser explotadas (y de hecho lo están siendo en la actualidad) una vez concluido el proceso académico de la defensa de esta tesis.



En cuanto al Estudio Framingham y al cálculo del riesgo cardiovascular, quedan todavía muchos genes y polimorfismos por estudiar para descubrir nuevas interacciones gen\*gen y gen\*ambiente. Si bien el ritmo de la investigación estaba limitada, hasta no hace mucho tiempo, por la limitación de capacidad de proceso del instrumental del laboratorio genómico y la propia capacidad del personal investigador, con la incorporación desde 2002 de un instrumento de análisis de expresión genética mediante micromatrices de ADN o de SNP, los estudios de expresión se ha acelerado y han bajado los costes inherentes a reactivos y otros recursos específicos necesarios. Con la tecnología de Affymetrix se pueden genotipar o expresar (que son dos enfoques genómicos distintos) del orden de 15.000 a 30.000 genes por individuo y placa de micromatriz. O con la tecnología de Illumina se pueden genotipar o expresar del orden de 2.000 a 5.000 individuos por gen y placa de micromatriz. Entonces, el problema a abordar no es precisamente el de mejorar la tecnología de micromatrices, sino el de mejorar los instrumentos de análisis de la ingente cantidad de información que produce un experimento con micromatrices de ADN. Esta es una de las vías de trabajo pendientes de abordar por el doctorando.

Por otra parte, y dado que las tecnologías mencionadas cada día que pasa van reduciendo sensiblemente su coste, ahora ya se pueden realizar experimentos con micromatrices de ADN en la propia Universidad de València. Y así, otra de las líneas abiertas es la continuación de los estudios de genes relacionados con las ECV en población, general o no, de la Comunidad Valenciana. Esto se supedita a la colaboración que tiene el doctorando con el grupo EPIGEM del Departamento de Medicina Preventiva de la Universidad de València a través de varios proyectos de investigación. Se trata igualmente de aplicar soluciones bioinformáticas y nuevas estrategias de análisis de los datos de expresión genética producidos en los estudios epidemiológicos genómicos con el apoyo de tecnología de micromatrices.

Otra línea de trabajo abierta, también relacionada con los resultados de esta tesis, y en el ámbito de la Red INBIOMED, consiste en la mejora de las soluciones bioinformáticas existentes y en el desarrollo de otras para necesidades específicas en la investigación epidemiológica genómica. Ejemplos de las últimas pueden ser el cálculo y asignación de haplotipos, la propuesta de formatos formales para la estructuración de ficheros de secuencias, y la construcción de representaciones del conocimiento en Epidemiología Genómica Cardiovascular por medio de ontologías.

Con respecto a la propuesta de Auditoría Bioinformática, se trata de aplicar el modelo y las guías en tanto casos como sea posible y efectuar las revisiones metodológicas y de las guías para mejorar el modelo. También se deben incorporar

las actualizaciones del propio marco conceptual que es el COBIT que lleva a cabo la ISACA.

Estas líneas de investigación descritas se refieren a la trayectoria del doctorando a corto plazo. No obstante, esta tesis ha dado lugar a la formulación de nuevos proyectos cuya acometida está prevista a medio plazo. En la sección siguiente se describen brevemente dichos proyectos.

### 7.3 Trabajo futuro

El trabajo futuro tiene un amplio espectro de posibilidades debido a la naturaleza multidisciplinar de las investigaciones realizadas en esta tesis. Sin embargo, las que se pueden concretar a medio plazo es un número más reducido y se pasan a relatar.

En primer lugar, los conocimientos adquiridos por el doctorando a lo largo de sus investigaciones le pueden facultar, cuando alcance el grado de doctor, para poder organizar e impartir docencia en Bioinformática. Esto se traduce en dos proyectos particulares. El primero es la organización y desarrollo de docencia en Bioinformática en el seno de la Red INBIOMED y adscrito al Comité de Formación creado para tal fin. El segundo es la organización y desarrollo de un Curso de Postgrado de Bioinformática aplicada a la Epidemiología Genómica, Agrotecnología y Medio Ambiente, con la colaboración de la Fundación Universidad-Empresa de la Universidad Jaume I (FUE-UJI) cuyo inicio está previsto en noviembre de 2004. El tercero es la participación como docente (y antes asesor de programa y contenidos) en el Master de Ingeniería Biomédica que gestionarán la Universitat de València y la Universidad Politécnica de Valencia. El cuarto es la posibilidad de participar en de una iniciativa para la creación de un programa oficial y nacional de doctorado en Bioinformática, impulsado desde la Red INBIOMED.

En segundo lugar, se plantea el desarrollo de dos proyectos relacionados con la Epidemiología Genómica Cardiovascular en relación con el Estudio Framingham y en la Comunidad Valenciana. El primero es la creación de un portal Web sobre Epidemiología Genómica Cardiovascular que incorpore la tecnología de servicios Web y soporte la gestión de conocimientos, metodologías, procesos, publicaciones y otras informaciones en el ámbito exclusivo de esta disciplina, y con un diseño más formal, riguroso y normalizado que las bases de datos genéricas existentes. El segundo consiste en la participación en el diseño de una micromatriz específica para analizar la expresión genética en el campo de las ECV y en el diseño de nuevas estrategias para analizar la información producida y poder sintetizarla en recomendaciones médicas para la prevención. En tercer lugar, se

plantea un proyecto para aplicar enfoques relativamente avanzados de la Ingeniería del Software y la Inteligencia Artificial en el cálculo del riesgo cardiovascular a escala individual partiendo del perfil genómico y ambiental específico de cada individuo. Parte ya está desarrollado mediante el enfoque de los agentes software.

Como se puede observar, las posibilidades son inmensas. Pero el tiempo y los recursos son limitados ya que muchos de estos proyectos entran en competencia con propuestas de otras fuentes. En resumen, se tratará de revisar y mejorar los planteamientos y diseños para que tengan las debidas garantías de alcanzar el éxito deseado.

### **8.1 Introducción**

En este Capítulo se han incluido todas las referencias manejadas a lo largo de esta tesis. Siguiendo con el espíritu de querer facilitar el acceso a las mismas, se ha mantenido la agrupación de referencias por artículos publicados en las respectivas secciones. Además, se han creado nuevas secciones con el propósito de agrupar las referencias básicas, no relacionadas con ningún artículo, y las referencias técnicas relacionadas con las distintas tecnologías que se describen en la tesis.

### **8.2 Artículos publicados**

Coltell O, Arregui M., Chalmeta R., Corella D., Parnell L., Ordovás JM. “SeqPacker: a nucleotide sequence manipulation utility”. Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE2004) May 19-21, 2004, Taichung, Taiwan, ROC (<http://bibe2004.ece.uci.edu/>).

Coltell O, Arregui M., Chalmeta R., Corella D., Parnell L., Ordovás JM. “A Biologist-Friendly User Interface to Manipulate Nucleotide Sequences in Genomic Epidemiology”. 2nd International Workshop on Biological Data Management (BIDM’04). 1 - 3 September, 2004. Zaragoza, Spain.

Coltell O. “Auditoría Bioinformática”. I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 15-24. ISSN: 1579-8070. Editorial MIC.

Coltell O. “El marco de auditoría básico para las funciones bioinformáticas”. I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 25-28. ISSN: 1579-8070. Editorial MIC.

Coltell O. “La disciplina de Bioinformática: definición y caracterización”. I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 29-34. ISSN: 1579-8070. Editorial MIC.

- Coltell O. "La representación, almacenamiento e interrelación de entidades genómicas y proteómicas en las bases de datos públicas: un caso de estudio". Curso de Informática Biomédica. Instituto de Salud Carlos III (ISCIII), Madrid, 13 de octubre de 2003.
- Coltell O. "Sistemas de información en laboratorios de investigación en Biomedicina". I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud; 2003, 43: 9-14. ISSN: 1579-8070. Editorial MIC.
- Coltell O., Corella D., Sánchez JT., Chalmeta R., Ordovas JM. "Modelo Conceptual Bioinformático Aplicado al Análisis Genómico en las Enfermedades Cardiovasculares". Clínica e Investigación en Arteriosclerosis; 2004, 16(2): 43-52.
- Coltell O., Corella D., Tai E-S., Guillén M., Chalmeta R., Ordovas JM. "PLATEX: Una herramienta bioinformática para la conversión de datos en el estudio genético de la arteriosclerosis". Clínica e Investigación en Arteriosclerosis; 2004, 16(2): 53-60.
- Corella D, Tucker K, Lahoz C, Coltell O, Cupples LA, Wilson PWF, Schaefer EJ, Ordovas JM. "Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study". American Journal of Clinical Nutrition; 2001; 73(4):736-45. ISSN "0002-9165 (SCI: 2000, 5,012 p., 3/52 "NUTRITION AND DIETETICS").
- Doreen Osgood; Dolores Corella; Serkalem Demissie; L Adrienne Cupples; Peter WF Wilson; James B Meigs; Ernst J Schaefer; Oscar Coltell; Jose M Ordovas. "Genetic variation at the scavenger receptor class B type I (SR-BI) gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: The Framingham Study". J Clin Endocrinol Metab. 2003 Jun;88(6):2869-79. ISSN 0021-972X (SCI 2002, 5.199 p.; 13/88 "ENDOCRINOLOGY & METABOLISM").
- Lu Qi; Dolores Corella, José V. Sorlí, Olga Portolés, Haiqing Shen, Oscar Coltell, Diego Godoy, Andrew S. Greenberg, Jose M. Ordovas. "Genetic variation at the Perilipin (PLIN) locus is associated with obesity-related phenotypes in White women". Clinical Genetics 2004 Ref. CGE-00136-2004. ISSN 0009-9163 Online ISSN: 1339-0004 (SCI 2002, 2.237 p.; 57/115 "GENETICS & HEREDITY").
- Ordovás JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Polyunsaturated fatty acids modulate the effect of the APOA1-75(G/A) polymorphism on HDL-C levels in a specific fashion: The Framingham Study". American Journal of Clinical Nutrition, 2002; 75(1):38-46. ISSN "0002-9165 (SCI: 2000, 5,012 p., 3/52 "NUTRITION AND DIETETICS").
- Ordovás JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on HDL metabolism: Evidence of a strong dose-effect in this gene-nutrient interaction in the Framingham Study". Circulation, 2002; 18: 2315-2321. ISSN 0009-7322 (SCI: (2000, 10,893 p.), (2001, 10,517 p.); 1/63 "CARDIAC & CARDIOVASCULAR SYSTEMS"; 1/60 "HEMATOLOGY"; 1/45 "PERIPHERAL VASCULAR DISEASE").

Ordovas JM, Cupples A, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PWF, Schaefer EJ. "Association of CETP TaqIB polymorphism with variation in lipoprotein subclasses and coronary heart disease risk: The Framingham Study". *Arteriosclerosis Thrombosis and Vascular Biology*; 2000; 20: 1323-1329. ISSN 1079-5642 (SCI: 1999, 5,406 p., 6/45 "PERIPHERIAL VASCULAR DISEASE", 5/60 "HEMATOLOGY").

### 8.3 Bibliografía sobre Bioinformática y disciplinas asociadas

- Andrieu N., Goldstein AM. "Epidemiologic and genetic Approaches in the Study of Gene-environment Interaction: an Overview of Available Methods". *Epidemiol Rev*, 1998; 20, 2: 137-147.
- Backofen R., Gilbert D. "Bioinformatics and Constraints". *Constraints*, 6; 2001: 141- 156.
- Collins F. S., Green E. D., Guttmacher A. E., Guyer M. S. "A vision for the future of genomics research A blueprint for the genomic era". *Nature* 2003, 422: 835-847.
- Collins F. S., Morgan M., Patrinos A. "The Human Genome Project: Lessons from Large-Scale Biology". *Science*, 2003, 300: 286-290.
- Coltell O., Corella D. "Bioinformática y Epidemiología Genómica". *I+S. Informática y Salud. Revista de la Sociedad Española de Informática y Salud*; 2002, 37: 15-26.
- Cook S. A., Rosenzweig A. "DNA Microarrays. Implications for Cardiovascular Medicine". *Circ Res*. 2002;91:559-564.
- INBIOMED. Red Temática de Investigación Cooperativa de Informática Biomédica. <http://www.inbiomed.retics.net>.
- Jackson D.G., Healy M.D., Davison D.B. Bioinformatics: not just for sequences anymore. *BIOSILICO* 2003; I(3): 103-111.
- Jones R.L. "The Internet and Healthcare Information Systems: How Safe Will Patient Data Be?". *IS Audit & Control Journal*, I; 1998: 25-30.
- Khoury M. "Genetic Epidemiology". Rothman K.J., Greenland S. (eds.) *Modern Epidemiology* (2nd ed.) Little, Brown and Co., Boston, MA (USA), 1997.
- Khoury M., Beaty T.H., Cohen B.H. *Fundamentals of Genetic Epidemiology*. Oxford University Press, New York, 1993.
- Marrugat J., Solanas P., D'Agostino R., Sullivan L., Ordovás J., et al. "Coronary risk estimation in Spain using a calibrated Framingham function". *Rev Esp Cardiol* 2003; 56: 253-261.
- Martín Sánchez F., López V., Sánchez J. P., Liébana I. Red Temática de Investigación Cooperativa en Informática Biomédica, INBIOMED. *Revista I+S. Informática y Salud*, 2004, 46: 7-13.

- Martín-Sánchez F. et al. Facilitating Genomic Medicine for Future Healthcare. *J. Biomed. Inform.* Artículo en prensa, doi:10.1016/j.jbi.2003.09.003 .
- McKusick V.A. "The Human Genome Organisation: History, Purposes and Membership". *Genomics*, 5; 1989: 385-387.
- Ordovás J. M. et al. "Identification and molecular analysis of two apob gene mutations causing low plasma cholesterol levels". *Circulation*, 92, 1995: 2036-2040.
- Ordovás JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PWF, Schaefer EJ, Tucker K. "Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on HDL metabolism: Evidence of a strong dose-effect in this gene-nutrient interaction in the Framingham Study". *Circulation*, 2002; 18: 2315-2321.
- Roberts R. "Bioinformatics Analysis of Gene Banks Provides a Treasure Trove for the Functional Genomist". *J. Mol. Cell. Cardiol.*; 32; 2000: 1917-1919.
- Sackman H. *Biomedical Information Technology. Global Social Responsibilities for the Democratic Age.* Academic Press, San Diego, CA (USA), 1997.
- Science's News staff. "A History of the Human Genome Project". *Science Magazine*; 291 (5507); 2001: 1195-1261.
- Van Bemmelen J.H., Musen M.A.(eds.) *Handbook of Medical Informatics.* Springer-Verlag, Heidelberg, 1997.
- Wilson PWF., D'Agostino RB., Levy D., Belanger A., Silbershatz H., Kannel W. "Prediction of Coronary Heart Disease Using Risk Factor Categories". *Circulation*, 1998; 97,18: 1837-1847.

## 8.4 Bibliografía sobre técnicas de determinación

- Kleinbaum, D.G., Kupper, L.L. and Muller, K.E. (1988) *Applied regression analysis and other multivariable methods.* 2ª ed. Boston: PWS Kente Publishing company.
- Rothman, K. (1986). *Modern Epidemiology*, eds. Little Brown and Company. Boston.
- Tilzer, L.L., Thomas, S.M. and Moreno, R.F. (1989). Use of silica gel polymer for DNA extraction with organic solvents. *Anal Biochem* 183:13-15.
- Wayne, S.D. (1990). *Bioestadística.* 3ª ed. Mexico D.F. Limusa.

## 8.5 Bibliografía sobre técnicas de desarrollo de software

- ANSI/IEEE Std. 830-1993. «Recommended Practice for Software Requirements», IEEE, 1993.
- Boehm B. Software Engineering Economics, Prentice-Hall, 1981.
- Booch G., Rumbaugh J., Jacobson I. El Lenguaje Unificado de Modelado, Addison-Wesley, Madrid, 1999.
- Bruegge B., Dutoit A.H. Ingeniería de Software Orientado a Objetos. Prentice Hall – Pearson educación, México, 2002.
- Budd T. Introducción a la programación orientada a objetos, Addison-Wesley Iberoamericana, Wilmington, DW (USA), 1994.
- Bourque, Pierre; Dupuis, Robert; Abran, Alain; Moore, James W.; Tripp, Leonard; y Wolf, Sybille. Fundamental principles of software engineering – a journey. The Journal of Systems and Software 2002; 62, 59-70.
- Castagna G. Object-Oriented Programming. A Unified Foundation. Birkhäuser, Boston, MA (USA), 1996.
- Cleland D.I. *Project Management. Strategic Desing and Implementation*, TAB Professional and Reference Books, Blue Ridge Summit, PA (USA), 1990.
- Coltell O. “Ingeniería del Software. Introducción. Ciclos de vida, herramientas CASE. Generación automática de código. Reusabilidad de componentes”. García F. J., del Valle F. J. (Eds.) Temario para Oposiciones a Analista de Sistemas de Información. Volumen I. Colegio Oficial de Ingenieros en Informática de la Comunidad Valenciana, Valencia, 2003: 657-707. ISBN 84-607-9851-8, 84-607-9852-6 (Vol. I).
- Harmon P., Hall C. Intelligent Software Systems Development. An IS Managers Guide. John Wiley, New York-USA, 1993.
- ISO home page. <http://www.iso.com>. Accedido el 4 de junio de 2004.
- Jacobson I., Booch G., Rumbaugh J. El Proceso Unificado de Desarrollo de Software. Addison-Wesley, Madrid, 2000.
- OMG home page. <http://www.omg.org>. Accedido el 4 de junio de 2004.
- Piattini M., Calvo-Manzano J., Cervera J., Fernández L. Análisis y Diseño Detallado de Aplicaciones Informáticas de Gestión .RA-MA, Madrid, 1993.
- Pressman R.S. Ingeniería del Software. Un enfoque práctico (5ª ed.) Mc Graw-Hill; New York, 2001.
- Rumbaugh J., Jacobson I., Booch G. El Lenguaje Unificado de Modelado. Manual de Referencia. Addison-Wesley, Madrid, 2000.
- Sommerville I. Ingeniería de software. 6ª edición. PrenticeHall – Pearson educación, México, 2002.



UML home page. <http://www.uml.org>. Accedido el 4 de junio de 2004.

Yourdon E., Whitehead K., Thomann J., Oppel K., Nevermann P. *Mainstream Objects: An Analysis and Design Approach for Business*. Prentice-Hall, Upper Saddle River, NJ (USA), 1995.

## 8.6 Bibliografía de CETP

- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; 62:1198-1211.
- Bernard S, Moulin P, Lagrost L, Picard S, Elchebly M, Ponsin G, et al. Association between plasma HDL-cholesterol concentration and TaqIB CETP gene polymorphism in non-insulin-dependent diabetes mellitus. *J Lipid Res* 1998; 39:59-65.
- Boerwinkle E, Utermann G. Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am J Hum Genet* 1988; 42:104-112.
- Cupples LA, Gagnon DR, Kannel WB. Long- and short-term risk of sudden coronary death. *Circulation* 1992; 85:111-118.
- Dawber TR, Meadors GF, Moore R. Epidemiological approaches to heart disease: The Framingham Study. *Am J Public Health* 1951; 41:279-286.
- Drayna D, Jarnagin AS, McLean J, Henzel W, Kohr W, Fielding C, et al. Cloning and sequencing of human cholesteryl ester transfer protein cDNA. *Nature* 1987; 327:632-634.
- Drayna D, Lawn R. Multiple RFLP's at the human cholesteryl ester transfer protein (CETP) locus. *Nucleic Acids Res* 1987; 15:4698.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975; 4:518-525.
- Freeman D, Packard CJ, Shepherd J, Gaffney D. Polymorphisms in the gene coding for cholesteryl ester transfer protein are related to plasma high density lipoprotein cholesterol and transfer ester activity. *Clin Sci* 1990; 79:575-581.
- Freeman D, Shepherd J, Packard CJ, Humphries SE, Gaffney D. An *StuI* RFLP at the human cholesteryl ester transfer protein (CETP) locus. *Nucleic Acids Res* 1989; 17:2880.
- Freeman DJ, Griffin BA, Holmes AP, Lindsay GM, Gaffney D, Packard CJ, et al. Regulation of plasma HDL cholesterol and subfraction distribution by genetic and environmental factors: Associations between the TaqI B RFLP in the CETP gene and smoking and obesity. *Arterioscler Thromb* 1994; 14:336-344.

- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972; 18:499-502.
- Fumeron F, Betoulle D, Luc G, Behague I, Ricard B, Poirier O, et al. Alcohol intake modulates the effect of a polymorphism of the cholesteryl ester transfer protein gene on plasma high density lipoprotein and the risk of myocardial infarction. *J Clin Invest* 1995; 96:1664-1671.
- Gordon DJ, Rifkind BM. High-density lipoprotein. The clinical implications of recent studies. *N Engl J Med* 1989; 321:1311-1316.
- Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* 1977; 62:707-714.
- Gudnason V, Kakko S, Nicaud V, Savolainen MJ, Kesaniemi YA, Thavanainen E, et al. Cholesteryl ester transfer protein gene effect on CETP activity and plasma high density lipoprotein in European populations. *Eur J Clin Invest* 1999; 29:116-128.
- Hannuksela ML, Liinamaa MJ, Kesäniemi YA, Savolainen MJ. Relation of polymorphisms in the cholesteryl ester transfer protein gene to transfer protein activity and plasma lipoprotein levels in alcohol drinkers. *Atherosclerosis* 1994; 110:35-44.
- Hill SA, Nazir DJ, Jayaratne P, Bamford KS, McQueen MJ. Mutations in cholesteryl ester transfer protein and hepatic lipase in a North American population. *Clin Biochem* 1997; 30:413-418.
- Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res* 1990; 31:545-548.
- Inazu A, Brown ML, Hesler CB, Agellon LB, Koizumi J, Takata K, et al. Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *N Engl J Med* 1990; 323:1234-1238.
- Inazu A, Jiang X-C, Haraki T, Yagi K, Kamon N, Koizumi J, et al. Genetic cholesteryl ester transfer protein deficiency caused by two prevalent mutations as a major determinant of increased levels of high density lipoprotein cholesterol. *J Clin Invest* 1994; 94:1872-1882.
- Jiang XC, Agellon LB, Walsh A, Breslow JL, Tall A. Dietary cholesterol increases transcription of the human cholesteryl ester transfer protein gene in transgenic mice. Dependence on natural flanking sequences. *J Clin Invest* 1992; 90:1290-1295.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol* 1979; 110:281-290.
- Kauma H, Savolainen MJ, Heikkilä R, Rantala AO, Lilja M, Reunanen A, et al. Sex difference in the regulation of plasma high density lipoprotein cholesterol by genetic and environmental factors. *Hum Genet* 1996; 97:156-162.

- Koizumi J, Inazu A, Yagi K, Koizumi I, Uno Y, Kajinami K, et al. Serum lipoprotein lipid concentration and composition in homozygous and heterozygous patients with cholesteryl ester transfer protein deficiency. *Atherosclerosis* 1991; 90:189-196.
- Kondo I, Berg K, Drayna DT, Lawn RM. DNA polymorphism at the locus for human cholesteryl ester transfer protein (CETP) is associated with high density lipoprotein cholesterol and apolipoprotein levels. *Clin Genet* 1989; 35:49-56.
- Kuivenhoven JA, de Knijff P, Boer JMA, Smalheer HA, Botma GJ, Seidell JC, et al. Heterogeneity at the CETP gene locus - Influence on plasma CETP concentrations and HDL cholesterol levels. *Arterioscler Thromb Vasc Biol* 1997; 17:560-568.
- Kuivenhoven JA, Jukema JW, Zwinderman AH, de Knijff P, McPherson R, Bruschke AV, et al. The role of a common variant of the cholesteryl ester transfer protein gene in the progression of coronary atherosclerosis. The Regression Growth Evaluation Statin Study Group. *N Engl J Med* 1998; 338:86-93.
- Logan AG, Greenwood CM. Salt, blood pressure, and cointervention [letter; comment]. *Archives of Internal Medicine* 1997; 157:2666-2667.
- Manninen V, Elo MO, Frick MH, Haapa K, Heinonen OP, Heinsalmi P, et al. Lipid alterations and decline in the incidence of coronary heart disease in the Helsinki Heart Study. *JAMA* 1988; 260:641-651.
- McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and apolipoprotein fractions. *Clin Chim Acta* 1987; 166:1-9.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1989; 16:1215
- Mitchell RJ, Earl L, Williams J, Bisucci T, Gasiamis H. Polymorphisms of the gene coding for the cholesteryl ester transfer protein and plasma lipid levels in Italian and Greek migrants to Australia. *Human Biology* 1994; 66:13-25.
- Ordovas JM, Peterson JP, Santaniello P, Cohn J, Wilson PWF, Schaefer EJ. Enzyme linked immunosorbent assay for human apolipoprotein B. *J Lipid Res* 1987; 28:1216
- Otvos JD, Jeyarajah EJ, Bennett DW, Krauss RM. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. *Clin Chem* 1992; 38:1632-1638.
- Otvos JD. Measurement of lipoprotein subclass profile by nuclear magnetic resonance. In: Rifai N, Warnick GR, Dominiczak MH, editors. *Handbook of lipoprotein testing*. Washington: AACC Press, 1997:497-508.
- Schaefer EJ, Ordovas JM. Metabolism of the apolipoproteins A-I, A-II, and A-IV. In: Segrest J, Albers J, editors. *Methods in Enzymology, Plasma Lipoproteins, Part B: Characterization, Cell Biology and Metabolism*. Academic Press, 1986:420-442.
- Stevenson CG. Cholesterol ester transfer protein: a molecule with three faces? *Crit Rev Clin Lab Sci* 1998; 35:517-546.

- Takegoshi T, Haba T, Kitoh C, Inazu A, Koizumi J, Mabuchi H, et al. Compound heterozygote of cholesteryl-ester transfer protein deficiency in a patient with hyperalphalipoproteinemia. *Atherosclerosis* 1992; 96:83-85.
- Tamminen M, Kakko S, Kesäniemi YA, Savolainen MJ. A polymorphic site in the 3' untranslated region of the cholesteryl ester transfer protein (CETP) gene is associated with low CETP activity. *Atherosclerosis* 1996; 124:237-247.
- Tenkanen H, Koskinen P, Kontula K, Aalto-Setälä K, Mänttari M, Manninen V, et al. Polymorphisms of the gene encoding cholesterol ester transfer protein and serum lipoprotein levels in subjects with and without coronary heart disease. *Hum Genet* 1991; 87:574-578.
- Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg<sup>2+</sup> precipitation procedure for quantitation of high-density-lipoprotein cholesterol. *Clin Chem* 1982;28:1379-88.
- Yamashita S, Sakai N, Hirano K, Arai T, Ishigami M, Maruyama T, et al. Molecular genetics of plasma cholesteryl ester transfer protein. *Curr Opin Lipidol* 1997; 8:101-110.
- Zuliani G, Hobbs HH. EcoNI polymorphism in the human cholesteryl ester transfer protein (CETP) gene. *Nucleic Acids Res* 1990; 18:2834.

## 8.7 Bibliografía de APOE

- Aguilar CA, Talavera G, Ordovas JM, Barriguete JA, Guillen LE, Leco ME, Pedro-Botet J, Gonzalez-Barranco J, Gomez-Perez FJ, Rull JA. The apolipoprotein E4 allele is not associated with an abnormal lipid profile in a Native American population following its traditional lifestyle. *Athero* 1999;142:409-414.
- Aguilar CA, Talavera G, Ordovas JM, et al. The apolipoprotein E4 allele is not associated with an abnormal lipid profile in a Native American population following its traditional lifestyle. *Atherosclerosis* 1999;142:409-14.
- Alonso-Villaverde C, Sardà P, Vallbé JC, Heras M, Pérez-Jimenez F, Pedro-Botet JC, Ros-Rahola E, Masana L. Manifestaciones clínicas de la hipercolesterolemia familiar en una población mediterránea. *Med Clin (Barc)* 1999;113:521-525.
- Ayaori M. Beneficial effects of alcohol withdrawal on LDL particle size distribution and oxidative susceptibility in subjects with alcohol-induced hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 1997;17:2540-7.
- Barter PJ. Cholesterol and cardiovascular disease: Basic science. *Aust N Z J Med* 1994;24:83-88.
- Bercedo A, Gonzalez-Lamuno D, Muñoz O, et al. Association between lipid profile and Apo E genotype in Spanish children (8-15 years old). *Ann Esp Pediatr* 1998;42:120-4.

- Bétard C, Kessling AM, Roy M, Davignon J. Influence of genetic variability in the nondeletion LDL- receptor allele on phenotypic variation in French-Canadian familial hypercholesterolemia heterozygotes sharing a 'null' LDL-receptor gene defect. *Athero* 1996;119:43-55.
- Bijnen FCH, Feskens EJM, Caspersen CJ, Giampaoli S, Nissinen AM, Menotti A, Mosterd WL, Kromhout D. Physical activity and cardiovascular risk factors among elderly men in Finland, Italy, and the Netherlands. *Am J Epidemiol* 1996;143:553-561.
- Bodurtha JN, Chen CW, Mosteller M, Nance WE, Schieken RM, Segrest J. Genetic and environmental contributions to cholesterol and its subfractions in 11-year-old twins: The Medical College of Virginia twin study. *Arterioscler Thromb* 1991;11:844-850.
- Boer JMA, Ehnholm C, Menzel H-J, Havekes L, Rosseneu M, O'Reilly DS, Tiret L. Interaction between lifestyle-related factors and the apoE polymorphism on plasma lipids and apolipoproteins. The EARS study. *Arterioscler Thromb Vasc Biol* 1997;17:1675-1681.
- Boerwinkle E, Utermann G. Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am J Hum Genet* 1988;42:104-12.
- Bottoni A, Cannella C, Del Balzo V. Lifestyle and dietary differences in smokers and non-smokers from an Italian employee population. *Public Health* 1997;111:161-4.
- Braeckman L, De Bacquer D, Rosseneu M, De Backer G. Apolipoprotein E polymorphism in middle-aged Belgian men: Phenotype distribution and relation to serum lipids and lipoproteins. *Atherosclerosis* 1996;120:67-73.
- Breslow JL, McPherson J, Nussbaum AL, Williams HW, Lofquist-Kahl F, Karathanasis SK, Zannis VI. Identification and DNA sequence of a human apolipoprotein E cDNA clone. *J Biol Chem* 1982;257:14639-14641.
- Cariolou MA, Kokkofitou A, Manoli P, Christou S, Karagrigoriou A, Middleton L. Underexpression of the apolipoprotein E2 and E4 alleles in the Greek Cypriot population of Cyprus. *Genet Epidemiol* 1995;12:489-497.
- Chaves FJ, Puig O, Garcia-Sogo M, Real J, Gil JV, Ascaso JF, Carmena R, Armengod ME. Seven DNA polymorphisms in the LDL receptor gene: application to the study of familial hypercholesterolemia in Spain. *Clin Genet* 1996;50:28-35.
- Colditz GA, Giovannucci E, Rimm EB, et al. Alcohol intake in relation to diet and obesity in women and men. *Am J Clin Nutr* 1991;54:49-55.
- Connelly PW, Maguire GF, Lee M, Little JA. Plasma lipoproteins in familial hepatic lipase deficiency. *Arterio* 1990;10:40-48.
- Corella D, Sáiz C, Guillén M, Portolés O, Mulet F, González JI, Ordovas JM. Association of TaqIB polymorphism in the cholesteryl ester transfer protein gene with plasma lipid levels in a healthy Spanish population. *Athero* 2000;(in press).
- Coughlin SS. The intersection of genetics, public health, and preventive medicine. *Am J Prev Med* 1999;16:89-90.

- Cumming AM, Robertson FW. Polymorphism at the apoprotein-E locus in relation to risk of coronary disease. *Clin Genet* 1984;25:310-313.
- Cupples LA, Gagnon DR, Kannel WB. Long- and short-term risk of sudden coronary death. *Circulation* 1992;85:111-18.
- Davignon J, Cohn JS, Mabile L, Bernier L. Apolipoprotein E and atherosclerosis: insight from animal and human studies. *Clin Chim Acta* 1999;286:115-143.
- Davignon J, Gregg RE, Sing CF. Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 1988;8:1-21.
- Davignon J, Gregg RE, Sing CF. Apolipoprotein E polymorphism and atherosclerosis. *Arterio* 1988;8:1-21.
- Dawber TR, Meadors GF, Moore R. Epidemiological approaches to heart disease: The Framingham Study. *Am J Public Health* 1951;41:279-86.
- Deiana L, Pes GM, Carru C, Errigo A, Pettinato S, Carcassi C, Baggio G, Contu L. Lack of influence of apolipoprotein E4 on lipoprotein levels in the island population of Sardinia. *Eur J Clin Invest* 1998;28:290-294.
- Dixon LB, Shannon BM, Tershakovec AM, Bennett MJ, Coates PM, Cortner JA. Effects of family history of heart disease, apolipoprotein E phenotype, and lipoprotein(a) on the response of children's plasma lipids to change in dietary lipids. *Am J Clin Nutr* 1997;66:1207-17.
- Eichner JE, Kuller LH, Ferrell RE, Meilahn EN, Kamboh MI. Phenotypic effects of apolipoprotein structural variation on lipid profiles.III. Contribution of apolipoprotein E phenotype to prediction of total cholesterol, apolipoprotein B, and low density lipoprotein cholesterol in the healthy women study. *Arteriosclerosis* 1990; 10: 379-85.
- Ellsworth DL, Sholinsky P, Jaquish C, Fabsitz RR, Manolio TA. Coronary heart disease. At the interface of molecular genetics and preventive medicine. *Am J Prev Med* 1999;16:122-133.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975;4:518-25.
- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972;18:499-502.
- Frohlich JJ. Effects of alcohol on plasma lipoprotein metabolism. *Clin Chim Acta* 1996;246:39-49.
- Fujishima M, Kiyohara Y, Ueda K, Hasuo Y, Kato I, Iwamoto H. Smoking as cardiovascular risk factor in low cholesterol population: The Hisayama Study. *Clin Exp Hypertens [A]* 1992;14A:99-108.
- Garry PJ, Baumgartner RN, Brodie SG, et al. Estrogen replacement therapy, serum lipids, and polymorphism of the apolipoprotein E gene. *Clin Chem* 1999;45:1214-23.

- Gaziano JM, Buring JE, Breslow JL, et al. Moderate alcohol intake, increased levels of high-density lipoprotein and its subfractions, and decreased risk of myocardial infarction. *N Engl J Med* 1993;329:1829-34.
- Gené M, Moreno P, Esquerre M, Prat A, Huguet E, Adroer R, Oliva R. Low apolipoprotein E E4 allele frequency in the population of Catalonia (Spain) determined by PCR-RFLP and laser fluorescent sequencer. *Eur J Epidemiol* 1997;13:841-843.
- Gerdes LU, Klausen IC, Sihm I, Faergeman O. Apolipoprotein E polymorphism in a Danish population compared to findings in 45 other study populations around the World. *Genet Epidemiol* 1992;9:155-167.
- Glueck CJ, Hogg E, Allen C, Gartside PS. Effects of alcohol ingestion on lipids and lipoproteins in normal men: isocaloric metabolic studies. *Am J Clin Nutr* 1980;33:2287-93.
- Gómez-Coronado D, Alvarez JJ, Entrala A, Olmos JM, Herrera E, Lasuncion MA. Apolipoprotein E in men and women from a Spanish population: allele frequencies and influence on plasma lipids and apolipoproteins. *Athero* 1999;147:167-176.
- Gómez-Gerique JA, Gutiérrez-Fuentes JA, Montoya MT, Porres A, Rueda A, Avellaneda A, Rubio MA. Perfil lipídico de la población española: estudio DRECE (Dieta y Riesgo de Enfermedad Cardiovascular en España). *Med Clin (Barc)* 1999;113:730-735.
- Greenfield TK, Rogers JD. Who drinks most of the alcohol in the US? The policy implications. *J Stu Alcohol* 1999;60:78-89.
- Gronbaek M, Deis A, Becker U, et al. Alcohol and mortality: is there a U-shaped relation in elderly people? *Age & Ageing* 1998;27:739-44.
- Hagberg JM, Ferrel RE, Dengel DR, Wilund KR. Exercise training-induced blood pressure and plasma lipid improvements in hypertensives may be genotype dependent. *Hypertension* 1999;34:18-23.
- Hallman DM, Boerwinkle E, Saha N, Sandholzer C, Jürgen Menzel H, Császár A, Utermann G. The apolipoprotein E polymorphism: A comparison of allele frequencies and effects in nine populations. *Am J Hum Genet* 1991;49:338-349.
- Hegele RA, Evans AJ, Tu L, Ip G, Brunt JH, Connelly PW. A gene-gender interaction affecting plasma lipoproteins in a genetic isolate. *Arterioscler Thromb* 1994;14:671-8.
- Hegele RA. Gene-environment interactions in atherosclerosis. *Mol Cell Biochem* 1992;113:177-186.
- Hein HO, Suadicani P, Gyntelberg F. Alcohol consumption, serum low density lipoprotein cholesterol concentration, and risk of ischaemic heart disease: Six year follow up in the Copenhagen male study. *Brit Med J* 1996;312:736-41.
- Hendriks HFJ, Veenstra J, Van Tol A, Groener JE, Schaafsma G. Moderate doses of alcoholic beverages with dinner and postprandial high density lipoprotein composition. *Alcohol* 1998;33:403-10.

- Higuchi M, Iwaoka K, Ishii K, et al. Plasma lipid and lipoprotein profiles in pre- and post-menopausal middle-aged runners. *Clin Physiol* 1990;10:69-76.
- Hirano K, Yamashita S, Sakai N, et al. Low-density lipoproteins in hyperalphalipoproteinemic heavy alcohol drinkers have reduced affinity for the low-density lipoprotein receptor. *Clin Biochem* 1992;25:357-62.
- Hixson JE, PDAY Research Group. Apolipoprotein E polymorphisms affect atherosclerosis in young males. *Arterioscler Thromb* 1991;11:1237-1244.
- Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res* 1990;31:545-548.
- Howard BV, Gidding SS, Liu K. Association of apolipoprotein E phenotype with plasma lipoproteins in African-American and white young adults. *Am J Epidemiol* 1998;148:859-868.
- Jarvik GP, Austin MA, Fabsitz RR, et al. Genetic influences on age-related change in total cholesterol, low density lipoprotein-cholesterol, and triglyceride levels: Longitudinal apolipoprotein E genotype effects. *Genet Epidemiol* 1994;11:375-84.
- Jarvik GP, Beaty TH, Gallagher PR, Coates PM, Cortner JA. Genotype at a major locus with large effects on apolipoprotein B levels predicts familial combined hyperlipidemia. *Genet Epidemiol* 1993;10:257-270.
- Jones DY, Judd JT, Taylor PR, Campbell WS, Nair PP. Menstrual cycle effect on plasma lipids. *Metabolism* 1988;37:1-2.
- Kamboh MI, Aston CE, Hamman RF. The relationship of APOE polymorphism and cholesterol levels in normoglycemic and diabetic subjects in a biethnic population from the San Luis Valley, Colorado. *Athero* 1995;112:145-159.
- Kamboh MI, Evans RW, Aston CE. Genetic effect of apolipoprotein(a) and apolipoprotein E polymorphisms on plasma quantitative risk factors for coronary heart disease in American black women. *Athero* 1995;117:73-81.
- Kamboh MI, Weiss KM, Ferrell RE. Genetic studies of human apolipoproteins. XVI. APOE polymorphism and cholesterol levels in the Mayans of the Yucatan Peninsula, Mexico. *Clin Genet* 1991;39:26-32.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol* 1979;110:281-90.
- Kauma H, Savolainen MJ, Rantala AO, et al. Apolipoprotein E phenotype determines the effect of alcohol on blood pressure in middle-aged men. *Am J Hypertens* 1998;11:1334-43.
- Keil U, Chambless LE, Döring A, Filipiak B, Stieber J. The relation of alcohol intake to coronary heart disease and all-cause mortality in a beer-drinking population. *Epidemiology* 1997;8:150-6.



- Khoury MJ, Yang Q. From genes to public health: the application of genetic technology in disease prevention. *Am J Public Health* 1996;86:1717-1722.
- Kiechl S, Willeit J, Rungger G, Egger G, Oberhollenzer F, Bonora E. Alcohol consumption and atherosclerosis: what is the relation? Prospective results from the Bruneck Study. *Stroke* 1998;29:900-7.
- Lahoz C, Ordovás JM. ApoE: lípidos plasmáticos, cardiopatía isquémica y enfermedad de Alzheimer. *Med Clin (Barc)* 1997;109:31-36.
- Langer RD, Criqui MH, Reed DM. Lipoproteins and blood pressure as biological pathways for effect of moderate alcohol consumption on coronary heart disease. *Circulation* 1992;85:910-5.
- Lefevre M, Ginsberg HN, Kris-Etherton PM, et al. ApoE genotype does not predict lipid response to changes in dietary saturated fatty acids in a heterogeneous normolipidemic population. The DELTA Research Group. Dietary Effects on Lipoproteins and Thrombogenic Activity. *Arterioscler Thromb Vasc Biol* 1997;17:2914-23.
- Lehtimäki T, Frankberg-Lakkala H, Solakivi T, et al. The effect of short-term fasting, apolipoprotein E gene polymorphism, and sex on plasma lipids. *Am J Clin Nutr* 1997;66:599-605.
- Lehtimäki T, Moilanen T, Nikkari T, Solakivi T, Porkka K, Ehnholm C, Rönnemaa T, Åkerblom HK, Uhari M, Matti Nuutinen E, Kaprio EA, Pesonen E, Pietikäinen M, Dahl M, Salo MK, Viikari J. Regional differences in apolipoprotein E polymorphism in Finland. *Ann Med* 1991;23:61-66.
- Lehtimäki T, Porkka K, Viikari J, Ehnholm C, Åkerblom HK, Nikkari T. Apolipoprotein E phenotypes and serum lipids in newborns and 3-year-old children: The cardiovascular risk in young Finns study. *Pediatrics* 1994;94:489-93.
- Liinamaa MJ, Hannuksela ML, Kesaniemi YA, Savolainen MJ. Altered transfer of cholesteryl esters and phospholipids in plasma from alcohol abusers. *Arterioscler Thromb Vasc Biol* 1997;17:2940-7.
- Lin RC, Miller BA, Kelly TJ. Concentrations of apolipoprotein AI, AII, and E in plasma and lipoprotein fractions of alcoholic patients: Gender differences in the effects of alcohol. *Hepatology* 1995;21:942-9.
- Lopez-Miranda J, Ordovas JM, Mata P, et al. Effect of apolipoprotein E phenotype on diet-induced lowering of plasma low density lipoprotein cholesterol. *J Lipid Res* 1994;35:1965-75.
- Lucotte G, Loirat F, Hazout S. Pattern of gradient of apolipoprotein E allele \*4 frequencies in western Europe. *Human Biology* 1997;69:253-262.
- Mannisto S, Uusitalo K, Roos E, Fogelholm M, Pietinen P. Alcohol beverage drinking, diet and body mass index in a cross-sectional survey. *Eur J Clin Nutr* 1997;51:326-32.

- Martin MJ, Hulley SB, Browner WS, Kuller LH, Wentworth D. Serum cholesterol, blood pressure, and mortality: implications from a cohort of 361,662 men. *Lancet* 1986;2:933-936.
- McConnell MW, Vavouranakis I, Wu LL, Vaughan DE, Ridker PM. Effects of a single, daily alcoholic beverage on lipid and hemostatic markers of cardiovascular risk. *Am J Cardiol* 1997;80:1226-28.
- McElduff P, Dobson AJ. How much alcohol and how often? population based case-control study of alcohol consumption and risk of major coronary event. *Brit Med J* 1997;314:1159-64.
- McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and apolipoprotein fractions. *Clin Chim Acta* 1987;166:1-9.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1989;16:1215.
- Muros M, Rodríguez-Ferrer C. Apolipoprotein E polymorphism influence on lipids, apolipoproteins and Lp(a) in a Spanish population underexpressing apo E4. *Athero* 1996;121:13-21.
- Myers RH, Schaefer EJ, Wilson PWF, D'Agostino R, Ordovas JM, Espino A, Au R, White RF, Knoefel JE, Cobb JL, McNulty KA, Beiser A, Wolf PA. Apolipoprotein E e4 association with dementia in a population- based study: The Framingham study. *Neurology* 1996;46:673-677.
- Nakanishi N, Nakamura K, Ichikawa S, Suzuki K, Tataru K. Relationship between lifestyle and serum lipid and lipoprotein levels in middle-aged Japanese men. *Eur J Epidem* 1999;15:341-8.
- Nishiwaki M, Ishikawa T, Ito T, et al. Effects of alcohol on lipoprotein lipase, hepatic lipase, cholesteryl ester transfer protein, and lecithin:cholesterol acyltransferase in high-density lipoprotein cholesterol elevation. *Atherosclerosis* 1994;111:99-109.
- Okada T, Sato Y, Iwata F, Hara M, Kim H, Harada K. Relationship of apolipoprotein E phenotypes to serum lipid and lipoprotein levels in Japanese schoolchildren. *Acta Paediatr* 1998;87:460-1.
- Ordovas JM, Lopez-Miranda J, Perez-Jimenez F, Rodriguez CR, Park J-S, Cole T, Schaefer EJ. Effect of apolipoprotein E and A-IV phenotypes on the low density lipoprotein response to HMG-CoA reductase inhibitor therapy. *Athero* 1995.
- Ordovas JM, Schaefer EJ. Genes, variation of cholesterol and fat intake and serum lipids. *Current Opinion In Lipidology* 1999;10:15-22.
- Ordovas JM, Schaefer EJ. Treatment of dyslipidemia: genetic interactions with diet and drug therapy. *Current Atherosclerosis Reports* 1999;1:16-23.
- Ordovas JM. The genetics of serum lipid responsiveness to dietary interventions. *Proc Nutr Soc* 1999;58:171-87.

- Ordovas JM. The genetics of serum lipid responsiveness to dietary interventions. *Proc Nutr Soc* 1999;58:171-187.
- Pablos-Mendez A, Mayeux R, Ngai C, Shea S, Berglund L. Association of apo E polymorphism with plasma lipid levels in a multiethnic elderly population. *Arterioscler Thromb Vasc Biol* 1997;17:3534-41.
- Pérez-Estrada E, Escolar JL, Valdivielso P, Andrade R, Pérez-Estrada M, González-Santos P. Fenotipos de la apoproteína E. Un estudio de la población de Málaga. *Rev Clin Esp* 1990;187:3-5.
- Posner BM, Cupples LA, Miller DR, Cobb JL, Lutz KJ, D'Agostino RB. Diet, menopause, and serum cholesterol levels in women: The Framingham Study. *Am Heart J* 1993;125:483-89.
- Posner BM, Martin-Munley SS, Smigelski C, et al. Comparison of the techniques for estimating nutrient intake: The Framingham Study. *Epidemiology* 1992;3:171-7.
- Rakic V, Puddey IB, Dimmitt SB, Burke V, Beilin LJ. A controlled trial of the effects of pattern of alcohol intake on serum lipid levels in regular drinkers. *Atherosclerosis* 1998;137:243-52.
- Reina M, Brunzell JD, Deeb SS. Molecular basis of familial chylomicronemia: Mutations in the lipoprotein lipase and apolipoprotein C-II genes. *J Lipid Res* 1992;33:1823-1832.
- Renaud SC, Gueguen R, Schenker J, d'Houtaud A. Alcohol and mortality in middle-aged men from eastern France. *Epidemiology* 1998;9:184-8.
- Rimm EB, Klatsky A, Grobbee D, Stampfer MJ. Review of moderate alcohol consumption and reduced risk of coronary heart disease: Is the effect due to beer, wine, or spirits. *Brit Med J* 1996;312:731-6.
- Rimm EB, Williams P, Criqui M, Stampfer MJ. Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. *Brit Med J* 1999;319:1523-8.
- Romelsjo A, Leifman H, Nystrom SA. A comparative study of two methods for the measurement of alcohol consumption in the general population. *Int J Epidemiol* 1995;24:929-36.
- Sandholzer C, Delport R, Vermaak H, Utermann G. High frequency of the apo epsilon 4 allele in Khoi San from South Africa. *Hum Genet* 1995;95:46-8.
- Schaefer EJ, Lamon-Fava S, Johnson S, Ordovas JM, Schaefer MM, Castelli WP, Wilson PWF. Effects of gender and menopausal status on the association of apolipoprotein E phenotype with plasma lipoprotein levels: Results from the Framingham Offspring Study. *Arterioscler Thromb* 1994;14:1105-1113.
- Shaper AG, Wannamethee SG. The J-shaped curve and changes in drinking habit. *Novartis Found Symp* 1998;216:173-88.

- Shaper AG. Alcohol and mortality: a review of prospective studies. *Br J Addict* 1990;85:837-47.
- Simon JA, Fong J, Bernert JT, Jr., Browner NS. Relation of smoking and alcohol consumption to serum fatty acids. *Am J Epidemiol* 1996;144:325-34.
- Sing CF, Davignon J. Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am J Hum Genet* 1985;37:268-285.
- Sociedad Espanola de Arteriosclerosis, Sociedad Española de Medicina Interna, y Liga Lucha contra la Hipertensión Arterial. Recomendaciones para la prevención primaria de la enfermedad cardiovascular. *Clin Invest Arteriosclerosis* 1994;6:62-102.
- Sytkowski PA, Kannel WB, D'Agostino RB. Changes in risk factors and the decline in mortality from cardiovascular disease. The Framingham Heart Study. *N Engl J Med* 1990;322:1635-1641.
- Talmud P, Tybjærg-Hansen A, Bhatnagar D, Mbewu A, Miller JP, Durrington P, Humphries S. Rapid screening for specific mutations in patients with a clinical diagnosis of familial hypercholesterolaemia. *Athero* 1991;89:137-141.
- The Expert Panel. Summary of the second report of the National Cholesterol Education Program (CNEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). *J Am Med Assoc* 1993;269:3015-3023.
- Tiret L, de Knijff P, Menzel H-J, Ehnholm C, Nicaud V, Havekes LM. ApoE polymorphism and predisposition to coronary heart disease in youths of different European populations: The EARS Study. *Arterioscler Thromb* 1994;14:1617-1624.
- Tucker KL, Rich S, Rosenberg I, et al. Plasma vitamin B12 concentrations relate to intake source in the Framingham Offspring study. *Am J Clin Nutr* 2000; 71: 514-22.
- Utermann G. Apolipoprotein E polymorphism in health and disease. *Am Heart J* 1987;113:433-440.
- Van Tol A, Van der Gaag MS, Scheek LM, Van Gent T, Hendriks HFJ. Changes in postprandial lipoproteins of low and high density caused by moderate alcohol consumption with dinner. *Atherosclerosis* 1998;141:S101-S3.
- Vasisht S, Pant MC, Srivastava LM. Effect of alcohol on serum lipids & lipoproteins in male drinkers. *Indian J Med Res* 1992;96:333-7.
- Wannamethee G, Shaper AG. Blood lipids: the relationship with alcohol intake, smoking, and body weight. *J Epidemiol Commun Health* 1992;46:197-202.
- Warnick R, Benderson J, Albers JJ. Dextran Sulfate-Mg precipitation procedure for quantitation of high density lipoprotein cholesterol. *Clin Chem* 1982;28:1379-88.
- Whichelow MJ, Prevost AT. Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. *Br J Nutr* 1996;76:17-30.

- Willett WC, Sampson LA, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51-65.
- Wilson PWF, Myers RH, Larson MG, Ordovas JM, Wolf PA, Schaefer EJ. Apolipoprotein E alleles, dyslipidemia, and coronary heart disease: The Framingham Offspring Study. *JAMA* 1994;272:1666-1671.
- Wilson PWF, Schaefer EJ, Larson MG, Ordovas JM. Apolipoprotein E alleles and risk of coronary disease - A meta-analysis. *Arterioscler Thromb Vasc Biol* 1996;16:1250-1255.
- Zerba KE, Ferrell RE, Sing CF. Genotype-environment interaction: Apolipoprotein E (ApoE) gene effects and age as an index of time and spatial context in the human . *Genetics* 1996;143:463-78.

## 8.8 Bibliografía de APOA1

- Akita H, Chiba H, Tsuji M, Hui SP, Takahashi Y, Matsuno K, Kobayashi K. Evaluation of G-to-A substitution in the apolipoprotein A-I gene promoter as a determinant of high-density lipoprotein cholesterol level in subjects with and without cholesteryl ester transfer protein deficiency. *Hum Genet* 1995;96:521-526.
- Barre DE, Guerra R, Verstraete R, Wang Z, Grundy SM, Cohen JC. Genetic analysis of a polymorphism in the human apolipoprotein A-I gene promoter: Effect on plasma HDL-cholesterol levels. *J Lipid Res* 1994;35:1292-1296.
- Barth JD, Jansen H, Kromhout D, Reiber JH, Birkenhager JC, Arntzenius AC. Progression and regression of human coronary atherosclerosis. The role of lipoproteins, lipases and thyroid hormones in coronary lesion growth. *Athero* 1987;68:51-58.
- Bruckert E, Von Eckardstein A, Funke H, Beucler I, Wiebusch H, Turpin G, Assmann G. The replacement of arginine by cysteine at residue 151 in apolipoprotein A-I produces a phenotype similar to that of apolipoprotein A-IMilano. *Athero* 1997;128:121-128.
- Carmena-Ramon RF, Ordovas JM, Ascaso JF, Real J, Priego MA, Carmena R. Influence of genetic variation at the apoA-I gene locus on lipid levels and response to diet in familial hypercholesterolemia. *Athero* 1998;139:107-113.
- Civeira F, Pocovi M, Cenarro A, Garces C, Ordovas JM. Adenine for guanine substitution - 78 base pairs to the apolipoprotein(APO) A-I gene: relation with high density lipoprotein cholesterol and apoA-I concentrations. *Clin Genet* 1993;44:307-312.
- Cupples LA, Gagnon DR, Kannel WB. Long- and short-term risk of sudden coronary death. *Circ* 1992;85:111-118.

- Danek GM, Valenti M, Baralle FE, Romano M. The A/G polymorphism in the 278 position of the apolipoprotein A-I promoter does not have a direct effect on transcriptional efficiency. *Biochim Biophys Acta* 1998;1398:67-74.
- Dawber TR, Meadors GF, Moore R. Epidemiological approaches to heart disease: The Framingham Study. *Am J Public Health* 1951;41:279-286.
- Deckelbaum RJ, Fisher EA, Winston M, Kumanyika S, Laurer RM, Pi-Sunyer FX, Jeor S, Schaefer EJ, Weinstein IB. Summary of a scientific conference on preventive nutrition: pediatrics to geriatrics. *Circ* 1999;100:450-456.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975;4:518-525.
- Fielding CJ, Shore VG, Fielding PE. A protein co-factor of lecithin:cholesterol acyltransferase. *Biochem Biophys Res Comm* 1972;46:1493-1498.
- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972;18:499-502.
- Jeenah M, Kessling A, Miller N, Humphries SE. G to A substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI and high density lipoprotein cholesterol concentrations. *Mol Biol Med* 1990;7:233-241.
- Jump DB, Clarke SD, Thelen A, Liimatta M, Ren B, Badin M. Dietary fat, genes, and human health. *Adv Exp Med Biol* 1997;422:176.
- Juo SHH, Wyszynski DF, Beaty TH, Huang HY, Bailey-Wilson JE. Mild association between the A/G polymorphism in the promoter of the apolipoprotein A-I gene and apolipoprotein A-I levels: A meta-analysis. *Am J Med Genet* 1999;82:235-241.
- Kamboh MI, Aston CE, Nestlerode CM, McAllister AE, Hamman RF. Haplotype analysis of two APOA1/MspI polymorphisms in relation to plasma levels of apo A-I and HDL-cholesterol. *Athero* 1996;127:255-262.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol* 1979;110:281-290.
- Karathanasis SK. Apolipoprotein multigene family: tandem organization of human apolipoprotein A-I, C-III and A-IV genes. *Proc Natl Acad Sci USA* 1985;82:6374-6378.
- Lopez-Miranda J, Ordovas JM, Espino A, Marin C, Salas J, Lopez-Segura F, Jimenez-Perez J, Perez-Jimenez F. Influence of mutation in human apolipoprotein A-I gene promoter on plasma LDL cholesterol response to dietary fat. *Lancet*, 1994;343:1246-1249.

- Mata P, Lopez-Miranda J, Pocovi M, Alonso R, Lahoz C, Marin C, Garces C, Cenarro A, Perez-Jimenez F, De Oya M, Ordovas JM. Human apolipoprotein A-I gene promoter mutation influences plasma low density lipoprotein cholesterol response to dietary fat saturation. *Athero* 1998;137:367-376.
- Matsunaga A, Sasaki J, Mori T, Moriyama K, Nishi K, Hidaka K, Arakawa K. Apolipoprotein A-I gene promoter polymorphism in patients with coronary heart disease and healthy controls. *Nutr Metab Cardiovasc Dis* 1995;5:275.
- Mattson FH, Grundy SM. Comparison of the effects of dietary saturated, monounsaturated, and polyunsaturated fatty acids on plasma lipids and lipoproteins in man. *J Lipid Res* 1985;26:194-202.
- McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and apolipoprotein fractions. *Clin Chim Acta* 1987;166:1-9.
- Meng QH, Pajukanta P, Valsta L, Aro A, Pietinen P, Tikkanen MJ. Influence of apolipoprotein A-I promoter polymorphism on lipid levels and responses to dietary change in Finnish adults. *J Intern Med* 1997;241:373-378.
- Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins: A meta-analysis of 27 trials. *Arterioscler Thromb* 1992;12:911-919.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1989;16:1215.
- Norum RA, Lakier JB, Goldstein S, Angel A, Goldberg RB, Black WD, Noffze DK, Dolphin PJ, Edelglass J, Borograd DD, Alaupovic P. Familial deficiency of apolipoprotein A-I and C-III and precocious coronary artery disease. *N Engl J Med* 1982;306:1513-1519.
- Ordovas JM, Civeira F, Genest JJ, Craig S, Robbins AH, Meade T, Pocovi M, Frossard P, Masharani U, Wilson PWF, Salem D, Ward RH, Schaefer EJ. Restriction fragment length polymorphisms of the apolipoprotein A-I, C-III, A-IV gene locus: Relationships with lipids, apolipoproteins, and premature coronary artery disease. *Athero* 1991.
- Ordovas JM, Peterson JP, Santaniello P, Cohn J, Wilson PWF, Schaefer EJ. Enzyme linked immunosorbent assay for human apolipoprotein B. *J Lipid Res* 1987;28:1216.
- Ordovas JM. The genetics of serum lipid responsiveness to dietary interventions. *Proc Nutr Soc* 1999;58:171-187.
- Pagani F, Sidoli A, Giudici GA, Barengi L, Vergani C, Baralle FE. Human apolipoprotein A-I gene promoter polymorphism: Association with hyperalphalipoproteinemia. *J Lipid Res* 1990;31:1371-1377.
- Paul-Hayase H, Rosseneu M, Robinson D, Van Bervliet JP, Deslypere JP, Humphries SE. Polymorphisms in the apolipoprotein (apo) AI-CIII-AIV gene cluster: Detection of genetic variation determining plasma apo AI, apo CIII and apo AIV concentrations. *Hum Genet* 1992;88:439-446.

- Rader DJ, Hoeg JM, Brewer HBJr. Quantitation of plasma apolipoproteins in the primary and secondary prevention of coronary artery disease. *Ann Intern Med* 1994;102:1012-1025.
- Saha N, Tay JSH, Low PS, Humphries SE. Guanidine to adenine (G/A) substitution in the promoter region of the apolipoprotein AI gene is associated with elevated serum apolipoprotein AI levels in Chinese non-smokers. *Genet Epidemiol* 1994;11:255-264.
- Schaefer EJ, Ordovas JM. Metabolism of the apolipoproteins A-I, A-II, and A-IV. In: Segrest J, Albers J, eds. *Methods in Enzymology, Plasma Lipoproteins, Part B: Characterization, Cell Biology and Metabolism*. Academic Press; 1986:420-442.
- Sigurdsson G, Jr., Gudnason V, Sigurdsson G, Humphries SE. Interaction between a polymorphism of the Apo A-I promoter region and smoking determines plasma levels of HDL and Apo A-I. *Arterioscler Thromb* 1992;12:1017-1022.
- Smith JD, Brinton EA, Breslow JL. Polymorphism in the human apolipoprotein A-I gene promoter region. Association of the minor allele with decreased production rate in vivo and promoter activity in vitro. *J Clin Invest* 1992;89:1796-1800.
- Smith JD, Brinton EA, Breslow JL. Polymorphism in the human apolipoprotein A-I gene promoter region. Association of the minor allele with decreased production rate in vivo and promoter activity in vitro. *J Clin Invest* 1992;89:1796-800.
- Talmud P, Ye S, Humphries SE. Polymorphism in the promoter region of the apolipoprotein AI gene associated with differences in apolipoprotein AI levels: the European Atherosclerosis Research Study. *Genet Epidemiol* 1994;11:265-280.
- Tuteja R, Tuteja N, Melo C, Casari G, Baralle FE. Transcription efficiency of human apolipoprotein A-I promoter varies with naturally occurring A to G transition. *FEBS Lett* 1992;304:98-101.
- Warnick R, Benderson J, Albers JJ. Dextran Sulfate-Mg precipitation procedure for quantitation of high density lipoprotein cholesterol. *Clin Chem* 1982;28:1379-1388.

## 8.9 Bibliografía de LIPC

- Berglund L, Oliver EH, Fontanez N, et al. HDL-subpopulation patterns in response to reduction in dietary total and saturated fat intakes in healthy subjects. *Am J Clin Nutr* 1999;70:992-1000.
- Campos H, Dreon DM, Krauss RM. Associations of hepatic and lipoprotein lipase activities with changes in dietary composition and low density lipoprotein subclasses. *J Lipid Res* 1995;36:462-472.



- Corella D, Tucker K, Lahoz C, et al. Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study. *Am J Clin Nutr* 2001;73:736-745.
- Couture P, Otvos JD, Cupples LA, et al. Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles. The Framingham Offspring Study. *Arterioscler Thromb Vasc Biol* 2000;20:815-822.
- Cupples LA, Gagnon DR, Kannel WB. Long- and short-term risk of sudden coronary death. *Circulation* 1992;85:111-118.
- Deeb S, Peng R. The C-514T polymorphism in the human hepatic lipase gene promoter diminishes its activity. *J Lipid Res* 2000;41:155-158.
- Dreon DM, Fernstrom HA, Campos H, et al. Change in dietary saturated fat intake is correlated with change in mass of large-low-density-lipoprotein particles in men. *Am J Clin Nutr* 1998;67:828-836.
- Dugi KA, Amar MJA, Haudenschild CC, et al. In vivo evidence for both lipolytic and nonlipolytic function of hepatic lipase in the metabolism of HDL. *Arterioscler Thromb Vasc Biol* 2000;20:793-800.
- Dugi KA, Brandauer K, Schmidt N, et al. Low hepatic activity is a novel risk factor for coronary artery disease. *Circulation* 2001;104:3057-3062.
- Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975;4:518-525.
- Freeman DS, Otvos JD, Jeyarajah EJ, et al. Relation of lipoprotein subclasses as measured by proton nuclear magnetic resonance spectroscopy to coronary artery disease. *Arterioscler Thromb Vasc Biol* 1998;18:1046-1053.
- Grundy SM, Vega GL, Otvos JD, et al. Hepatic lipase activity influences high density lipoprotein subclass distribution in normotriglyceridemic men: genetic and pharmacological evidence. *J Lipid Res* 1999;40:229-234.
- Guerra R, Wang J, Grundy SM, Cohen JC. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A*. 1997;94:4532-4537
- Guerra R, Wang J, Grundy SM, et al. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *PNAS (USA)* 1997;94:4532-4537.
- Hegele RA, Harris SB, Brunt JH, et al. Absence of association between genetic variation in the LIPC gene promoter and plasma lipoproteins in three Canadian populations. *Atherosclerosis* 1999;146:153-160.
- Jansen H, Verhoeven AJ, Weeks L, et al. A common C-to-T substitution at position -480 of the hepatic lipase promoter associated with a lowered lipase activity in coronary artery disease participants. *Arterioscler Thromb Vasc Biol* 1997;17:2837-2842.

- Juo SH, Han Z, Smith JD, et al. Promoter polymorphisms of hepatic lipase gene influence HDL(2) but not HDL (3) in African American men: CARDIA study. *J Lipid Res* 2001;42:258-264.
- Kasim SE, Martino S, Kim P-N, et al. Dietary and anthropometric determinants of plasma lipoproteins during a long-term low-fat diet in healthy women. *Am J Clin Nutr* 1993;57:146-153.
- Katan MB, Grundy SM, Willett WC. Should a low-fat, high-carbohydrate diet be recommended for everyone? Beyond low-fat diets. *N Engl J Med* 1997;337:563-566.
- Kong C, Nimmo L, Elatrozy T, et al. Smoking is associated with increased hepatic lipase activity, insulin resistance, dyslipaemia and early atherosclerosis in Type 2 diabetes. *Atherosclerosis* 2001;156:373-378.
- Lambert G, Amar MJA, Martin P, et al. Hepatic lipase deficiency decreases the selective uptake of HDL-cholesterol esters in vivo. *J Lipid Res* 2001;41:667-672.
- Mensink RP, Katan MB. Effect of dietary fatty acids on serum lipids and lipoproteins: A meta-analysis of 27 trials. *Arterioscler Thromb* 1992;12:911-919.
- Otvos JD, Jeyarajah EJ, Bennett DW, et al. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. *Clin Chem* 1992;38:1632-1638.
- Rimm EB, Giovannucci EL, Stampfer MJ, et al. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135:1114-1126.
- Santamarina-Fojo S, Haudenschild CC, Amar M. The role of hepatic lipase in lipoprotein metabolism and atherosclerosis. *Curr Opin Lipidol* 1998;9:211-219.
- Shohet RV, Vega GL, Anwar A et al. Hepatic lipase (LIPC) promoter polymorphism in men with coronary artery disease. Allele frequency and effects on hepatic lipase activity and plasma HDL-C concentrations. *Arterioscler Thromb Vasc Biol* 1999;19:1975-1978.
- Tan K-C, Shiu SW, Chu BY. Effects of gender, hepatic lipase gene polymorphism and type 2 diabetes mellitus on hepatic lipase activity in Chinese. *Atherosclerosis* 2001;157:233-239.
- Thuren T. Hepatic lipase and HDL metabolism. *Curr Opin Lipidol* 2000;11:277-283.
- Tu A-Y, Albers JJ. Glucose regulates the transcription of human genes relevant in HDL metabolism: responsive elements for peroxisomes proliferator-activated receptor are involved in the regulation of phospholipid transfer protein. *Diabetes* 2001;50:1851-1856.
- Van't Hooft FM, Lundahl B, Ragogna F, et al. Functional characterization of 4 polymorphisms in promoter region of hepatic lipase gene. *Arterioscler Thromb Vasc Bio* 2000; 20:1335-1339.

- Vega GL, Clark LT, Tang A, et al. Hepatic lipase activity is lower in African American than in white American men: effects of 5' flanking polymorphism in the hepatic lipase gene. *J Lipid Res* 1998;39, 228-232.
- Zambon A, Deeb S, Hokanson JE, et al. Common variants in the promoter of the hepatic lipase gene are associated with lower levels of hepatic lipase activity, buoyant LDL, and higher HDL2 cholesterol. *Arterioscler Thromb Vasc Biol* 1998;18:1723-1729.

## 8.10 Bibliografía de SR-BI

- Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler Thromb Vasc Biol* 1999; 19:1734-1743.
- Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, Krieger M. Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* 1996; 271:518-520.
- Arai T, Wang N, Bezouevski M, Welch C, Tall AR. Decreased atherosclerosis in heterozygous low density lipoprotein receptor-deficient mice expressing the scavenger receptor BI transgene. *J Biol Chem* 1999; 274:2366-2371.
- Bektas A, Suprenant ME, Wogan LT, Plengvidhya N, Rich SS, Warram JH, Krolewski AS, Doria A. Evidence of a novel type 2 diabetes locus 50 cM centromeric to NIDDM2 on chromosome 12q. *Diabetes* 1996; 48:2246-2251.
- Braun A, Trigatti BL, Post MJ, Sato K, Simons M, Edelberg JM, Rosenberg RD, Schrenzel M, Krieger M. Loss of SR-BI expression leads to the early onset of occlusive atherosclerotic coronary artery disease, spontaneous myocardial infarctions, severe cardiac dysfunction, and premature death in apolipoprotein E-deficient mice. *Circ Res* 2002; 90:270-276.
- Cao G, Garcia CK, Wyne KL, Schultz RA, Parker KL, Hobbs HH. Structure and localization of the human gene encoding SR-BI/CLA-1. Evidence for transcriptional control by steroidogenic factor 1. *J Biol Chem* 1997; 272:33068-33076.
- Cupples LA, Gagnon DR, Kannel WB. Long- and short-term risk of sudden coronary death. *Circulation* 1992; 85:111-118.
- de Beer FC, Connell PM, Yu J, de Beer MC, Webb NR, van der Westhuyzen DR. HDL modification by secretory phospholipase A(2) promotes scavenger receptor class B type I interaction and accelerates HDL catabolism. *J Lipid Res* 2000; 41:1849-57.
- de Beer MC, Durbin DM, Cai L, Jonas A, de Beer FC, van der Westhuyzen DR. Apolipoprotein A-I conformation markedly influences HDL interaction with scavenger receptor BI. *J Lipid Res* 2001; 42:309-313.

- Durst R, Colombo R, Shpitzen S, Avi LB, Friedlander Y, Wexler R, Raal FJ, Marais DA, Defesche JC, Mandelshtam MY, Kotze MJ, Leitersdorf E, Meiner V. Recent origin and spread of a common Lithuanian mutation, G197del LDLR, causing familial hypercholesterolemia: positive selection is not always necessary to account for disease incidence among Ashkenazi Jews. *Am J Hum Genet* 2001; 68:1172-1188.
- Erkelens DW Insulin resistance syndrome and type 2 diabetes mellitus. *Am J Cardiol* 2001; 88:38-42.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975; 4:518-525.
- Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA* 2002; 287:356-359.
- Franceschini G. Epidemiologic evidence for high-density lipoprotein cholesterol as a risk factor for coronary artery disease. *Am J Cardiol* 2001; 88:9-13.
- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma without use of preparative ultracentrifuge. *Clin Chem* 1972; 18:499-502.
- Gillotte-Taylor K, Boullier A, Witztum JL, Steinberg D, Quehenberger O. Scavenger receptor class B type I as a receptor for oxidized low density lipoprotein. *J Lipid Res* 2001; 42:1474-1482.
- Goff DC Jr, D'Agostino RB Jr, Haffner SM, Saad MF, Wagenknecht LE. Lipoprotein concentrations and carotid atherosclerosis by diabetes status: results from the Insulin Resistance Atherosclerosis Study. *Diabetes Care* 2000; 23:1006-1011.
- Greene DJ, Skeggs JW, Morton RE. Elevated triglyceride content diminishes the capacity of high density lipoprotein to deliver cholesteryl esters via the scavenger receptor class B type I (SR-BI). *J Biol Chem* 2001; 276:4804-4811.
- Gu X, Trigatti B, Xu S, Acton S, Babitt J, Krieger M. The efficient cellular uptake of high density lipoprotein lipids via scavenger receptor class B type I requires not only receptor-mediated surface binding but also receptor-specific lipid transfer mediated by its extracellular domain. *J Biol Chem* 1998; 273:26338-26348.
- Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res.* 1990;31:545-548.
- Howard BV, Cowan LD, Go O, Welty TK, Robbins DC, Lee ET. Adverse effects of diabetes on multiple cardiovascular disease risk factors in women. The Strong Heart Study. *Diabetes Care* 1998; 21:1258-1265.
- Huszar D, Varban ML, Rinninger F, Feeley R, Arai T, Fairchild-Huntress V, Donovan MJ, Tall AR. Increased LDL cholesterol and atherosclerosis in LDL receptor-deficient mice with attenuated expression of scavenger receptor B1. *Arterioscler Thromb Vasc Biol* 2000; 20:1068-1073.

- Iwashima Y, Eto M, Hata A, Kaku K, Horiuchi S, Ushikubi F, Sano H. Advanced glycation end products-induced gene expression of scavenger receptors in cultured human monocyte-derived macrophages. *Biochem Biophys Res Commun* 2000; 277:368-80.
- Kannel WB, Feinlieb M, McNamara PM, Garrison NJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol* 1979; 110:281-290.
- Kozarsky KF, Donahee MH, Glick JM, Krieger M, Rader DJ. Gene transfer and hepatic overexpression of the HDL receptor SR-BI reduces atherosclerosis in the cholesterol-fed LDL receptor-deficient mouse. *Arterioscler. Thromb. Vasc. Biol* 2000; 20:721-727.
- Kozarsky KF, Donahee MH, Rigotti A, Iqbal SN, Edelman ER, Krieger M. Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels. *Nature* 1997; 387:414-417.
- Krieger M. Charting the fate of the "good cholesterol": Identification and characterization of the high-density lipoprotein receptor SR-BI. *Annu Rev Biochem* 1999; 68: 523-558.
- Krieger M. Scavenger receptor class B type I is a multiligand HDL receptor that influences diverse physiologic systems. *J Clin Invest* 2001; 108:793-797.
- Lewontin RC. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 1964; 49:49-67.
- Lindgren CM, Mahtani MM, Widen E, McCarthy MI, Daly MJ, Kirby A, Reeve MP, Kruglyak L, Parker A, Meyer J, Almgren P, Lehto M, Kanninen T, Tuomi T, Groop LC, Lander ES. Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am J Hum Genet* 2002; 70:509-516.
- Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, Brayer J, Bryant B, Chan G, Daly M, Forsblom C, Kanninen T, Kirby A, Kruglyak L, Munnely K, Parkkonen M, Reeve-Daly MP, Weaver A, Brettin T, Duyk G, Lander ES, Groop LC. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat Genet* 1996; 14:90-94.
- Meigs JB, Larson MG, D'Agostino RB, Levy D, Clouse ME, Nathan DM, Wilson PW, O'Donnell CJ 2002. Coronary artery calcification in type 2 diabetes and insulin resistance: the Framingham Offspring Study. *Diabetes Care* 25:1313-1319.
- Meigs JB, Ordovas JM, Cupples LA, Singer DE, Nathan DM, Schaefer EJ, Wilson PW. Apolipoprotein E isoform polymorphisms are not associated with insulin resistance: the Framingham Offspring Study. *Diabetes Care* 2000; 23:669-674.
- Milliat F, Gripois D, Blouquit M-E, Ferezou J, Serougne C, Fidge NH, Lutton C. Short and long-term effects of streptozotocin on dietary cholesterol absorption, plasma lipoproteins and liver lipoprotein receptors in RICO rats. *Exp Clin Endocrinol Diabetes* 2000; 108:436-46.

- Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. *Arterioscler Thromb Vasc Biol* 2000; 20:1323-1329.
- Ordovas JM, Peterson JP, Santaniello P, Cohn J, Wilson PWF, Schaefer EJ. Enzyme linked immunosorbent assay for human apolipoprotein B. *J Lipid Res* 1987; 28:1216.
- Osgood-McWeeney D, Galluzzi JR, Ordovas JM. Allelic discrimination for single nucleotide polymorphisms in the human scavenger receptor class B type 1 gene locus using fluorescent probes. *Clin Chem* 2000; 46:118-119.
- Otvos JD, Jeyarajah EJ, Bennett DW, Krauss RM. Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement. *Clin Chem* 1992; 38:1632-1638.
- Plat J, Mensink RP. Relationship of genetic variation in genes encoding apolipoprotein A-IV, scavenger receptor BI, HMG-CoA reductase, CETP and apolipoprotein E with cholesterol metabolism and the response to plant stanol ester consumption. *Eur J Clin Invest* 2002; 32:242-250.
- Rigotti A, Trigatti BL, Penman M, Rayburn H, Herz J, Krieger M. A targeted mutation in the murine gene encoding the high density lipoprotein (HDL) receptor scavenger receptor class B type I reveals its key role in HDL metabolism. *Proc Natl Acad Sci U S A* 1997; 94:12610-12615.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score Tests for Association between Traits and Haplotypes when Linkage Phase Is Ambiguous. *Am. J. Hum. Genet.* 2002; 70:425-434.
- Silver DL, Jiang XC, Arai T, Bruce C, Tall AR. Receptors and lipid transfer proteins in HDL metabolism. *Ann N Y Acad Sci* 2000; 902:103-111.
- Stein O, Stein Y. Atheroprotective mechanisms of HDL. *Atherosclerosis* 1999; 144:285-301.
- Tan KC, Ai VH, Chow WS, Chau MT, Leong L, Lam KS; Influence of low density lipoprotein (LDL) subfraction profile and LDL oxidation on endothelium-dependent and independent vasodilation in patients with type 2 diabetes. *J Clin Endocrinol Metab* 1999; 84:3212-3216.
- Taskinen MR. Diabetic dyslipidemia. *Atherosclerosis*; 2002; S 3:47-51.
- Trigatti BL, Rigotti A, Braun A. Cellular and physiological roles of SR-BI, a lipoprotein receptor which mediates selective lipid uptake. *Biochim Biophys Acta* 2000; 1529:276-86.
- Ueda Y, Gong E, Royer L, Cooper PN, Francone OL, Rubin EM. Relationship between expression levels and atherogenesis in scavenger receptor class B, type I transgenics. *J Biol Chem* 2000; 275:20368-20373.

- Ueda Y, Royer L, Gong E, Zhang J, Cooper PN, Francone O, Rubin EM. Lower plasma levels and accelerated clearance of high density lipoprotein (HDL) and non-HDL cholesterol in scavenger receptor class B type I transgenic mice. *J Biol Chem* 1999; 274:7165-7171.
- US adults: findings from the third National Health and Nutrition Examination Survey. *JAMA* 287:356-359.
- Varban ML, Rinninger F, Wang N, Fairchild-Huntress V, Dunmore JH, Fang Q, Gosselin ML, Dixon KL, Deeds JD, Acton SL, Tall AR, Huszar D. Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc Natl Acad Sci USA* 1998; 95:4619-4624.
- Vaxillaire M, Boccio V, Philippi A, Vigouroux C, Terwilliger J, Passa P, Beckmann JS, Velho G, Lathrop GM, Froguel P. A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nat Genet* 1995; 9:418-423.
- Wang N, Arai T, Ji Y, Rinninger F, Tall AR. Liver-specific overexpression of scavenger receptor BI decreases levels of very low density lipoprotein ApoB, low density lipoprotein ApoB, and high density lipoprotein in transgenic mice. *Biol Chem* 1998; 273:32920-32926.
- Xu S, Laccotripe M, Huang X, Rigotti A, Zannis VI, Krieger M. Apolipoproteins of HDL can directly mediate binding to the scavenger receptor SR-BI, an HDL receptor that mediates selective lipid uptake. *J Lipid Res* 1997;38:1289-1298.

## 8.11 Bibliografía de PLIN

- Antonarakis SE. Recommendations for a nomenclature system for human gene mutations. Nomenclature Working Group. *Hum Mutat* 1998; 11:1-3.
- Arner P. Genetic variance and lipolysis regulation: implications for obesity 2001 *Ann Med* 2001; 33:542-546.
- Arner P. Insulin resistance in type 2 diabetes: role of fatty acids. *Diabetes Metab Res Rev* 2002; 18:S5-S9.
- Blanchette-Mackie EJ, Dwyer NK, Barber T, Coxey RA, Takeda T, Rondinone CM, Theodorakis JL, Greenberg AS, Londos C. Perilipin is located on the surface layer of intracellular lipid droplets in adipocytes. *J Lipid Res* 1995; 36:1211-1226.
- Brasaemle DL, Rubin B, Harten IA, Gruia-Gray J, Kimmel AR, Londos C. Perilipin A increases triacylglycerol storage by decreasing the rate of triacylglycerol hydrolysis. *J Biol Chem* 2000; 275:38486-38493.

- Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Cortina S, Folch J, Gonzalez JI, Ordovas JM. Environmental factors modulate the effect of the APOE genetic polymorphism on plasma lipid concentrations: Ecogenetic studies in a Mediterranean Spanish population. *Metabolism* 2001; 50:936-944.
- Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Folch J, Ordovas JM. Associations of LPL and APOC3 gene polymorphisms on plasma lipids in a Mediterranean population: Interaction with tobacco smoking and the APOE locus. *J Lipid Res* 2002; 43:416-427.
- Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'connell P, Stern MP. A major susceptibility locus influencing plasma triglyceride concentrations is located on chromosome 15q in mexican americans. *Am J Hum Genet* 2000; 66:1237-1245.
- Frayn KN, Karpe F, Fielding BA, Macdonald IA, Coppack SW. Integrative physiology of human adipose tissue. *Int J Obes Relat Metab Disord* 2003; 27:875-888.
- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; 18:499-502.
- Garcia A, Sekowski A, Subramanian V, Brasaemle DL. The central domain is required to target and anchor perilipin A to lipid droplets. *J Biol Chem* 2003; 278:625-635.
- Gasteyger C, Tremblay A. Metabolic impact of body fat distribution. *J Endocrinol Invest* 2002; 25:876-883.
- Greenberg AS, Egan JJ, Wek SA, Garty NB, Blanchette-Mackie EJ, Londos C. Perilipin, A major hormonally regulated adipocyte-specific phosphoprotein associated with the periphery of lipid storage droplets. *J Biol Chem* 1991; 266:11341-11346.
- Greenberg AS, Egan JJ, Wek SA, Moos MC Jr, Londos C, Kimmel AR. Isolation of cDNAs for perilipins a and b: sequence and expression of lipid droplet-associated proteins of adipocytes. *Proc Natl Acad Sci USA* 1993; 90:12035-12039.
- Hellstrom L, Langin D, Reynisdottir S, Dauszats M, Arner P. Adipocyte lipolysis in normal weight subjects with obesity among first-degree relatives. *Diabetologia* 1996; 39:921-928.
- Hoffstedt J, Arner P, Schalling M, Pedersen NL, Sengul S, Ahlberg S, Iliadou A, Lavebratt C. A common hormone-sensitive lipase i6 gene polymorphism is associated with decreased human adipocyte lipolytic function. *Diabetes* 2001; 50:2410-2413.
- Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000; 26:163-175.



- Kolehmainen M, Vidal H, Ohisalo JJ, Pirinen E, Alhava E, Uusitupa MI. Hormone sensitive lipase expression and adipose tissue metabolism show gender difference in obese subjects after weight loss. *Int J Obes Relat Metab Disord* 2002; 26:6-16.
- Lavebratt C, Ryden M, Schalling M, Sengul S, Ahlberg S, Hoffstedt J. The hormone-sensitive lipase i6 gene polymorphism and body fat accumulation. *Eur J Clin Invest* 2002; 32:938-942.
- Lewis GF, Carpentier A, Adeli K, Giacca A. Disordered Fat Storage And Mobilization in the pathogenesis of insulin resistance and type 2 diabetes. *Endocr Rev* 2002; 23:201-229.
- Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, Hanson J, Khoury M, Lau J, O'Brien TR, Rothman N, Stroup D, Taioli E, Thomas D, Vainio H, Wacholder S, Weinberg C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002; 156:300-310.
- Lofgren P, Hoffstedt J, Ryden M, Thorne A, Holm C, Wahrenberg H, Arner P. Major gender differences in the lipolytic capacity of abdominal subcutaneous fat cells in obesity observed before and after long-term weight reduction. *J Clin Endocrinol Metab* 2002; 87:764-771.
- Londos C, Brasaemle DL, Gruia-Gray J, Servetnick DA, Schultz CJ, Levin DM, Kimmel AR. Perilipin: Unique proteins associated with intracellular neutral lipid droplets in adipocytes and steroidogenic cells. *Biochem Soc Trans* 1995; 23:611-615.
- Londos C, Gruia-Gray J, Brasaemle DL, Rondinone CM, Takeda T, Dwyer NK, Barber T, Kimmel AR, Blanchette-Mackie EJ. Perilipin: Possible roles in structure and metabolism of intracellular neutral lipids in adipocytes and steroidogenic cells. *Int J Obes Relat Metab Disord* 1996; 20 S3:S97-S101.
- Lu X, Gruia-Gray J, Copeland NG, Gilbert DJ, Jenkins NA, Londos C, Kimmel AR. The murine perilipin gene: The lipid droplet-associated perilipins derive from tissue-specific, mRNA splice variants and define a gene family of ancient origin. *Mamm Genome* 2001; 12:741-749.
- Martinez-Botas J, Anderson JB, Tessier D, Lapillonne A, Chang BH, Quast MJ, Gorenstein D, Chen KH, Chan L. Absence of perilipin results in leanness and reverses obesity in *lepr(db/db)* mice. *Nat Genet* 2000; 26:474-479.
- Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoecur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P. Genome-wide search for type 2 diabetes in japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* 2002; 51:1247-1255.
- Mottagui-Tabar S, Ryden M, Lofgren P, Faulds G, Hoffstedt J, Brookes AJ, Andersson I, Arner P. Evidence for an important role of perilipin in the regulation of human adipocyte lipolysis. *Diabetologia* 2003; 46:789-797.
- Nielsen S, Guo Z, Albu JB, Klein S, O'Brien PC, Jensen MD. Energy Expenditure, sex, and endogenous fuel availability in humans. *J Clin Invest* 2003; 111:981-988.

- Nishiu J, Tanaka T, Nakamura Y. Isolation and chromosomal mapping of the human homolog of perilipin (plin), A rat adipose tissue-specific gene, by differential display method. *Genomics* 1998; 48:254-257.
- Schiffelers SL, Akkermans JA, Saris WH, Blaak EE. Lipolytic and nutritive blood flow response to beta-adrenoceptor stimulation in situ in subcutaneous abdominal adipose tissue in obese men. *Int J Obes Relat Metab Disord* 2003; 27:227-231.
- Servetnick DA, Brasaemle DL, Gruia-Gray J, Kimmel AR, Wolff J, Londos C. Perilipins are associated with cholesteryl ester droplets in steroidogenic adrenal cortical and leydig cells. *J Biol Chem* 1995; 270:16970-16973.
- Sorli JV, Velert R, Guillen M, Portoles O, Ramirez JV, Iborra J, Corella D. Effects of the apolipoprotein E polymorphism on plasma lipid levels and cardiovascular disease risk in a Mediterranean population. *Med Clin* 2002; 118:569-574.
- Souza SC, De Vargas LM, Yamamoto MT, Lien P, Franciosa MD, Moss LG, Greenberg AS. Overexpression of perilipin A and B blocks the ability of tumor necrosis factor alpha to increase lipolysis in 3T3-L1 adipocytes. *J Biol Chem* 1998; 273:24665-24669.
- Sztalryd C, Xu G, Dorward H, Tansey JT, Contreras JA, Kimmel AR, Londos C. Perilipin A is essential for the translocation of hormone-sensitive lipase during lipolytic activation. *J Cell Biol* 2002; 161:1093-1103.
- Tansey JT, Huml AM, Vogt R, Davis KE, Jones JM, Fraser KA, Brasaemle DL, Kimmel AR, Londos C. Functional studies on native and mutated forms of perilipins: A role in protein kinase A-mediated lipolysis of triacylglycerols in CHO cells. *J Biol Chem* 2003; 278:8401-8406.
- Tansey JT, Sztalryd C, Gruia-Gray J, Roush,DL, Zee JV, Gavrilova O, Reitman ML, Deng CX, Li C, Kimmel AR, Londos C. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proc Natl Acad Sci USA* 2001; 98:6494-6499.
- Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992; 135:1019-1028.

## 8.12 Bibliografía de resultados de los proyectos bioinformáticos

### 8.12.1 Bibliografía sobre comunicación de instrumentos: PLATEX

- Applied Biosystems. ABI PRISM® 3100 Genetic Analyzer. User's Manual 2001. Applied Biosystems at <http://docs.appliedbiosystems.com/genindex.taf> (access data 10/31/2002).

### 8.12.2 Bibliografía sobre visualización de secuencias: seqpacker

- Booch G., Rumbaugh J., Jacobson I. The Unified Modeling Language User Guide. Addison-Wesley, Reading, MA, 1999.
- Coltell O, Arregui M., Chalmeta R., Corella D., Parnell L., Ordovás JM. "SeqPacker: a nucleotide sequence manipulation utility". Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE2004) May 19-21, 2004, Taichung, Taiwan, ROC (<http://bibe2004.ece.uci.edu/>).
- Coltell O, Arregui M., Chalmeta R., Corella D., Parnell L., Ordovás JM. "A Biologist-Friendly User Interface to Manipulate Nucleotide Sequences in Genomic Epidemiology". 2nd International Workshop on Biological Data Management (BIDM'04). 1 - 3 September, 2004. Zaragoza, Spain.
- Fischer, C., Schweigert, S., Spreckelsen, C., Vogel, F. "Programs, databases, and expert systems for hu-man geneticists-a survey". Hum. Genet. (1997), 97:129-137.
- GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.
- GenBank-ftp. [ftp://ftp.ncbi.nih.gov/genomes/H\\_sapiens/](ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/). Accedido el 4 de junio de 2004.
- <http://iubio.bio.indiana.edu/soft/molbio/mac/plasmid-maker.html>. Accedido el 4 de junio de 2004.
- <http://www.accessexcellence.org/AE/AEC/CC/restriction.html>. Accedido el 4 de junio de 2004.
- <http://www.arches.uga.edu/~lace52/rplasmid.html>. Accedido el 4 de junio de 2004.
- OMG home page. <http://www.omg.org>. Accedido el 4 de junio de 2004.
- Ordovas, J.M., Corella, D., Cupples, L.A., Demissie, S., Kelleher, A., Coltell, O., Wilson, P.W., Schaefer, E.J., Tucker, K. "Polyunsaturated fatty acids modulate the effects of the APOA1 G-A polymorphism on HDL-cholesterol concentrations in a sex-specific manner: the Framingham Study". Am J Clin Nutr., Jan;75(1); 2002: 38-46."
- Rumbaugh J., Jacobson I., Booch G. El Lenguaje Unificado de Modelado. Manual de Referencia. Addison-Wesley, Madrid, 2000.

### 8.12.3 Bibliografía sobre búsqueda de polimorfismos: PREDICARD

- Christian Blaschke, Alfonso Valencia. The Frame-Based Module of the SUISEKI Information Extraction System. IEEE Intelligent Systems. March 2002: 14-20.
- F. Olken. Bioinformatics Databases 1. 18th International Conference on Data Engineering (ICDE'02). February 2002: 649.

Lada A, Adamic, Dennis Wilkinson, Bernardo A. Huberman, Eytan Adar. A Literature Based Method for Identifying Gene-Disease Connections. IEEE Computer Society Bioinformatics Conference (CSB'02). August 2002: 109.

NCBI. The NCBI Handbook (electronic book). (2003). <http://www.ncbi.nlm.nih.gov/>.

#### 8.12.4 Bibliografía sobre genes relacionados con las ECV

Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler Thromb Vasc Biol.* 1999; 19:1734-43.

Agerholm-Larsen B, Tybjaerg-Hansen A, Schnohr P, Steffensen R, Nordestgaard BG. Common cholesteryl ester transfer protein mutations, decreased HDL cholesterol, and possible decreased risk of ischemic heart disease: The Copenhagen City Heart Study. *Circulation.* 2000; 102:2197-203.

Andersen RV, Wittrup HH, Tybjaerg-Hansen A, Steffensen R, Schnohr P, Nordestgaard BG. Hepatic lipase mutations, elevated high-density lipoprotein cholesterol, and increased risk of ischemic heart disease: the Copenhagen City Heart Study. *J Am Coll Cardiol.* 2003; 41:1972-82.

Blankenberg S, Rupprecht HJ, Bickel C, Jiang XC, Poirier O, Lackner KJ, Meyer J, Cambien F, Tiret L; AtheroGene Investigators. Common genetic variation of the cholesteryl ester transfer protein gene strongly predicts future cardiovascular death in patients with coronary artery disease. *J Am Coll Cardiol.* 2003; 41:1983-9.

Clee SM, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, Jukema JW, van Wijland M, van Dam M, Hudson TJ, Brooks-Wilson A, Genest J Jr, Kastelein JJ, Hayden MR. Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation.* 2001; 103:1198-205.

Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Folch J, Ordovas JM. Associations of LPL and APOC3 gene polymorphisms on plasma lipids in a Mediterranean population: interaction with tobacco smoking and the APOE locus. *J Lipid Res.* 2002; 43:416-27.

Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet.* 1998; 20:284-7.

Deeb SS, Zambon A, Carr MC, Ayyobi AF, Brunzell JD. Hepatic lipase and dyslipidemia: interactions among genetic variants, obesity, gender, and diet. *J Lipid Res.* 2003; 44:1279-86.

- Flavell DM, Jamshidi Y, Hawe E, Pineda Torra I, Taskinen MR, Frick MH, Nieminen MS, Kesaniemi YA, Pasternack A, Staels B, Miller G, Humphries SE, Talmud PJ, Syvanne M. Peroxisome proliferator-activated receptor alpha gene variants influence progression of coronary atherosclerosis and risk of coronary artery disease. *Circulation*. 2002; 105:1440-5.
- Gagne SE, Larson MG, Pimstone SN, Schaefer EJ, Kastelein JJ, Wilson PW, Ordovas JM, Hayden MR. A common truncation variant of lipoprotein lipase (Ser447X) confers protection against coronary heart disease: the Framingham Offspring Study. *Clin Genet*. 1999; 55:450-4.
- Herrington DM, Howard TD, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Brosnihan KB, Meyers DA, Bleecker ER. Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease.[comment]. [Clinical Trial. Journal Article] *New England Journal of Medicine*; 2002; 346:967-74, 2002
- Hines L. M., Stampfer M. J., Ma J., Gaziano J. M., Ridker P. M., Hankinson S. E., Sacks F., Rimm E. B., Hunter D. J. N. Genetic Variation in Alcohol Dehydrogenase and the Beneficial Effect of Moderate Alcohol Consumption on Myocardial Infarction *Engl J Med* 2001; 344:549-555.
- Kastelein JJ, Ordovas JM, Wittekoek ME, Pimstone SN, Wilson WF, Gagne SE, Larson MG, Schaefer EJ, Boer JM, Gerdes C, Hayden MR. Two common mutations (D9N, N291S) in lipoprotein lipase: a cumulative analysis of their influence on plasma lipids and lipoproteins in men and women. *Clin Genet*. 1999; 56:297-305.
- Martin S, Nicaud V, Humphries SE, Talmud PJ; EARS group. Contribution of APOA5 gene variants to plasma triglyceride determination and to the response to both fat and glucose tolerance challenges. *Biochim Biophys Acta*. 2003;1637:217-25.
- Mata P, Ordovas JM, Lopez-Miranda J, Lichtenstein AH, Clevidence B, Judd JT, Schaefer EJ. ApoA-IV phenotype affects diet-induced plasma LDL cholesterol lowering. *Arterioscler Thromb*. 1994; 14:884-91.
- McCarthy JJ, Lehner T, Reeves C, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ; Genequest investigators. Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. *J Med Genet*. 2003; 40:453-8.
- Olivieri O, Stranieri C, Bassi A, Zaia B, Girelli D, Pizzolo F, Trabetti E, Cheng S, Grow MA, Pignatti PF, Corrocher R. ApoC-III gene polymorphisms and risk of coronary artery disease. *J Lipid Res*. 2002;43:1450-7.
- Ordovas JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PW, Schaefer EJ, Tucker K. Polyunsaturated fatty acids modulate the effects of the APOA1 G-A polymorphism on HDL Cholesterol concentrations in a sex-specific manner: the Framingham Study. *Am J Clin Nutr*. 2002; 75:38-46.

- Ordovas JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PW, Schaefer EJ, Tucker KL. Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study. *Circulation*. 2002; 106:2315-21.
- Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. *Arterioscler Thromb Vasc Biol*. 2000; 20:1323-9.
- Ordovas JM, Mooser V. The APOE locus and the pharmacogenetics of lipid response. *Curr Opin Lipidol*. 2002; 13:113-7. Review.
- Osgood D, Corella D, Demissie S, Cupples LA, Wilson PW, Meigs JB, Schaefer EJ, Coltell O, Ordovas JM. Genetic variation at the scavenger receptor class B type I gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: the Framingham study. *J Clin Endocrinol Metab*. 2003; 88:2869-79.
- Tai ES, Demissie S, Cupples LA, Corella D, Wilson PW, Schaefer EJ, Ordovas JM. Association between the PPARA L162V polymorphism and plasma lipid levels: the Framingham Offspring Study. *Arterioscler Thromb Vasc Biol*. 2002; 22:805-10.
- Talmud PJ, Hawe E, Martin S, Olivier M, Miller GJ, Rubin EM, Pennacchio LA, Humphries SE. Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. *Hum Mol Genet*. 2002; 11:3039-46.
- Zambon A, Deeb SS, Pauletto P, Crepaldi G, Brunzell JD. Hepatic lipase: a marker for cardiovascular disease risk and response to therapy. *Curr Opin Lipidol*. 2003;14:179-89.

### 8.12.5 Bibliografía sobre compresión de secuencias

- Applied Biosystems home page. <http://www.appliedbiosystems.com/>. Accedido el 4 de junio de 2004.
- Chromas home page. <http://www.technelysium.com.au/chromas.html>. Accedido el 4 de junio de 2004.
- FASTA format Description web page. <http://ngfnblast.gbf.de/docs/fasta.html>. Accedido el 4 de junio de 2004.
- GenBank Home page. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>. Accedido el 4 de junio de 2004.

## 8.13 Bibliografía de auditoría y control en Bioinformática

- Benal R., Coltell O. Auditoría de los Sistemas de Información (reimpresión). Servicio de Publicaciones de la Universidad Politécnica de Valencia, Valencia, 1999.
- Cass S., Riezenmann M.J. "Improving Security, Preserving Privacy". IEEE Spectrum, Jan.; 2002: 44-49.
- Coltell O., Chalmeta R. "Auditoría Bioinformática". Actas del V Congreso Nacional de Informática y Salud, INFORSALUD 2002. Madrid, 2002.
- Cornell M., Paton N.W., Wu S., Goble C.A., Miller C.J., Kirby P., Eilbeck K., Brass A., Hayes A., Oliver S.G. "GIMS-A Data Warehouse for Storage and Analysis of Genome Sequence and Functional Data". EBI, the European Bioinformatics Institute (EMBL Outstation, Hinxton, UK), <http://www.ebi.ac.uk/> (accedido 6/02/2002).
- Ermolaeva O., Rastogi M., Pret K.D., Schuler G.D., Bittner M.L., Chen Y., Simon R., Meltzer P., Trent J.M., Boguski M.S. "Data management and analysis for gene expression arrays". Nature genetics, 20; 1998: 19-23.
- Hass L.M., Kodali P., Rice J.E. "Integrating Life Sciences Data – With a Little Garlic". Proc. IEEE Int. Symp. on Bio-Informatics and Biomedical Engineering. IEEE; 2000: 5-12.
- ISACA home page. Information Systems Audit and Control Association. <http://www.isaca.org>. Accedido el 4 de junio de 2004.
- ISACAF-B. COBIT. Framework. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- ISACAF-D. COBIT. Control Objectives. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- ISACAF-E. COBIT. Audit Guidelines. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- ISACAF-F. COBIT. Implementation Tool Set. 3rd ed. ISACA, Rolling Meadows, IL (USA), 2000.
- Piattini M., Del Peso E. (eds.) Auditoría Informática. Un enfoque práctico. Ra-Ma, Madrid, 1998.
- Rindfleisch T.C. "Privacy, Information Technology, and Health Care". Communications of the ACM, 40-8; 1997: 93-100.
- Rondel R. K., Varley S. A., Webb C. (eds.) Clinical Data Management. John Wiley, New York, 1993.
- Sackman H. Biomedical Information Technology. Global Social Responsibilities for the Democratic Age. Academic Press, San Diego, CA (USA), 1997.
- Science's News staff. "A History of the Human Genome Project". Science Magazine; 291 (5507); 2001: 1195-1261.

---

Shrotliffe E.H., Blois M.S. "The Computer Meets Medicine and Biology: Emergence of a Discipline". Computer Applications in Health Care and Biomedicine; <http://smi-web.stanford.edu/textbook/ChapterOne.htm> (accedido el 22/01/2002).

Van Bemmel J.H., Musen M.A.(eds.) Hadbook of Medical Informatics. Springer-Verlag, Heidelberg, 1997.





En el contenido de esta tesis hay un gran número de conceptos provenientes de las distintas disciplinas tratadas que no se han definido explícitamente o cuya definición está diluida en el desarrollo de las secciones. A continuación se incluye un pequeño glosario con los términos considerados más importantes.

**ADN:**

La molécula en forma de doble hélice, portadora de los genes y formada por subunidades nucleotídicas. Es la molécula hereditaria principal en la mayoría de las especies.

**Alelo dominante:**

Un alelo cuyo efecto fenotípico está expresado teniendo en cuenta si el organismo es homocigoto o heterocigoto para dicho alelo.

**Alelo recesivo:**

Un alelo cuyo efecto fenotípico está expresado solamente en el tipo homocigoto.

**Alelo:**

Una de las dos o más formas alternativas de un gen en un locus determinado, que da lugar características hereditarias alternativas.

**Alelos codominantes:**

Un par de alelos alternativos que en conjunto están expresados completamente en los heterocigotos.

**Análisis de asociación:**

Un método de análisis genético que compara la frecuencia de los alelos entre individuos afectados y no afectados. Un alelo dado se considera que está asociado con la

enfermedad si dicho alelo se da con una alta frecuencia estadísticamente significativa entre individuos afectados.

**Análisis de ligamiento:**

Un método estadístico para la detección de ligamientos entre distintos loci usando estructuras familiares (individuos relacionados por lazos familiares).

**Animal transgénico:**

Un animal cuyo genoma se ha modificado para que contenga nuevo ADN no propio e introducido artificialmente.

**Anticipación:**

La ocurrencia de un rasgo heredado que incrementa progresivamente su severidad (se manifiesta como síntomas más severos, mayor ocurrencia de riesgo, y/o menor edad en su aparición) sobre sucesivas generaciones.

**ARN mensajero (mRNA):**

Una molécula de ARN que se transcribe desde la secuencia de ADN en el núcleo y que funciona en la traslación en el citoplasma de una secuencia específico de aminoácido de un polipéptido naciente.

**ARN:**

Tipo de ácido nucleico que se parece mucho al ADN, pero que tiene ribosa en lugar de desoxirribosa, y uracilo en lugar de timina.

**Autosomas:**

Todos los cromosomas excepto los cromosomas sexuales. En los humanos, todos los cromosomas excepto los cromosomas X e Y.

**Codon:**

En la molécula de ARN mensajero, una secuencia de tres bases que codifica un aminoácido en particular o una señal de Sto. (una señal que le dice al mecanismo de generación de proteínas que pare de producirlas).

**Confiabilidad de un sistema:**

Es el grado de cumplimiento satisfactorio de las características siguientes: (1) Corrección de su diseño. El diseño está realizado según unos requisitos expresados en consonancia con las necesidades de los usuarios y para resolver el problema planteado. (2) Correcta correspondencia entre el diseño y su aplicación: la construcción resultante debe reflejar fielmente el diseño previo. (3) Fiabilidad de sus componentes: la tasa de

fiabilidad de sus componentes debe ser la mayor posible de forma que la tasa de fiabilidad conjunta sea de un nivel parecido. (4) Regida por el tiempo que tardan los componentes en desgastarse: la calidad de los componentes debe ser tal que alargue en lo posible la vida útil del sistema.

**Dato:**

Registro de un hecho (almacenamiento de un suceso), en cualquier tipo de soporte.

**Desequilibrio de ligamiento:**

Una situación en la que la frecuencia de un cierto haplotipo en una población no es igual al producto de sus respectivas frecuencias alélicas. También se refiere a la asociación entre alelos en diferentes loci en la población.

**Enfoque de ingeniería:**

El enfoque de ingeniería se basa en la aplicación rigurosa y metódica de técnicas y tecnologías suficientemente probadas dentro de un ámbito teórico y práctico para obtener un resultado de calidad.

**Enzimas de restricción:**

Enzimas que reconocen secuencias de nucleótidos cortas específicas y cortan el ADN entre los sitios de dichas secuencias y en su vecindad.

**Equilibrio de Hardy-Weinberg:**

Dados los alelos "A" y "a" en la población con frecuencias  $p$  para **A** y  $q$  para **a**, se trata del fenómeno que, en ausencia de mutación, migración, selección natural, o derivación genética, y bajo apareamientos aleatorios, la distribución de frecuencias de los genotipos **AA**, **Aa** y **aa** permanece constante en  $p^2$ ,  $2pq$  y  $q^2$  respectivamente.

**Estado:**

Es el conjunto de valores de las propiedades que caracterizan una entidad que tenga existencia (objetiva en el caso de objetos, y subjetiva en el caso de conceptos) durante un intervalo de tiempo dado.

**Exón:**

Secuencia de ADN que se traduce en una proteína.

**Factor:**

Características que influyen directa o indirectamente sobre el CHDR.

**Fenotipo:**

Las características perceptibles asociadas a un genotipo particular.

**Gen:**

La unidad física y funcional de la herencia, que se transmite de una generación a la siguiente, y puede transcribirse en un polipéptido o proteína.

**Genoma:**

La dotación genética de un organismo o individuo.

**Genotipo:**

La constitución genética de un individuo que subyace a un rasgo o constelación de rasgos específicos.

**Haplotipo:**

combinación particular de alelos que se encuentran en una posición particular de un cromosoma.

**Herramientas:**

Es el conjunto de los elementos que, mediante la estructuración, clasificación y automatización de determinados procedimientos de ingeniería y diseño, facilitan el trabajo del ingeniero al descargarle de tareas rutinarias, repetitivas o extremadamente exhaustivas, y le permiten centrarse en aspectos cualitativos o fundamentales. Suministran en resumen un soporte automático o semiautomático a los métodos.

**Heterocigoto:**

Un individuo que tiene diferentes alelos de un gen particular.

An individual having different alleles of a particular gene.

**Heterogeneidad alélica (heterogeneidad intralocus):**

Una forma de expresión genética en la que distintos alelos mutantes conducen al mismo fenotipo de enfermedad en el mismo locus.

**Heterogeneidad genética:**

Una situación en la cual distintos alelos mutantes conducen al mismo fenotipo de enfermedad.

**Homocigoto:**

Un individuo que tiene idénticos alelos de un gen particular.

**Impronta genómica:**

Expresión diferencial de un gen dependiendo del sexo del pariente transmisor.

**Ingeniería de las Tecnologías de Información:**

Es la disciplina que tiene como meta la definición de arquitecturas que permitan a las empresas emplear la información y las tecnologías asociadas eficazmente. En consecuencia, trabaja en la creación de un plan global para implementar dichas arquitecturas.

**Ingeniería de las Tecnologías de la Información y la Comunicación:**

Es el conjunto de ingenierías cuyo propósito es el desarrollo y la aplicación de las distintas tecnologías de información y comunicación relacionadas con los sistemas de información.

**Ingeniería del Software (como enfoque de ingeniería):**

Es el establecimiento y uso de principios de ingeniería robustos, orientados a obtener software económico que sea fiable y funcione de manera eficiente sobre máquinas reales, mediante la aplicación de los elementos y actividades siguientes: Métodos; Planificación y estimación de proyectos; Análisis de los requisitos del sistema y del software; Diseño de estructuras de datos; Arquitectura de programas y procedimientos algorítmicos; Codificación; Pruebas; Instalación y Mantenimiento; Herramientas; y Procedimientos.

**Ingeniería del Software (como proceso de modelado):**

Es una subdisciplina de la Ingeniería de Sistemas de Información. Es un proceso de modelado para definir los procesos que satisfagan las necesidades de la vista global, del elemento o detallada; para representar el comportamiento de los procesos y los supuestos en los que se basa el comportamiento, para definir explícitamente las entradas exógenas y endógenas de información al modelo, y representar todas las uniones que permitan al ingeniero entender mejor la visión correspondiente.

**Ingeniería:**

La ingeniería es una disciplina del conocimiento humano que busca y aplica soluciones técnicas y tecnológicas a problemas reales, a partir de conocimientos y planteamientos científicos.

**Intrón:**

Secuencia de ADN no codificadora que se encuentra dentro de un gen y que se transcribe inicialmente en ARN mensajero, pero que más tarde se elimina de este.

**Ligamiento:**

La tendencia de los genes a heredarse juntos como resultado de su localización en el mismo cromosoma.

**Loci:**

El plural de Locus siguiendo las reglas de formación de plurales en latín. El conjunto de posiciones específicas de un gen en un cromosoma.

**Locus del gen:**

La localización específica del gen en el cromosoma.

**Locus:**

Posición específica de un gen en un cromosoma. Ver Locus de gen.

**Mapa de transcripción:**

Un mapa genómico que ofrece información sobre la localización física de las secuencias de ADN expresadas (marcas o tags de secuencias expresadas), así como de la secuencia parcial de una región expresada.

**Metodologías:**

Son los sistemas estructurados y organizados de principios, reglas y prácticas que se aplican a ramas del conocimiento específicas.

**Métodos:**

Es el conjunto de principios, reglas y prácticas que suministran la forma de construir técnicamente (el cómo) el software. Los métodos están implicados en las categorías de actividades del desarrollo del software siguientes: Planificación y estimación de proyectos; Análisis de los requerimientos del sistema y del software; Diseño de estructuras de datos, arquitectura de programas y procedimientos algorítmicos; Generación de código; Pruebas y comprobaciones; e Implantación y mantenimiento.

**Modelo:**

Es una abstracción de la realidad o de un sistema real tomando los elementos más representativos con un propósito determinado. Por lo tanto, de un mismo sistema puede haber más de un modelo, porque, según el propósito del mismo, los elementos

representativos pueden ser distintos. Los elementos a considerar en la construcción de modelos son los siguientes: (1) Los *supuestos* son elementos para la construcción de modelos que reducen el número de permutaciones y variaciones posibles, permitiendo al modelo reflejar el problema de manera razonable. (2) Las *simplificaciones* son elementos para la construcción de modelos que permiten crear el modelo a tiempo. (3) Las *limitaciones* o *restricciones* son elementos para la construcción de modelos que ayudan a delimitar el problema. (4) Las *preferencias* son elementos para la construcción de modelos que indican la arquitectura preferida para toda la información, funciones y tecnología; conflictos con otros factores restrictivos. Es recomendable tenerlas en cuenta para obtener un resultado aceptado, además de correcto.

**Modelos animales knockout:**

Animales que se han modificado genéticamente, mediante ingeniería genética, para que contengan una mutación nula (una mutación que resulta en la ausencia completa de un producto genético funcional) de un gen de interés. Otra forma de decirlo es que son animales genéticamente modificados para que expresen el gen sin la mutación específica.

**Multifactorial:**

Un rasgo que está influido por múltiples genes y/o múltiples factores ambientales.

**Mutación de eliminación o borrado:**

Una mutación que resulta de la eliminación de un segmento de ADN en un gen. El segmento puede estar formado por una sola base o por varias.

**Mutación de splicing:**

Una mutación que interrumpe la formación de un producto de gen funcional mediante la disrupción del adecuado splicing de dicho gen.

**Mutación:**

Un cambio en el material genético que se traduce como cambio heredable en una molécula de ADN.

**Mutagénesis:**

Cualquier proceso que lleve a cambios en el material genético.

**Paradigma:**

Es en general una forma distinta de ver las cosas. En la Ingeniería del Software, es un enfoque de ingeniería cuyas técnicas y tecnologías corresponden a una filosofía y a un ámbito particular en la representación de sistemas y de sus soluciones tecnológicas. Los más importantes paradigmas, centrados en el producto final, son el Paradigma de la



Programación Estructurada, el Paradigma de la Programación Lógica, el Paradigma de la Programación Funcional y el Paradigma de la Programación Orientada a Objetos.

**Par de bases (pb):**

Una medida que se refiere al par de nucleótidos que están ligados en las hebras complementarias del ADN (también del ARN). Por facilidad operativa, se considera también la posición que ocupa un nucleótido en una cadena de ADN (de una de las hebras) y se utiliza como medida del tamaño de las secuencias de ADN y de ARN.

**Pares de kilobases (kb):**

Un millar de pares de bases. Es un múltiplo muy utilizado en las medidas de genes cortos y fragmentos.

**Pares de megabases (Mb):**

Un millón de pares de bases. Es un múltiplo muy utilizado en las medidas de genes largos, cromosomas y genomas.

**Penetrancia dependiente de la edad:**

La dependencia de la manifestación de un genotipo a nivel de fenotipo en función de la edad del individuo.

**Penetrancia:**

Proporción de individuos con un genotipo específico que lo manifiestan a nivel fenotípico.

**Penetrancia:**

La proporción de individuos con un genotipo específico quienes manifiestan dicho genotipo a nivel fenotípico.

**Perturbación:**

Alteración introducida en un sistema.

**Población:**

Grupo local de organismos que pertenecen a la misma especie que pueden aparearse entre sí.

**Polimorfismo:**

Variación en la secuencia de AND entre regiones homólogas de ADN de individuos diferentes: al menos deben estar presentes en la población dos alelos con frecuencias mayores del 1%.

**Polimorfismos de longitud de fragmento de restricción (RFLP):**

Variaciones en la longitud de fragmentos de ADN específicos generados por enzimas de restricción que cortan moléculas de ADN en las secuencias específicas, pero son incapaces de cortar el ADN si el sitio de reconocimiento ha sido alterado por una mutación. Estas secuencias polimórficas se utilizan como marcadores genéticos en el análisis de ligamiento.

**Procedimientos:**

Es el conjunto de facilidades que integran métodos y herramientas en unidades metodológicas operativas. Entre otras cosas, estas unidades definen las secuencias de aplicación de los métodos; describen y establecen los resultados de la culminación de cada etapa de aplicación de los métodos, denominadas entregas (documentos, informes, diagramas, etc.); definen los controles para asegurar la calidad y gestionar los cambios; y establecen las directrices que ayudan a los gestores del software en la evaluación del progreso en el desarrollo.

**Proceso de ingeniería:**

Es la forma unitaria en que se desarrolla el enfoque de ingeniería. El conjunto de procesos de ingeniería constituyen el desarrollo global del enfoque de ingeniería.

**Proceso:**

Conjunto de actividades relacionadas entre sí a través de un flujo de información y materiales, y que tienen un objetivo común.

**Propósito de la ingeniería:**

El propósito de la ingeniería es la de resolver problemas reales de tipo técnico y tecnológico de la manera más eficaz (alcanzando los objetivos marcados) y eficiente (con el menor coste posible).

**Proyecto de ingeniería informática:**

Es el conjunto de actividades coordinadas cronológicamente para alcanzar un subconjunto de objetivos a partir de la definición de un subconjunto de necesidades, todo ello enmarcado dentro de un Plan Estratégico de Sistemas de Información. En el aspecto cronológico, un proyecto tiene un inicio, una fecha inicial, y un final, la fecha de finalización; pero también tiene unos hitos temporales intermedios para la distribución de las actividades. En el aspecto económico, un proyecto necesita de unos recursos, tanto para su puesta en marcha, como para su continuación y culminación. Además, el

proyecto debe ofrecer determinados resultados que justifiquen el coste de los recursos empleados. En el aspecto organizativo, un proyecto parte de una planificación previa y necesita de una gestión que adecue las actividades a la planificación y que controle las desviaciones.

**Proyecto de ingeniería:**

Es la unidad de acción de la ingeniería.

**Rasgo complejo:**

Un rasgo que tiene múltiples determinantes, que pueden ser genéticos, ambientales o ambos. Las ECV son un ejemplo típicos del tercer caso.

**Reacción en cadena de la polimerasa (PCR):**

Una técnica de Biología molecular que se usa para amplificar segmentos particulares de ADN para su detección y/o posterior manipulación.

**Señal de stop:**

Un codon que dice al mecanismo de producción de proteínas de la célula que pare de generar proteínas y que no incorpore más aminoácidos en la cadena polipeptídica en construcción.

**Software confiable:**

Es el software que cumple las características siguientes: Debe estar ajustado a las especificaciones establecidas previamente; no debe producir resultados incorrectos; no debe permitir su corrupción; debe reaccionar de manera útil y significativa en situaciones inesperadas; y sólo debe fallar por completo cuando sea imposible seguir avanzando.

**Software de calidad:**

Es el software que tiene los atributos siguientes: (1) Es *mantenible*: está escrito y documentado de forma que los cambios puedan ser realizados sin costes adicionales. (2) Es *fiable*: ha de funcionar según las expectativas de los clientes y no debería fallar más de lo indicado en las especificaciones; es eficiente: aprovecha al máximo los recursos hardware y de otro tipo del sistema que lo soporta y alcanza los objetivos establecidos. (3) Es *amigable*: tiene una apropiada interfaz de usuario que facilita la utilización del sistema o producto y en donde intervienen tanto elementos de visualización e intervención, como elementos de ayuda y asistencia. (4) Es *parcialmente sustituible*: la sustitución parcial de determinados elementos por otros con la misma interfaz no debe alterar el comportamiento general al menos en sentido negativo. (5) Es *parcialmente reutilizable*: determinados elementos deben poderse aplicar en nuevos desarrollos manteniendo como mínimo el mismo nivel de funcionalidad de cada uno de dichos elementos. (6) Es *controlable* y *auditable*: está escrito y documentado de forma que se pueda registrar los cambios sucesivos y se pueda registrar fiablemente su comportamiento para inspección posterior. (7) Es *seguro*: se ha diseñado para que

cumpla al máximo las premisas de asegurar la integridad, la confidencialidad y la disponibilidad.

**Software:**

El software es el conjunto de las categorías de elementos siguientes: (1) Programas de ordenador asociados con alguna aplicación o producto, junto con toda la información necesaria para instalar, usar, desarrollar y mantener estos programas. (2) Instrucciones de ordenador que, cuando se ejecutan, proporcionan la función y el comportamiento deseado. (3) Estructuras de datos que facilitan a los programas manipular adecuadamente la información. (4) Documentos que describen la operación y el uso de los programas y las estructuras de datos manipuladas por los mismos.

**Técnica:**

Mientras que la ciencia se dedica primordialmente al saber, la técnica es el arte de hacer. Entonces, la técnica es la aplicación de un conjunto de procedimientos derivados del saber científico o artístico, y de los recursos necesarios para poder producir resultados. La caracterización de una buena técnica se basa en los siguientes aspectos fundamentales: (1) *Utilidad*: los resultados obtenidos corresponden a determinados objetivos establecidos previamente antes de la aplicación de la técnica. (2) *Habilidad*: los procedimientos y recursos se usan con pericia o habilidad en casos semejantes. (3) *Reproducibilidad*: bajo las mismas condiciones y con los mismos elementos iniciales, los resultados obtenidos deben ser siempre los mismos. (4) *Perdurabilidad*: el conocimiento del proceso de aplicación de los procedimientos y recursos se puede guardar para recuperarlo en otro momento futuro. (5) *Transmisibilidad*: el conocimiento del proceso de aplicación de los procedimientos y recursos se puede enseñar y aprender. (6) *Eficiencia*: los resultados se obtienen con el menor coste de recursos y en el tiempo más corto.

**Tipo natural (o salvaje):**

El genotipo o fenotipo predominante que se encuentra en la naturaleza o en las muestras de laboratorio estándar para un organismo dado.



|   |    |
|---|----|
| Figura 1.1. La Bioinformática como convergencia multidisciplinar .....  | 30 |
| Figura 1.2. La Epidemiología Genética se adscribe en el contexto de las ciencias<br>médicas y es una parte de la epidemiología general, pero con una<br>contribución del Proyecto del Genoma Humano ..... | 32 |
| Figura 1.3. La penetrancia en las enfermedades genéticas .....  | 34 |
| Figura 1.4. Las interacciones de varios factores en las enfermedades cardiovasculares .....   | 35 |
| Figura 1.5. Estadísticas de mortalidad de las EVC en todo el mundo según datos de las<br>OMS en 1998 .....  | 37 |
| Figura 1.6. Los determinantes de la salud en ECV (adaptado de la propuesta general de<br>Laframboise (1973)) .....  | 38 |
| Figura 1.7. El esquema del que parte la Epidemiología Genómica en el estudio de las<br>ECV .....  | 39 |
| Figura 1.8. Estructura esquemática del gen CETP con sus polimorfismos .....   | 40 |
| Figura 1.9. Estructura esquemática del gen APOE y sus polimorfismos .....   | 42 |
| Figura 1.10. Disminución de Norte a Sur de la prevalencia del alelo $\epsilon 4$ en Europa .....  | 43 |
| Figura 1.11. Estructura esquemática del gen APOA (Fuente: GenBank NCBI Map<br>Viewer) .....   | 46 |
| Figura 1.12. Estructura esquemática del gen LIPC (Fuente: GenBank NCBI Map<br>Viewer) .....   | 48 |
| Figura 1.13. Estructura esquemática del gen SR-BI y situación en el cromosoma<br>(Fuente: GenBank NCBI Map Viewer) .....  | 49 |
| Figura 1.14. Estructura esquemática del gen PLIN (Fuente: GenBank NCBI Map<br>Viewer) .....   | 52 |
| Figura 1.15. Localización geográfica del Estudio Framingham .....   | 55 |
| Figura 1.16. Las matrices genómicas como fuentes de datos del modelo de riesgo .....  | 62 |
| Figura 1.17. Los análisis clínicos como fuentes de datos del modelo de riesgo .....   | 63 |

|  |    |
|--|----|
| Figura 1.18. Las exploraciones clínicas como fuentes de datos del modelo de riesgo.....  | 63 |
| Figura 1.19. Modelo de interacciones: matriz gen-gen. *: interacción por defecto de cada par de genes idénticos. #: interacción efectiva de cada par de genes distintos .....  | 65 |
| Figura 1.20. Modelo de interacciones: matriz gen-ambiente. #: interacción efectiva de cada par de factores distintos.....  | 66 |
| Figura 1.21. Modelo de interacciones: matriz ambiente-ambiente. *: interacción por defecto de cada par de factores idénticos. #: interacción efectiva de cada par de factores distintos .....  | 67 |
| Figura 1.22. Modelo de interacciones: combinación de matrices gen-gen y gen-ambiente .   | 68 |
| Figura 1.23. Modelo de interacciones: combinación de matrices gen-gen, gen-ambiente y ambiente-ambiente .....  | 69 |
| Figura 1.24. Arquitectura de La Medicina Genómica (Fuente: Fernando Martín, SemBioGen de Zaragoza 2003).....   | 70 |
| Figura 1.25. Arquitectura de integración de la Bioinformática como soporte de la Epidemiología Genómica .....  | 72 |
| Figura 1.26. Estructura de datos del registro de placa de fragmentos de ADN. En la columna “Well”, cada letra (A, B, C, etc.) designa uno de los pocillos de la placa. En la columna “Color Number”, se incluyen los colores del conjunto Dye (hasta 5). Las columnas “Sample Name” y “Color Info” tienen información inespecífica ..... | 75 |
| Figura 1.27. Estructura de datos de los fragmentos de ADN etiquetados químicamente que contiene la hoja de cálculo de la rejilla de PCR (Son 8 filas de 12 columnas, es decir, 96 muestras como máximo, aunque puede haber “faltas” representadas por “—”).....  | 75 |
| Figura 1.28. Estructura de datos del fichero de placa que es leído por los programas de secuenciación de ADN y análisis de polimorfismos. Las columnas “Sample Name” y “Color Info” tienen información relativa a las etiquetas de cada muestra. El valor “99999” indica “faltas” en los pocillos .....                                  | 76 |
| Figura 1.29. El proceso de conversión manual que produce los ficheros de datos (plate record) para la secuenciación y análisis de ADN .....  | 77 |
| Figura 1.30. Submodelo de objetos del formato ABI de los registros que almacenan las secuencias de nucleótido producidos por el algoritmo de lectura y asignación del secuenciador .....   | 80 |
| Figura 1.31. Estructura de la secuencia correspondiente a cada polimorfismo .....  | 83 |
| Figura 1.32. Estructura funcional del proyecto .....   | 84 |
| Figura 1.33. Condificación y almacenamiento ordinarios de las secuencias de bases .....  | 88 |
| Figura 1.34. Compactación y almacenamiento de las secuencias de bases por cada byte....  | 89 |

|   |     |
|---|-----|
| Figura 1.35. Compactación en cada byte de las secuencias de bases con tres bits por base.....   | 89  |
| Figura 3.1. Tamaños de partículas de VLDL, HDL y LDL. (A) Ejemplo en la ruta metabólica de la Lipasa Hepática (Fuente: Ordovás et al., 2002, Circulation). (B) Distintos tamaños y morfologías de las partículas HDL.....   | 124 |
| Figura 3.2. Obtención del ADN de los individuos y análisis de variaciones en epidemiología genómica.....  | 126 |
| Figura 3.3. Esquema del gen PLIN y localización de polimorfismos .....  | 131 |
| Figura 3.4. Estructura formal de la Ingeniería del Software.....  | 151 |
| Figura 3.5. Estructura formal de la Ingeniería del Software Bioinformático .....  | 152 |
| Figura 3.6. Interfaz principal de Rational Rose Enterprise .....  | 155 |
| Figura 3.7. El ciclo de vida del desarrollo del software (Fuente: Rational Corporation)...  | 157 |
| Figura 3.8. Recursos de TI, Objetivos de Negocio y Dominios de COBIT “Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute.” .....            | 164 |
| Figura 3.9. El cubo de COBIT. “Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute.” .....   | 165 |
| Figura 3.10. Objetivos de control de COBIT definidos genéricamente. “Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute.” .....             | 166 |
| Figura 3.11. Tabla resumen de COBIT. “Copyright 1996, 1998, 2000 Information Systems Audit and Control Foundation. Reprinted with the permission of the Information Systems Audit and Control Foundation and IT Governance Institute.” .....  | 167 |
| Figura 3.12. Arquitectura funcional y de responsabilidades del laboratorio .....  | 169 |
| Figura 4.1. El proceso de conversión manual que produce los ficheros de datos (plate record) para la secuenciación y análisis de ADN .....  | 179 |
| Figura 4.2. El proceso de conversión que obtiene los ficheros de datos para la secuenciación y análisis de ADN. Parte A: proceso manual. Parte B: proceso automático.....   | 180 |
| Figura 4.3. Interfaz de la hoja de cálculo que actúa de área de trabajo con un ejemplo de la tarea fundamental en que se asigna etiquetas de muestra a los conjuntos de muestras. Algunos parámetros se pueden modificar para permitir el trabajo con distintos ficheros y placas ..... | 180 |



|   |     |
|---|-----|
| Figura 4.4. Fichero de rejillas de PCR donde se puede ver la distribución irregular de las rejillas .....   | 181 |
| Figura 4.5. Representación habitual de la salida de un secuenciador ABI PRISM con el código de colores para las bases. Se denomina cromatograma.....  | 185 |
| Figura 4.6. Interfaz general de SeqPacker.....  | 186 |
| Figura 4.7. Un caso de búsqueda de una pequeña secuencia.....   | 187 |
| Figura 4.8. Flujo de trabajo de SeqPacker.....  | 188 |
| Figura 4.9. Flujo de trabajo para la visualización de secuencias.....   | 189 |
| Figure 4.10. El modelo de objetos de la interfaz gráfica .....  | 190 |
| Figure 4.11. El modelo de objetos de aplicación.....  | 190 |
| Figura 4.12. Esquema lógico de la base de datos .....   | 192 |
| Figura 4.13. Interfaz de la base de datos para la documentación de polimorfismos .....  | 201 |
| Figura 4.14. Composición de las secuencias de SNP en formato FASTA .....  | 203 |
| Figura 4.15. Interfaz de la base de datos para la documentación de polimorfismos .....  | 203 |
| Figura 4.16. Estructura interna del formato BSD.....  | 205 |
| Figura 4.17. Interfaz de la utilidad Squeezer2B .....   | 206 |
| Figura 4.18. Análisis de Sensibilidad. Coeficientes de regresión e intervalos de confianza al 95% para los genotipos B1B2 y B2B2 respectivamente, comparados con B1B1 cuando cada variable indicada se ha ido incluyendo progresivamente en los modelos de regresión lineal. Modelo 1: genotipo CETP. Modelo 2: modelo 1 + sexo. Modelo 3: modelo 2 + IMC. Modelo 4: modelo 3+ consumo de tabaco. Modelo 5: modelo 4 + consumo de alcohol. Modelo 6: modelo 5 + genotipo APOE. Se ha incluido $R^2$ en la figura para mostrar la variabilidad considerada para cada modelo de regresión ..... | 212 |
| Figura 4.19. Concentraciones de LDL-C medias según los genotipos de APOE y el consumo de alcohol en hombres (A) y en mujeres (B). Los valores de P se han obtenido en los tests ANOVA para la comparación de las medias entre genotipos por consumo de alcohol. Las barras de error muestran el error estándar de las medias .....  | 221 |
| Figura 4.20. Medias ajustadas de HDL-C (A) y apolipoproteína A-I (B) según el genotipo de APOA1 y el consumo de PUFA ( $\leq 4\%$ , $4-8\%$ y $\geq 8\%$ ) en mujeres. Las medias se han ajustado por edad, IMC, consumo de alcohol, tabaco, energía, consumo de SAFA, MUFA y PUFA. Los valores de P se han obtenido para el término de interacción entre el genotipo de APOA1 y PUFA en los modelos de regresión lineal multivariada adaptados como se ha indicado en los métodos.....   | 237 |

- Figura 4.21. Concentraciones medias de HDL-C según el polimorfismo de LIPC y las categorías de ingesta de grasa total. Se han ajustado las medias por relaciones familiares, sexo, edad, IMC, tabaco, alcohol, estrógenos, Betabloqueantes y energía. Las barras de error indican en error estándar de la media ..... 240
- Figura 4.22. Valores predichos (símbolos abiertos) del tamaño de HDL-C (A) y HDL según los genotipos de HL dependientes de la grasa total consumida (en continuo). Los valores predichos se han calculado a partir de los modelos de regresión que contienen la ingesta de grasa total, el polimorfismo de LIPC, su término de interacción y las variables de confusión potenciales. Los valores de P y R<sup>2</sup> se han obtenido de los modelos de regresión. Los símbolos sólidos representan la media de la variable dependiente por genotipo en las cuatro categorías de grasa total (valores medios: <20% [n = 166], 20 a 30% [n = 951], 30 a 40% [n = 882] y >= 40% [n = 125]) ..... 245
- Figura 4.23. Valores predichos (símbolos abiertos) de las concentraciones de gran HDL (A) y partículas intermedias y pequeñas de HDL (B) según los genotipos de HL dependientes de la cantidad de grasa animal consumida (en continuo). Los valores predichos se han calculado a partir de los modelos de regresión que contienen grasa animal, el polimorfismo de LIPC, su término de interacción, y las variables de confusión potenciales. Los valores de P y R<sup>2</sup> se han obtenido de los modelos. Los símbolos sólidos representan las medias de la variable dependiente para el genotipo y las cuatro categorías de grasa animal (valores medios: <10% [n = 303], 10 a 20% [n = 1419], 20 a 30% [n = 369] y >=30% [n = 39]; no se han encontrado sujetos TT en la categoría de máximo consumo de grasa ..... 247
- Figura 4.24. Análisis de Sensibilidad. Coeficientes de regresión e intervalos de confianza al 95% para los genotipos B1B2 y B2B2 respectivamente, comparados con B1B1 cuando cada variable indicada se ha ido incluyendo progresivamente en los modelos de regresión lineal. Modelo 1: genotipo CETP. Modelo 2: modelo 1 + sexo. Modelo 3: modelo 2 + IMC. Modelo 4: modelo 3+ consumo de tabaco. Modelo 5: modelo 4 + consumo de alcohol. Modelo 6: modelo 5 + genotipo APOE. Se ha incluido R<sup>2</sup> en la figura para mostrar la variabilidad considerada para cada modelo de regresión ..... 256
- Figura 4.25. : Concentraciones estimadas de HDL-C (A) y la tasa de gran HDL-C / pequeño HDL-C (B) según la diabetes dependiente de los haplotipos más comunes exón 1/ intrón 5/ exón 8 (1/1/1, 1/1/2, y 2/1/1, con una frecuencia de 0,37, 0,42 y 0,06, respectivamente) después de ajustar por edad, sexo, IMC, tabaco, consumo de alcohol, genotipo de apoE, uso de betabloqueantes, y estado menopáusico y terapia por estrógenos en las mujeres según el procedimiento GLM. .... 258
- Figura 4.26. Nomenclatura de los polimorfismos de PLIN. Las posiciones de los polimorfismos examinados se indican con líneas verticales cortas, con los nombres debajo. El cuadrado debajo del diagrama del gen muestra la secuencia que acompaña al nucleótido denominado “+1” en esta

nomenclatura. El A del codon iniciador ATG de la Metionina se indica en letra cursiva negrita, con la posición genómica en la secuencia de referencia (número de acceso de GenBank GI2143119 etiquetado debajo. Los aminoácidos correspondientes también están incluidos. El cuadrado con barras indica la región donde puede ocurrir el splicing alternativo. ....260

Figura 4.27. Medias de IMC dependientes de los genotipos combinados en el locus de PLIN en mujeres de la población general. Las medias ajustadas por edad de las combinaciones los SNP de PLIN1 y PLIN4 después de controlar por PLIN5 y PLIN6 (A), y combinaciones de lo SNP de PLIN5 y PLIN6 después de controlar por PLIN1 y PLIN4 (B). ....267

Figura 4.28. Fases de un proyecto de auditoría .....287

Figura 5.1. Replicación de AND con secuencias de plásmidos .....321

Figura 5.2. Nomenclaturas paralelas en aminoácidos .....323

Figura 5.3. Interfaz de la base de datos NCBI SNP .....324

Figura 12.1. Vistas de la arquitectura de un sistema .....427

Figura 12.2. Diagrama de clases .....430

Figura 12.3. Diagrama de objetos .....430

Figura 12.4. Diagrama de casos de uso .....431

Figura 12.5. Diagrama de secuencia de sucesos.....431

Figura 12.6. Diagrama de colaboración .....432

Figura 12.7. Diagrama de estados .....432

Figura 12.8. Diagrama de actividades .....433

Figura 12.9. Diagrama de componentes .....433

Figura 12.10. Diagrama de despliegue .....434

Figura 12.11. Especificación de una clase.....436

Figura 12.12. Ejemplos de adornos .....437

Figura 12.13. Ejemplo de dicotomía clase/objeto .....437

Figura 12.14. Ejemplo de dicotomía interfaz/implementación .....438

Figura 12.15. Estereotipo como mecanismo común en UML .....439

Figura 12.16. El ciclo de vida del desarrollo del software .....441

Figura 12.17. Las iteraciones son distintas en el ciclo de vida.....442

Figura 12.18. Vistas de la arquitectura de un sistema .....446

|  |     |
|--|-----|
| Tabla 1.1. Etiología de las enfermedades desde un punto de vista genético .....  | 33  |
| Tabla 1.2. Distribución por edad y sexo de los participantes del Framingham Offspring Study al inicio del estudio (1971).....                        | 56  |
| Tabla 1.3. Principales hitos alcanzados en la cohorte del Framingham Offspring Study (Fuente: Institutos Nacionales de Salud de Estados Unidos)..... | 57  |
| Tabla 1.4A. Elementos del Espacio del Problema de Auditoría y su relación con las Áreas y Problemas de Bioinformática .....                          | 92  |
| Tabla 1.4B. Elementos del Espacio del Problema de Auditoría y su relación con las Áreas y Problemas de Bioinformática (continuación).....            | 93  |
| Tabla 3.1. Secuencias de los primers y sondas del gen SR-BI (Fuente: Osgood et al., 2000).....   | 130 |
| Tabla 3.2. Descripción de los SNP, primers y sondas del gen PLIN .....   | 132 |
| Tabla 3.3. Vistas y diagramas de UML.....  | 155 |
| Tabla 3.4. Modelos y flujos de trabajo del Proceso Unificado.....  | 158 |
| Tabla 3.5. Genes y Polimorfismos Identificados. Proceso preliminar (parcial) .....   | 161 |
| Tabla 3.6. Dominios de COBIT .....   | 165 |
| Tabla 3.7. Estructura orgánica estable del Nutrition and Genomics Laboratory.....  | 168 |
| Tabla 3.8. Estructura de sistemas informáticos del Nutrition and Genomics Laboratory ..  | 170 |
| Tabla 4.1. Descripción de las macros que componen LATEX .....  | 182 |
| Tabla 4.2. Código fuente de la macro MakePlateRecord.....  | 183 |
| Tabla 4.3. Genes y Polimorfismos Identificados. Proceso preliminar.....  | 194 |
| Tabla 4.4. Actividad 2: Lista de 20 genes (parcial).....   | 199 |
| Tabla 4.5. Actividad 3: Lista de 20 genes (parcial).....   | 200 |
| Tabla 4.6. Actividad 4: Lista de 20 genes (parcial).....   | 202 |
| Tabla 4.7. Características demográficas, genotípicas y bioquímicas de los participantes del FOS según su sexo.....                                   | 209 |

|   |     |
|---|-----|
| Tabla 4.8. Niveles en plasma de lípidos, lipoproteínas y apolipoproteínas de los sujetos del FOS según los genotipos TaqIB CETP.....  | 211 |
| Tabla 4.9. Distribuciones de las subclases de lipoproteínas de los sujetos del FOS según los genotipos de TaqIB CETP .....  | 214 |
| Tabla 4.10. Diámetros de lipoproteínas (nm) de los sujetos del FOS según los genotipos del TaqIB CETP .....   | 215 |
| Tabla 4.11. Características demográficas, bioquímicas y dietéticas de la población (media y desviación estándar) .....  | 218 |
| Tabla 4.12. Concentraciones en plasma de lípidos, consumo de alcohol, tabaco y grasa en dieta según los genotipos de APOE y sexo.....   | 219 |
| Tabla 4.13. Coeficientes de correlación de Pearson entre grasa de la dieta y el aporte de energía y el consumo de alcohol por sexo y genotipo de APOE .....   | 220 |
| Tabla 4.14. Efecto del genotipo de APOE, alcohol, grasa saturada y la interacción APOE*alcohol en el LDL-C del plasma . Análisis de regresión lineal por sexo.....  | 224 |
| Tabla 4.15. Concentraciones medias de LDL-C plasmático según el genotipo de APOE, estratificadas por el consumo de alcohol .....  | 225 |
| Tabla 4.16. Asociación entre ingesta de alcohol y concentraciones de LDL-C plasmático por genotipo de APOE ajustado por edad, IMC, grasa aporte de energía y tabaco .....   | 226 |
| Tabla 4.17. Características demográficas, bioquímicas y dietéticas de la población .....  | 227 |
| Tabla 4.18. Concentraciones en plasma de lípidos, lipoproteínas, apolipoproteínas, e ingesta de grasa en dieta según los genotipos de APOA1 y sexo.....   | 228 |
| Tabla 4.19. Interacción del consumo de grasa con los efectos de los alelos de APOA1 (-75 bp) en los niveles de HDL-C, controlados por edad, IMC, consumo de alcohol y de tabaco. Se aplica análisis de regresión lineal múltiple en mujeres ..... | 231 |
| Tabla 4.20. Interacción del consumo de grasa con los efectos de los alelos de APOA1 (-75 bp) en los niveles de HDL-C, controlados por edad, IMC, consumo de alcohol y de tabaco. Se aplica análisis de regresión lineal múltiple en hombres ..... | 232 |
| Tabla 4.21. HDL-C plasmático y niveles de apo A-I por los genotipos de APOA1, ingesta de grasa poliinsaturada y sexo.....   | 233 |
| Tabla 4.22. Características demográficas y bioquímicas y de los datos dietéticos y genéticos de los participantes según el sexo.....  | 239 |
| Tabla 4.23. Media de lípidos plasmáticos e ingesta dietaria según los genotipos de LIPC   | 241 |
| Tabla 4.24. Interacción del consumo de grasa con los efectos de los genotipos de lipasa hepática sobre las concentraciones de HDL-C. Análisis de regresión  |     |

|  |     |
|--|-----|
| combinado múltiple (centrado y no centrado por grasa) por el tipo de grasa consumida.....  | 242 |
| Tabla 4.25. Valores de P para los términos de interacción entre ingesta de grasa (en continuo) y el genotipo de la lipasa hepática (tres categorías) en la determinación de las concentraciones de HDL-C y la distribución de subclases. Análisis de regresión multivariados por el tipo de grasa consumida..... | 243 |
| Tabla 4.26. Coeficientes de correlación de Pearson entre tipos de grasa dietaria en los participantes del estudio .....  | 246 |
| Tabla 4.27. Características demográficas, bioquímicas, dietarias y genotípicas de los participantes según el sexo y la diabetes de tipo 2 .....  | 249 |
| Tabla 4.28. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 1 de SR-BI para hombres y mujeres .....  | 250 |
| Tabla 4.29. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 1 de SR-BI dependientes de la diabetes 2 .....   | 251 |
| Tabla 4.30. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 8 de SR-BI para hombres y mujeres. Interacción con la diabetes 2 .....   | 252 |
| Tabla 4.31. Niveles plasmáticos de lípidos, lipoproteínas, apolipoproteínas y tamaño de partículas por el genotipo del exón 8 de SR-BI para hombres y mujeres. Interacción con la diabetes 2 .....   | 261 |
| Tabla 4.32. Características demográficas, bioquímicas y de estilo de vida de los sujetos del estudio.....  | 262 |
| Tabla 4.33. Distribución de genotipo, frecuencias alélicas y desequilibrio de ligamiento de las variantes genéticas polimórficas en el locus del PLIN en sujetos de la población general.....  | 263 |
| Tabla 4.34. Índice de masa corporal (IMC) y fenotipos de obesidad según el tipo de portador de la variante del alelo 2 en cada uno de los polimorfismos de PLIN en sujetos de la población general. Medias en hombres ajustadas por la edad .....  | 264 |
| Tabla 4.35. Índice de masa corporal (IMC) y fenotipos de obesidad según el tipo de portador de la variante del alelo 2 en cada uno de los polimorfismos de PLIN en sujetos de la población general. Medias en mujeres ajustadas por la edad .....  | 265 |
| Tabla 4.36. Prevalencia de los portadores del alelo 2 en PLIN1 y PLIN4 en sujetos obesos (casos) y no obesos (controles), y riesgo (OR y 95% CI) de obesidad en portadores del alelo 2 en comparación con los homocigotos del alelo común (11) .....   | 266 |

|  |     |
|--|-----|
| Tabla 4.37. Áreas de Auditoría y Auditorías Tecnologías que se relacionan con los elementos del Espacio del Problema de Auditoría identificados..... | 273 |
| Tabla 4.38. Identificación de riesgos asociados a los elementos del Espacio del Problema de Auditoría.....   | 274 |
| Tabla 4.39. Asignación de Dominios de COBIT a elementos del Espacio del Problema de Auditoría.....   | 277 |
| Tabla 4.40. Dominios y Procesos de COBIT que intervienen sobre las funciones principales de la Bioinformática.....                                   | 281 |
| Tabla 4.41. Propuesta para la extensión de COBIT en nuevos Dominios y Procesos para el contexto científico: el COBSIT .....                          | 283 |
| Tabla 4.42. Los nuevos Dominios y Procesos de COBSIT que intervienen sobre las funciones principales de la Bioinformática .....                      | 284 |
| Tabla 12.1. Categorías de presión sanguínea .....  | 416 |
| Tabla 12.2. Categorías de concentración de colesterol total (TC).....  | 416 |
| Tabla 12.3. Categorías de concentración de colesterol HDL (HDL-C).....   | 416 |
| Tabla 12.4. Categorías de concentración de colesterol LDL (LDL-C) .....  | 416 |
| Tabla 12.5. Descripción de los elementos estructurales de UML .....  | 423 |
| Tabla 12.6. Descripción de los elementos de comportamiento, agrupación y anotación de UML .....  | 424 |
| Tabla 12.6. Tipos de relaciones entre elementos de UML .....   | 425 |
| Tabla 12.7. Vistas de la arquitectura de un sistema.....   | 427 |
| Tabla 12.8. Diagramas de UML.....  | 428 |
| Tabla 12.8. Diagramas de UML (continuación).....   | 429 |
| Tabla 12.9. Vistas y diagramas de UML.....   | 435 |
| Tabla 12.10. Modelos y flujos de trabajo del Proceso Unificado.....  | 446 |
| Tabla 12.11. Modelos y flujos de trabajo del proceso: desarrollo técnico .....   | 447 |
| Tabla 12.12. Tabla de conversión de unidades entre el sistema no métrico y el sistema métrico.....   | 470 |
| Tabla 12.13. Prefijos para los múltiplos y divisores de las unidades en el SI .....  | 470 |
| Tabla 12.14. Genes y Polimorfismos Identificados. Proceso completado.....  | 471 |

### 12.1 Cálculo clásico del riesgo cardiovascular: la ecuación clásica de Framingham

En (Wilson et al., 1998) se da un modelo predictivo del riesgo cardiovascular basado en el estudio de una parte de la población de muestra del *Framingham Heart Study* a lo largo de 12 años. Este modelo es de tipo estadístico basado en el análisis de los parámetros siguientes:

1. *Presión sanguínea*: agrupa las medidas continuas por categorías (Tabla 12.1).
2. *Colesterol (TC)*: agrupa las medidas continuas por categorías (Tabla 12.2).
3. *HDL-C*: agrupa las medidas continuas por categorías (Tabla 12.3).
4. *LDL-C*: agrupa las medidas continuas por categorías (Tabla 12.4).
5. *Edad*: variable continua.
6. *Tratamiento de diabetes*: variable discontinua (Si/No).
7. *Tabaco*: variable discontinua (Si/No).
8. *Peso*: variable continua.
9. *Altura*: variable continua.
10. *Índice de Masa Corporal (IMC)*: variable continua dependiente de peso y altura. Se calcula como  $IMC = \text{Peso} / \text{Altura}^2$  (Kg/m<sup>2</sup>).
11. *Sexo*: variable discontinua. Sirve para diferenciar el modelo según los sexos ya que el riesgo es sensiblemente distinto.



**Tabla 12.1. Categorías de presión sanguínea**

| Categoría           | Sistólica (mm Hg) | Diastólica (mm Hg) |
|---------------------|-------------------|--------------------|
| Optima              | < 120             | < 80               |
| Normal              | 120-129           | 80-84              |
| Normal alta         | 130-139           | 85-89              |
| Hipert. Fase I      | 140-159           | 90-99              |
| Hipert. Fases II-IV | $\geq$ 160        | $\geq$ 100         |

**Tabla 12.2. Categorías de concentración de colesterol total (TC)**

| Categoría | Concentración (mg/dL) |
|-----------|-----------------------|
| 1         | < 160                 |
| 2         | 160-199               |
| 3         | 200-239               |
| 4         | 240-279               |
| 5         | $\geq$ 280            |

**Tabla 12.3. Categorías de concentración de colesterol HDL (HDL-C)**

| Categoría | Concentración (mg/dL) |
|-----------|-----------------------|
| 1         | < 35                  |
| 2         | 35-44                 |
| 3         | 45-49                 |
| 4         | 50-59                 |
| 5         | $\geq$ 60             |

**Tabla 12.4. Categorías de concentración de colesterol LDL (LDL-C)**

| Categoría | Concentración (mg/dL) |
|-----------|-----------------------|
| 1         | < 100                 |
| 2         | 100-129               |
| 3         | 130-159               |
| 4         | 160-189               |
| 5         | $\geq$ 190            |

No se consideran los factores siguientes:

1. Antecedentes familiares de CHD prematura.
2. Actividad física.

3. Obesidad.
4. ERT (*estrogen replacement therapy*) para mujeres posmenopáusicas.
5. Tratamiento para hipertensión.
6. Tratamiento para hipercolesterolemia.

Los coeficientes beta resultantes de las regresiones lineales y logísticas aplicadas combinadamente se incorporan en una función lineal para el cálculo del riesgo, que se refina en varias etapas:

### Etapa 1:

Se define la siguiente ecuación para los hombres (la estructura es la misma para las mujeres) ajustando por colesterol total (TC):

$$\begin{aligned}
 L\_Chol_{men} = & 0,04826 * age + \left\{ \begin{array}{ll} chol < 160 & -0,65945 \\ 160 \leq chol < 200 & 0,0 \\ 200 \leq chol < 240 & 0,17692 \\ 240 \leq chol < 280 & 0,50539 \\ 280 \leq chol & 0,65713 \end{array} \right\} * chol + \\
 & \left\{ \begin{array}{ll} hdl - c < 35 & 0,49744 \\ 35 \leq hdl - c < 45 & 0,24310 \\ 45 \leq hdl - c < 50 & 0,0 \\ 50 \leq hdl - c < 60 & -0,05107 \\ 60 \leq hdl - c & -0,48660 \end{array} \right\} * hdl-c + \left\{ \begin{array}{ll} bp : optimal & -0,00226 \\ bp : normal & 0,0 \\ bp : high\_n & 0,28320 \\ bp : hypert\_I & 0,52168 \\ bp : hypert\_II & 0,61859 \end{array} \right\} * bp \\
 & + \left\{ \begin{array}{ll} diab : YES & 0,42809 \\ diab : NO & 0,0 \end{array} \right\} * diab + \left\{ \begin{array}{ll} smokers : YES & 0,52337 \\ smokers : NO & 0,0 \end{array} \right\} * smokers
 \end{aligned}$$

### Etapa 2:

Se evalúa la ecuación anterior con los valores de las medias de cada variable para los hombres (la estructura es la misma para las mujeres):

$$G\_Chol_{men} = 0,04826 * \overline{age} + \sum_i \alpha_i * \overline{chol}_i + \sum_j \beta_j * \overline{hdl - c}_j + \sum_k \gamma_k * \overline{bp}_k + \sum_{a=1}^2 \delta_a * \overline{diab}_a + \sum_{b=1}^2 \omega_b * \overline{smoker}_b$$

**Etapas 3:**

Se repiten los mismos pasos 1 y 2, ajustando por LDL-C para los hombres (la estructura es la misma para las mujeres):

$$L\_LDL_{men} = 0,04808 * age + \left\{ \begin{array}{ll} ldl - c < 100 & -0,69281 \\ 100 \leq ldl - c < 130 & 0,0 \\ 130 \leq ldl - c < 160 & 0,00389 \\ 160 \leq ldl - c < 190 & 0,26755 \\ 190 \leq ldl - c & 0,56705 \end{array} \right\} * ldl-c + \left\{ \begin{array}{ll} hdl - c < 35 & 0,48598 \\ 35 \leq hdl - c < 45 & 0,21643 \\ 45 \leq hdl - c < 50 & 0,0 \\ 50 \leq hdl - c < 60 & -0,04710 \\ 60 \leq hdl - c & -0,34190 \end{array} \right\} * hdl-c + \left\{ \begin{array}{ll} bp : optimal & -0,02642 \\ bp : normal & 0,0 \\ bp : high\_n & 0,30104 \\ bp : hypert\_I & 0,55714 \\ bp : hypert\_II & 0,65107 \end{array} \right\} * bp + \left\{ \begin{array}{ll} diab : YES & 0,42146 \\ diab : NO & 0,0 \end{array} \right\} * diab + \left\{ \begin{array}{ll} smoker : YES & 0,54377 \\ smoker : NO & 0,0 \end{array} \right\} * smokers$$

Y también:

$$G\_LDL_{men} = 0,04808 * \overline{age} + \sum_i \alpha_i * \overline{ldl - c}_i + \sum_j \beta_j * \overline{hdl - c}_j + \sum_k \gamma_k * \overline{bp}_k + \sum_{a=1}^2 \delta_a * \overline{diab}_a + \sum_{b=1}^2 \omega_b * \overline{smoker}_b$$

**Etapa 4:**

Se calcula la diferencia entre los resultados de la primera y la segunda ecuación para los ajustes por colesterol (TC) y por LDL-C para los hombres (la estructura es la misma para las mujeres):

$$A\_Chol_{men} = L\_Chol_{men} - G\_Chol_{men}$$

$$A\_LDL_{men} = L\_LDL_{men} - G\_LDL_{men}$$

**Etapa 5:**

Se calculan los **Odd Ratios** relativos de cada una de las ecuaciones anteriores para los hombres (la estructura es la misma para las mujeres):

$$B\_Chol_{men} = e^{A\_Chol_{men}}$$

$$B\_LDL_{men} = e^{A\_LDL_{men}}$$

**Etapa 6:**

Finalmente, se calcula la probabilidad de CHD a 10 años para los hombres (la estructura es la misma para las mujeres):

$$P\_Chol_{men} = 1 - [sub\_Chol_{men}^{10}(t)]^{B\_Chol_{men}}$$

$$P\_LDL_{men} = 1 - [sub\_LDL_{men}^{10}(t)]^{B\_LDL_{men}}$$

donde

**sub\_Chol<sub>men</sub><sup>10</sup>(t)**: es el valor de supervivencia proyectado a 10 años, s(t), calculado para el ajuste por TC ( $s(t)_{men} = 0,90015$ ,  $s(t)_{women} = 0,96246$ ).

**sub\_LDL<sub>men</sub><sup>10</sup>(t)**: es el valor de supervivencia proyectado a 10 años, s(t), calculado para el ajuste por LDL-C ( $s(t)_{men} = 0,90017$ ,  $s(t)_{women} = 0,96280$ ).

## 12.2 El Lenguaje UML

La notación UML (*Unified Modeling Language*) fue diseñada principalmente para el modelado de sistemas de software orientado a objetos por James Rumbaugh y Grady Booch, a los que se unió Ivar Jacobson. Los tres son conocidos expertos de software y autores de sus propias metodologías orientadas a objetos. Por eso la notación UML es una fusión de las notaciones de Booch, OMT, OOSE y otras notaciones.

La primera versión pública de UML, UML ver. 0.8, se dio a conocer en 1995 y se mejoró en 1996 publicando versión 0.9. Los autores no establecieron ningún derecho de explotación sobre UML, salvo el de la propiedad intelectual, por lo que tuvo una buena acogida en la comunidad científica y técnica. Esto facilitó que se fuera extendiendo en su uso y que muchos usuarios aportaran sugerencias de mejora y de extensión a otros ámbitos hasta llegar a UML versión 1.0.

Así, en 1997, tras haberse convertido casi en un estándar de facto, la versión 1.0 se sometió a la evaluación de la organización OMG (*Object Management Group*) (OMG, 2004). En el proceso de revisión, las sucesivas mejoras sugeridas por la *UML Revision Task Force* (UML RTF) pasaron a ser las versiones 1.1 (1997) y 1.2 (1998). En 1999, después de la última revisión técnica de la UML RTF, se publicó la versión 1.3 como el estándar OMG. A partir de entonces, UML se sometió a la evaluación de una *Task Force* de ISO (*International Standard Organization*) (ISO, 2004) con la intención de publicar una versión UML ISO. La UML RTF sugirió unas mejoras de importancia que se incorporaron parcialmente en las versiones 1.4 en 2000 y 1.5 en 2001. Actualmente, aunque continua el proceso de revisión en ISO, se está procediendo paralelamente a la elaboración de UML versión 2.0 (UML, 2004), algunas de cuyas partes ya se están aplicando. Sin embargo, no se puede afirmar taxativamente que UML 2 es ISO.

### 11.2.1 La notación UML para el Modelo Orientado a Objetos

UML fue pensado para ser legible sobre soportes muy variados, como las pizarras, el papel, los manteles de los restaurantes, las pantallas de ordenador, las impresiones en blanco y negro, etc. Los diseñadores de la notación buscaron ante todo la simplicidad, la intuición, la homogeneidad y la coherencia. Se eliminaron en la medida de lo posible los símbolos embrollados, redundantes o superfluos a favor de un mejor aspecto visual. No obstante, el lenguaje necesita de una formación específica ya que es bastante extenso puesto que intenta cubrir la mayoría de aspectos de la orientación a objetos.

UML se concentra sobre la descripción de los artefactos del desarrollo de programa, en lugar de en la formalización del propio proceso de desarrollo: así, puede utilizarse para describir los elementos lógicos, obtenidos por la aplicación de diferentes procesos de desarrollo. UML no es una notación cerrada, sino que es genérica, extensible y configurable por el usuario. UML no busca la especificación a ultranza: no tiene una representación gráfica para todos los conceptos imaginables; en caso de necesidades particulares, pueden aportarse precisiones por medio de mecanismos de extensión y de comentarios textuales. Se da una gran libertad a las herramientas para el filtrado y la visualización de información. El uso de colores, de dibujos y de atributos gráficos se deja a la discreción del usuario.

El hecho de que UML se haya convertido en un lenguaje de notación estándar y que se haya ofrecido libremente a la comunidad informática, ha favorecido enormemente su expansión y la adopción del mismo por los grandes fabricantes de equipos y software, como IBM, Digital, Microsoft, etc. También se han desarrollado herramientas informáticas OOCASE (*Object-Oriented Computer Aided Software Engineering*) que soportan el diseño y desarrollo de sistemas orientados a objetos con UML. Por ejemplo: Rational Rose, Together, Poseidon, Argos, etc.

### 11.2.2 El modelo conceptual de UML

UML no es una metodología de diseño, a diferencia de OMT (de James Rumbaugh y otros colaboradores), Booch Method (de Grady Booch), u OOSE (de Ivar Jacobson). Por tanto, los propios Rumbaugh, Booch y Jacobson han diseñado una metodología que enseña a utilizar correctamente UML en el proceso de modelado de sistemas. Además, UML es bastante independiente del proceso, lo que significa que se puede utilizar con diferentes procesos de ingeniería del software. Por lo tanto, el modelo conceptual de UML debe entenderse solamente en los límites estrictos de la representación de sistemas orientados a objetos, pero no del proceso de representación.

El modelo conceptual del lenguaje lo componen tres elementos: bloques básicos de construcción, reglas de combinación de los bloques y mecanismos comunes. Los bloques básicos de construcción son los siguientes:

- *Elementos*: son las abstracciones de elementos del problema.
- *Relaciones*: son las ligaduras entre los elementos.
- *Diagramas*: son las agrupaciones de colecciones de elementos y relaciones.

Hay cuatro tipos de elementos que son los siguientes:

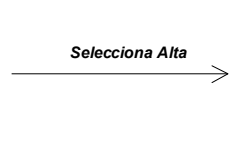
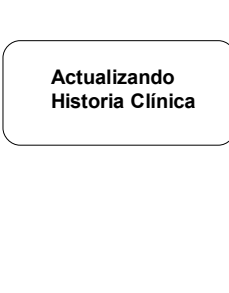
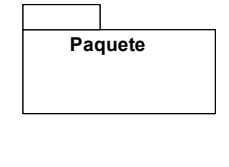
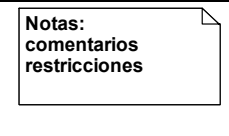
- *Elementos estructurales*: constituyen los nombres del modelo o las partes estáticas del mismo. Hay siete tipos distintos (Tabla 12.5).
- *Elementos de comportamiento*: constituyen los verbos del modelo o las partes dinámicas del mismo. Hay dos tipos distintos (Tabla 12.6).
- *Elementos de agrupación*: son cajas de descomposición del modelo que ayudan a estructurarlo en partes organizativas (Tabla 12.6).
- *Elementos de anotación*: son comentarios o las partes explicativas del modelo (Tabla 12.6).

Tabla 12.5. Descripción de los elementos estructurales de UML

| Nombre       | Descripción  | Símbolo |
|--------------|--|---------|
| Clase        | Descripción de un conjunto de objetos que comparten los mismos atributos, operaciones, relaciones y semántica. Implementa al menos una interfaz.   |         |
| Interfaz     | Colección de operaciones que especifican un servicio de una clase o componente. Describe el comportamiento visible y define el conjunto de especificaciones de operaciones: signatura de la operación.   |         |
| Colaboración | Es una sociedad de roles y otros elementos que colaboran para proporcionar un comportamiento cooperativo mayor que la suma de los comportamientos de sus elementos. Una clase puede participar en varias colaboraciones. También representa la implementación de patrones que forman un sistema.                     |         |
| Caso de uso  | Es una descripción de un conjunto de secuencias de acciones que un sistema ejecuta y que produce un resultado observable de interés para un actor en particular. Estructura los aspectos de comportamiento en un modelo. Se realiza por una colaboración.  |         |
| Clase activa | Es una clase cuyos objetos tienen uno o más procesos o hilos ( <i>threads</i> ) de ejecución. Los objetos pueden dar origen a actividades de control y representan a elementos se pueden comportar concurrentemente con respecto a otros elementos.  |         |
| Componente   | Es una parte física y reemplazable de un sistema que se combina con un conjunto de interfaces y proporciona la implementación del conjunto. Representa el empaquetamiento físico de diferentes elementos lógicos: clases, interfaces y colaboraciones. Son componentes de despliegue: COM+, JavaBeans, fuentes, etc. |         |
| Nodo         | Es un elemento físico que existe en tiempo de ejecución y representa un recurso computacional (memoria, proceso, i/o). Un conjunto de componentes puede residir en un nodo y también puede migrar de un nodo a otro.   |         |



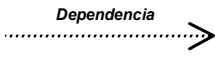
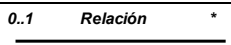
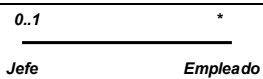
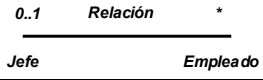
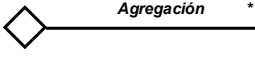
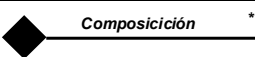


**Tabla 12.6. Descripción de los elementos de comportamiento, agrupación y anotación de UML**

| Nombre                                 | Descripción  | Símbolo   |
|--|--|---|
| Interacción<br>(comportamiento)        | Es un comportamiento que comprende un conjunto de mensajes intercambiados entre un conjunto de objetos, dentro de un contexto particular, para alcanzar un propósito específico. Implica: mensajes, secuencias de acción y enlaces. Especifica comportamiento de: sociedad de objetos, operación individual.   |    |
| Máquina de estados<br>(comportamiento) | Es un comportamiento que especifica lo siguiente: <ul style="list-style-type: none"> <li>las secuencias de estados por las que pasa un objeto.</li> <li>una interacción durante la vida del objeto en respuesta a sucesos, junto con las reacciones a estos sucesos.</li> <li>comportamiento de clase de objetos y colaboración de clases.</li> </ul> Implica: estados, transiciones, sucesos y actividades. |   |
| Paquete<br>(agrupación)                | Es la parte organizativa de los modelos UML puesto que es un mecanismo de propósito general para organizar elementos en grupos. Los propios paquetes puede agruparse en paquetes. Es un elemento puramente conceptual ya que sólo existe en tiempo de desarrollo.  |  |
| Notas<br>(anotación)                   | Es la parte explicativa de los modelos UML a base de comentarios para describir, clarificar y hacer observaciones sobre el resto de elementos. También es un símbolo que muestra restricciones y comentarios de un elemento o colección de elementos.  |  |

Hay cuatro tipos de relaciones que son las siguientes:

- *Dependencia*: relación semántica entre elementos (Tabla 12.6).
- *Asociación*: relación estructural entre elementos, principalmente entre clases y entre objetos (Tabla 12.6).
- *Generalización*: relación de especialización y/o generalización entre elementos, principalmente entre clases y entre objetos (Tabla 12.6).
- *Realización*: relación semántica entre clasificadores (Tabla 12.6).

Tabla 12.6. Tipos de relaciones entre elementos de UML

| Nombre         | Descripción   | Símbolo   |
|----------------|---|---|
| Dependencia    | Es la relación semántica entre dos elementos, uno de ellos es el elemento independiente, y el otro el dependiente. Un cambio en el elemento independiente puede afecta a la semántica del elemento dependiente.   |    |
| Asociación     | Es la relación estructural que describe un conjunto de enlaces, siendo cada enlace una conexión entre objetos.<br><br>Hay una asociación especial, la <i>agregación</i> , que es una asociación entre un todo y sus partes. Hay variantes de notación empleando nombres de etiqueta, nombres de rol y ambos. También se añade la multiplicidad en cada uno de los extremos de la asociación: «0..1», «0..n», «+», etc.<br><br>La <i>composición</i> es una agregación especial donde el todo posee fuertemente a sus partes. Si se copia o se borra un objeto de la clase todo, sus objetos partes se copian o borran solidariamente. |    |
|                |   |    |
|                |   |    |
|                |   |    |
|                |   |    |
| Generalización | Es la relación de especialización y/o generalización entre objetos padre e hijo de forma que: <ul style="list-style-type: none"> <li>los objetos hijo (especializados) sustituyen a los objetos padre (general).</li> <li>los objetos especializados comparten la estructura y comportamiento del objeto general.</li> <li>los objetos padre (generalizados) recogen la estructura y comportamiento comunes de los objetos hijos.</li> </ul>  |  |
| Realización    | Es la relación semántica entre clasificadores. Un clasificador especifica un contrato que otro clasificador garantiza que cumplirá. Hay realizaciones entre interfaces y clases /componentes, y realizaciones entre casos de uso y colaboraciones.  |  |

### 11.2.3 La arquitectura de un sistema en UML

La *arquitectura* de un sistema en UML es un artefacto que permite decidir sobre la organización de un sistema software, mediante la selección de elementos estructurales y sus interfaces; sobre el comportamiento de dicho sistema, mediante la especificación de operaciones en las colaboraciones entre los elementos estructurales; y el estilo arquitectónico, mediante la composición de elementos estáticos y dinámicos. Se denomina también arquitectura software porque

principalmente se aplica en el modelado y desarrollo de sistemas software. La arquitectura software también se relaciona con otros aspectos también importantes: la utilización, la funcionalidad, el rendimiento, la reutilización, las restricciones económicas y tecnológicas, etc.

Sin embargo, cualquier sistema software es tan complejo que debe ser estudiado desde diferentes perspectivas en distintos momentos del proyecto. Además, los actores y/o usuarios relacionados con dicho sistema tienen intereses distintos y siguen diferentes agendas con relación al proyecto. Tales actores son los usuarios finales, los analistas, los desarrolladores, los integradores de sistemas, etc. Los elementos conceptuales o (artefactos en UML) que encarnan estas perspectivas e intereses son las vistas. En cada una de las vistas se tiene en cuenta los aspectos estáticos y dinámicos, que son axiomas esenciales en el concepto general de objeto dentro del Paradigma de la Orientación a Objetos.

Pero las vistas no son suficientes para describir la arquitectura por su generalidad. Entonces, son necesarios otros artefactos que describan con mayor detalle y amplitud cada una de las vistas y que, además, sean independientes de las propias vistas. Estos artefactos son los diagramas.

#### **11.2.4 Las vistas de UML**

Las vistas son las proyecciones de la organización y la estructura del sistema que se centran en un aspecto particular del sistema. Así, en UML se presentan cinco vistas según sendos criterios (Figura 12.1): vista de casos de uso, vista de diseño, vista de procesos, vista de implementación y vista de despliegue.

Cada una de estas vistas presenta además los aspectos estáticos, mediante los diagramas estructurales de UML, y los aspectos dinámicos, mediante diagramas dinámicos de UML. Así, se puede trabajar con la vista de casos de uso estática y la vista de casos de uso dinámica, la vista de diseño estática y la vista de diseño dinámica, y así sucesivamente. En la Tabla 12.7 se define cada una de las vistas.

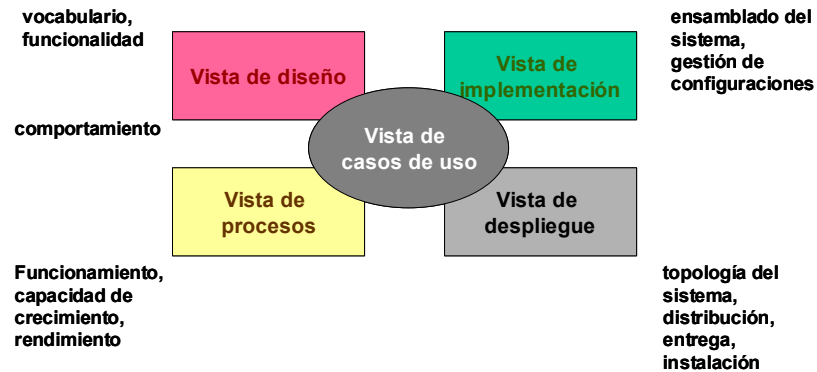


Figura 12.1. Vistas de la arquitectura de un sistema

Tabla 12.7. Vistas de la arquitectura de un sistema

| Nombre                  | Descripción   | Aspectos Estáticos                                    | Aspectos Dinámicos   |
|-------------------------|---|---|--|
| Vista de casos de uso   | Proyecta el comportamiento del sistema tal y como es percibido por los usuarios finales, analistas y encargados de las pruebas. Especifica las fuerzas que configuran la arquitectura del sistema.  | Diagramas de casos de uso                             | Diagramas de interacción<br>Diagramas de estados                             |
| Vista de diseño         | Soporta los requisitos funcionales del sistema: servicios proporcionados a los usuarios finales. Vocabulario del problema y su solución: clases, interfaces y colaboraciones  | Diagramas de clases<br>Diagramas de objetos           | Diagramas de interacción<br>Diagramas de estados<br>Diagramas de actividades |
| Vista de procesos       | Cubre el funcionamiento, capacidad de crecimiento y rendimiento del sistema. Mecanismos de sincronización y concurrencia del sistema: hilos y procesos  | Diagramas de clases (activas)<br>Diagramas de objetos | Diagramas de interacción<br>Diagramas de estados<br>Diagramas de actividades |
| Vista de implementación | Cubre la gestión de configuraciones de las distintas versiones de un sistema a partir de componentes y archivos quasi-independientes. Ensamblado y disponibilidad del sistema: componentes y archivos   | Diagramas de componentes                              | Diagramas de interacción<br>Diagramas de estados<br>Diagramas de actividades |
| Vista de despliegue     | Contiene los nodos que forman la arquitectura (topología) hardware sobre la que se ejecuta el sistema a través de sus componentes. Está destinada a representar la distribución, entrega e instalación de las partes que forman el sistema informático físico | Diagramas de despliegue                               | Diagramas de interacción<br>Diagramas de estados<br>Diagramas de actividades |

### 11.2.5 Los diagramas de UML

Los diagramas de UML son la representación gráfica de un conjunto de elementos como grafo conexo donde los nodos son los elementos y los arcos son las relaciones. Visualizan el sistema desde distintas perspectivas, que son las proyecciones de los modelos según diferentes criterios. Ocurre que puede haber elementos comunes entre distintos diagramas. Hay nueve diagramas en UML agrupados en cinco vistas que se definen y distribuyen en la Tabla 12.8.

**Tabla 12.8. Diagramas de UML**

| Diagrama                                   | Descripción  | Asociada a Vistas   | Ejemplo gráfico |
|--|--|---|-----------------|
| <b>1. Diagrama de clases</b>               | Es el conjunto de clases, interfaces, colaboraciones y relaciones entre ellas  | Vista de Diseño Estática<br>Vista de Proceso Estática (cuando hay algunas clases que son <i>clases activas</i> )                | Figura 12.2     |
| <b>2. Diagrama de objetos</b>              | Es el conjunto de objetos y sus relaciones. Constituyen las instantáneas de instancias de los elementos del diagrama de clases. Se puede utilizar en la Vista de Diseño Estática y en la Vista de Proceso Estática cuando hay algunos objetos que son instancias de clases activas | Vista de Diseño Estática<br>Vista de Proceso Estática (cuando hay algunos objetos que son instancias de <i>clases activas</i> ) | Figura 12.3     |
| <b>3. Diagrama de casos de uso</b>         | Es el conjunto de casos de uso, actores y sus relaciones. Se utiliza para el modelado del comportamiento del sistema   | Vista de Casos de Uso Estática  | Figura 12.4     |
| <b>4. Diagrama de secuencia de sucesos</b> | Es uno de los dos tipos de diagrama de interacción, entendido como el conjunto de objetos, sus relaciones y los mensajes que pueden ser enviados entre ellos. Este tipo de diagrama de interacción resalta la ordenación temporal de los mensajes                                  | Subvista dinámica de las cinco vistas de UML  | Figura 12.5     |

Tabla 12.8. Diagramas de UML (continuación)

| Diagrama                           | Descripción   | Asociada a Vistas   | Ejemplo gráfico |
|------------------------------------|---|---|-----------------|
| <b>5. Diagrama de colaboración</b> | Es uno de los dos tipos de diagrama de interacción, entendido como el conjunto de objetos, sus relaciones y los mensajes que pueden ser enviados entre ellos. Este tipo de diagrama de interacción resalta la organización estructural de los objetos que envían y reciben mensajes | Subvista dinámica de las cinco vistas de UML  | Figura 12.6     |
| <b>6. Diagrama de estados</b>      | Es el conjunto de estados, transiciones, eventos y actividades, denominado también máquina de estados. Este tipo de diagrama modela el comportamiento de una interfaz, una clase, una colaboración. También resalta el comportamiento dirigido por sucesos de un objeto             | Subvista dinámica de las cinco vistas de UML<br>Es más profusamente aplicado en la Vista de Diseño Dinámica | Figura 12.7     |
| <b>7. Diagrama de actividades</b>  | Es un tipo especial de diagrama de estados donde se representa es flujo de actividades dentro de un sistema. Modela el comportamiento del sistema resaltando el flujo de control entre objetos  | Subvista dinámica de las cinco vistas de UML  | Figura 12.8     |
| <b>8. Diagrama de componentes</b>  | Es un diagrama que muestra la organización y las dependencias entre un conjunto de componentes. Estos componentes tienen relación con los diagramas de clases de forma que, cada componente representa, por ejemplo, una clase y su interfaz, una clase y una colaboración, etc.    | Vista de Implementación Estática  | Figura 12.9     |
| <b>9. Diagrama de despliegue</b>   | Es un diagrama que muestra la configuración de nodos de proceso en tiempo de ejecución y los componentes que residen en ellos. Estos nodos tienen relación con los diagramas de componentes de forma que en cada nodo se pueden ejecutar uno o más de un componente                 | Vista de Despliegue Estática  | Figura 12.10    |

Ejemplos gráficos de diagramas de UML

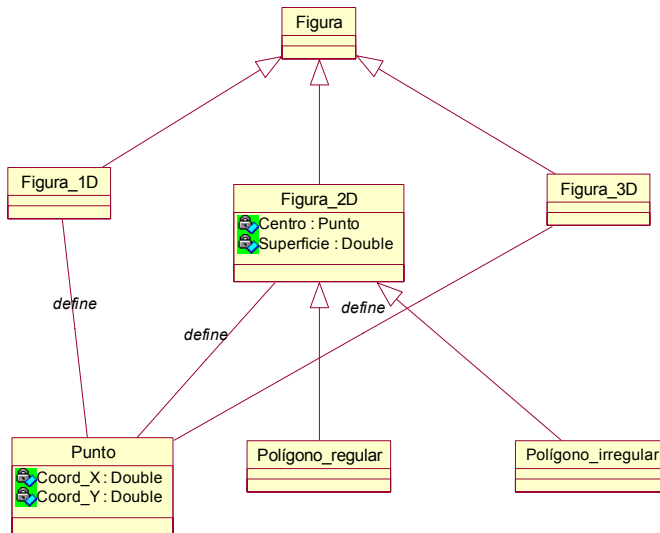


Figura 12.2. Diagrama de clases

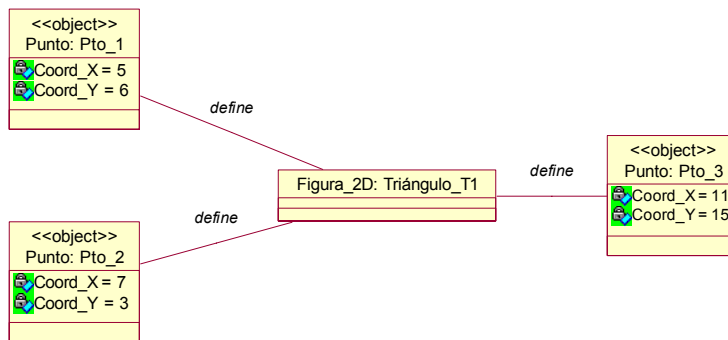


Figura 12.3. Diagrama de objetos

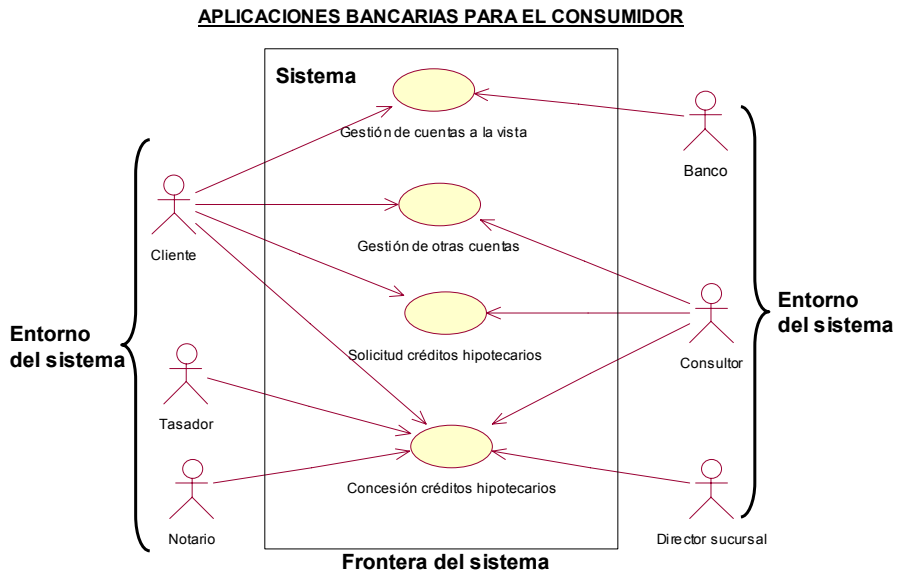


Figura 12.4. Diagrama de casos de uso

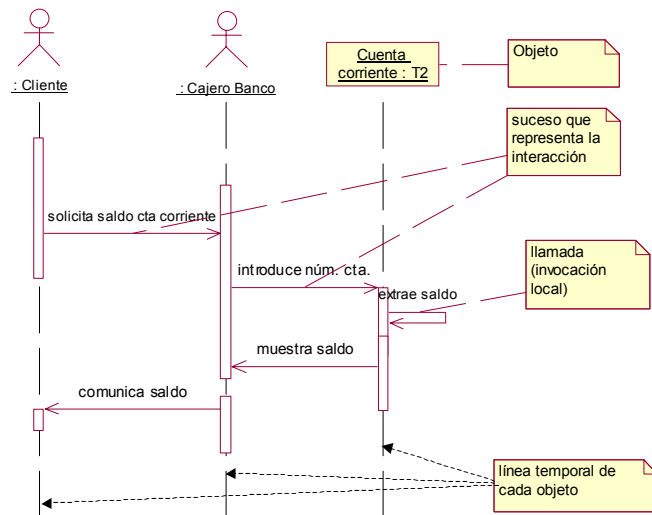


Figura 12.5. Diagrama de secuencia de sucesos



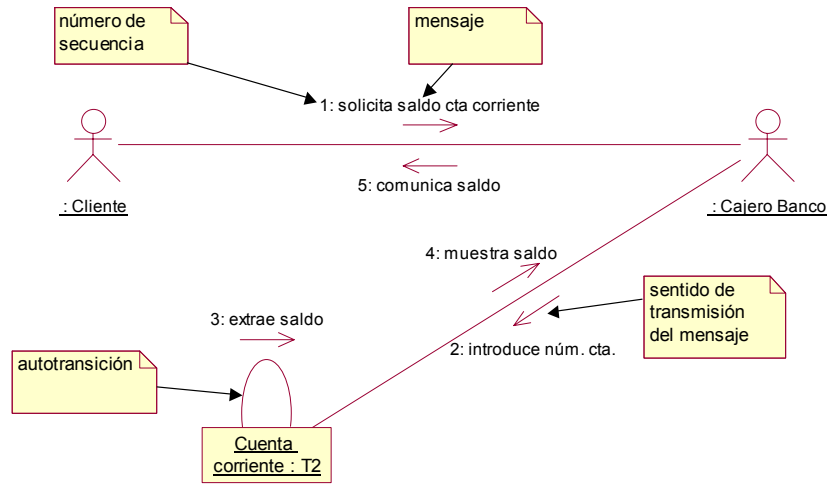


Figura 12.6. Diagrama de colaboración

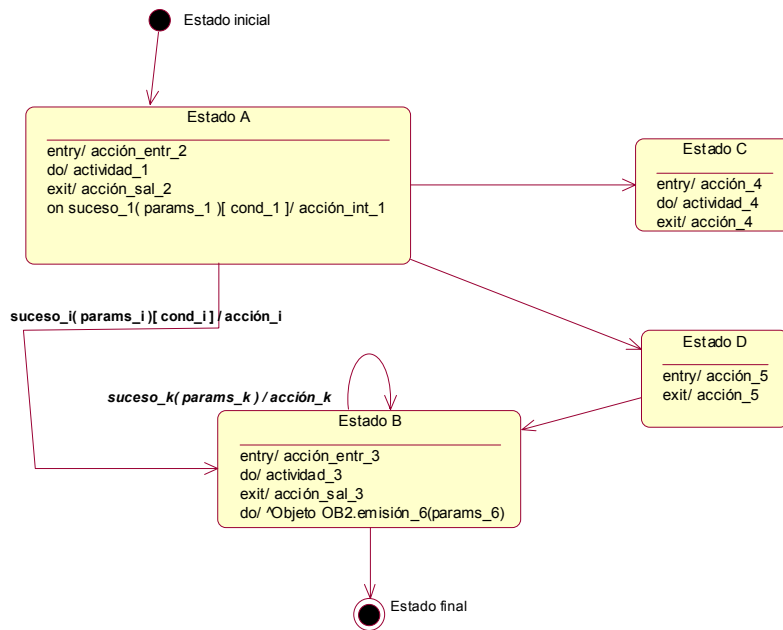


Figura 12.7. Diagrama de estados

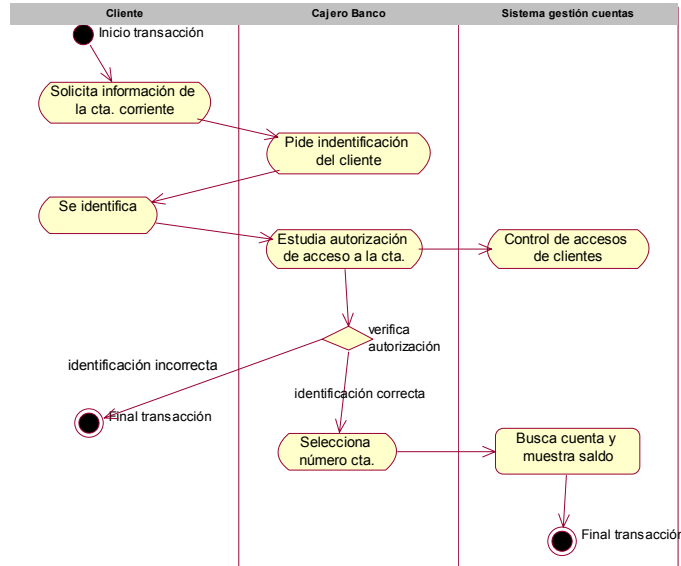


Figura 12.8. Diagrama de actividades

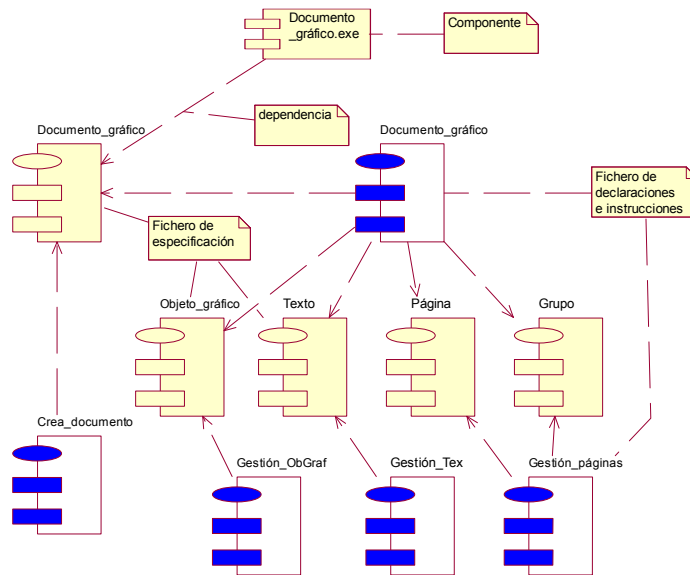


Figura 12.9. Diagrama de componentes

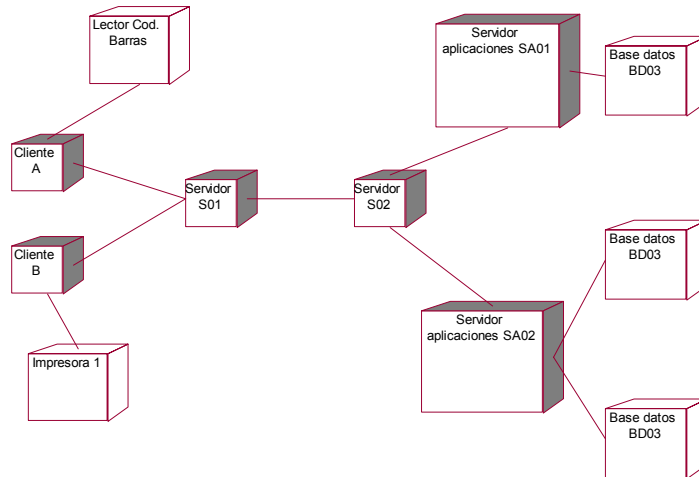


Figura 12.10. Diagrama de despliegue

En la Tabla 12.9 se muestra la adscripción de cada uno de los diagramas presentados a las cinco vistas que componen la arquitectura del sistema.

### 11.2.6 Reglas de combinación y mecanismos

Los bloques básicos de construcción se deben combinar según determinadas reglas, que se denominan *reglas de combinación de los bloques*, para obtener un modelo bien formado. Un modelo bien formado (abreviado como MBF) es un modelo semánticamente autoconsistente y en armonía con todos sus modelos relacionados. Por lo tanto, estas reglas son de tipo semántico y son las siguientes:

- *Nombres*: denominación de elementos, relaciones y diagramas.
- *Alcance*: contexto que da significado específico a un nombre.
- *Visibilidad*: forma de ver y utilizar los nombres.
- *Integridad*: forma de relacionar apropiada y consistentemente unos elementos con otros.
- *Ejecución*: consecuencias de la ejecución o simulación de un modelo dinámico.

Tabla 12.9. Vistas y diagramas de UML

|                                      | Vista de Casos de Uso |      | Vista de Diseño |      | Vista de Procesos |      | Vista de Implementación |      | Vista de Despliegue |      |
|--------------------------------------|-----------------------|------|-----------------|------|-------------------|------|-------------------------|------|---------------------|------|
|                                      | Est.                  | Din. | Est.            | Din. | Est.              | Din. | Est.                    | Din. | Est.                | Din. |
| Diagrama de Casos de Uso             | X                     |      |                 |      |                   |      |                         |      |                     |      |
| Diagrama de Interacción-Secuencia    |                       | X    |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Interacción-Colaboración |                       | X    |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Clases                   |                       |      | X               |      | X                 |      |                         |      |                     |      |
| Diagrama de Objetos                  |                       |      | X               |      | X                 |      |                         |      |                     |      |
| Diagrama de Estados                  |                       | X    |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Actividades              |                       |      |                 | X    |                   | X    |                         | X    |                     | X    |
| Diagrama de Componentes              |                       |      |                 |      |                   |      | X                       |      |                     |      |
| Diagrama de Despliegue               |                       |      |                 |      |                   |      |                         |      | X                   |      |

Est.: Estática

Din.: Dinámica

Estas reglas se aplican para controlar la obtención de modelos no bien formados, que son modelos construidos durante el desarrollo que no cumplen todos los requisitos de un MBF. Son de tres tipos:

- *Abreviados*: con elementos ocultos para simplificar la vista.
- *Incompletos*: con elementos ausentes.
- *Inconsistentes*: sin garantía de integridad.

Es normal que en el desarrollo de un sistema los modelos obtenidos en las iteraciones iniciales de cada fase sean modelos no bien formados, pero las reglas de combinación de bloques deben aplicarse para prever la aparición de los mismos y para controlar que los modelos finales sean MBF. De lo contrario, no se puede garantizar la calidad del software generado y/o construido en base a dichos modelos.

Los mecanismos comunes son los instrumentos para ajustar los modelos a patrones de características comunes. Hay cuatro tipos:

- *Especificaciones*: son las explicaciones textuales de la sintaxis y semántica de cada bloque de construcción. Por ejemplo, la especificación de clase es el conjunto completo de atributos, operaciones y comportamiento. La especificación es la base semántica que incluye a todos los modelos de todos los modelos de un sistema de forma que cada elemento está relacionado con otros de forma consistente (Figura 12.11).
- *Adornos*: son símbolos gráficos o textuales para incluir detalles complementarios en los elementos de notación gráfica básicos (Figura 12.12).
- *Divisiones comunes*: expresan las dicotomías asociadas a los bloques de construcción que se incorporan en el proceso de modelado. Hay dos tipos de dicotomías:
  - Dicotomía clase/objeto: elemento general contra elemento instancia (Figura 12.13).
  - Dicotomía interfaz/implementación: declaración de contrato contra realización concreta del contrato (Figura 12.14).

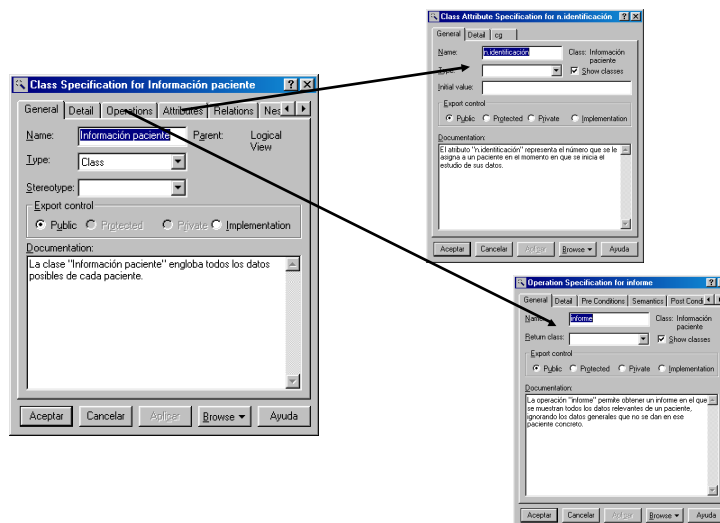


Figura 12.11. Especificación de una clase

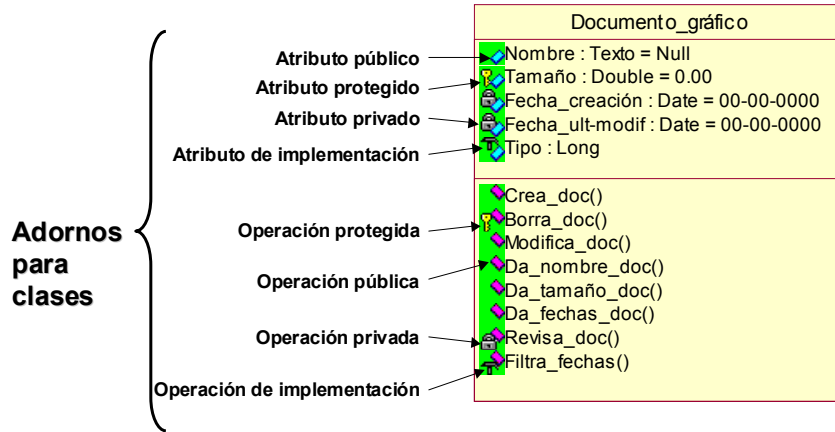


Figura 12.12. Ejemplos de adornos

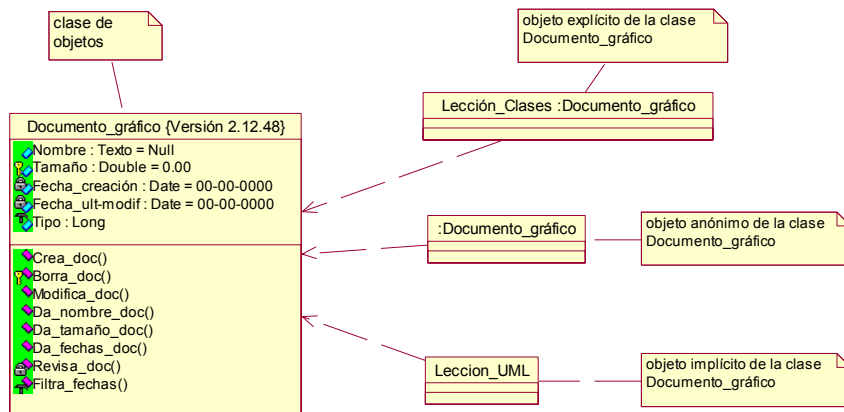


Figura 12.13. Ejemplo de dicotomía clase/objeto

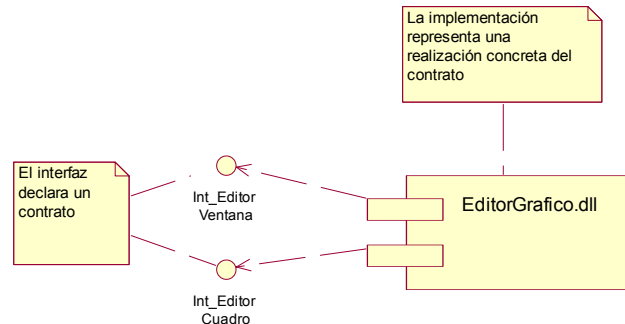


Figura 12.14. Ejemplo de dicotomía interfaz/implementación

- *Mecanismos de extensibilidad:* UML es un lenguaje abierto pero cerrado para determinados ámbitos. Sin embargo, al mismo tiempo es un lenguaje estándar con rica expresividad, pero no es suficiente para cubrir toda la expresividad de todos los modelos en todos los dominios y en todos los periodos de evolución. Por lo tanto, los mecanismos de extensibilidad son instrumentos semánticos que permiten extender el lenguaje de forma controlada. Los hay de tres tipos: Estereotipos, Valores etiquetados, y Restricciones. Se describen de la forma siguiente:
  - El *estereotipo* extiende el vocabulario de UML permitiendo la creación de nuevos tipos de bloques de construcción (Figura 12.15). Estos bloques son derivados de los existentes pero específicos a un problema dado.
  - El *valor etiquetado* extiende las propiedades de un bloque de construcción permitiendo añadir nueva información en la especificación del bloque. Y la *restricción* extiende la semántica de un bloque de construcción permitiendo añadir nuevas reglas o modificar las existentes.

Estos mecanismos permiten que UML sea adaptable a nuevas tecnologías de software y a nuevos ámbitos de aplicación. Por ejemplo, existe en la actualidad una variante de UML para el modelado de sistemas en tiempo real y otra variante para sistemas organizativos (organizaciones de empresas).

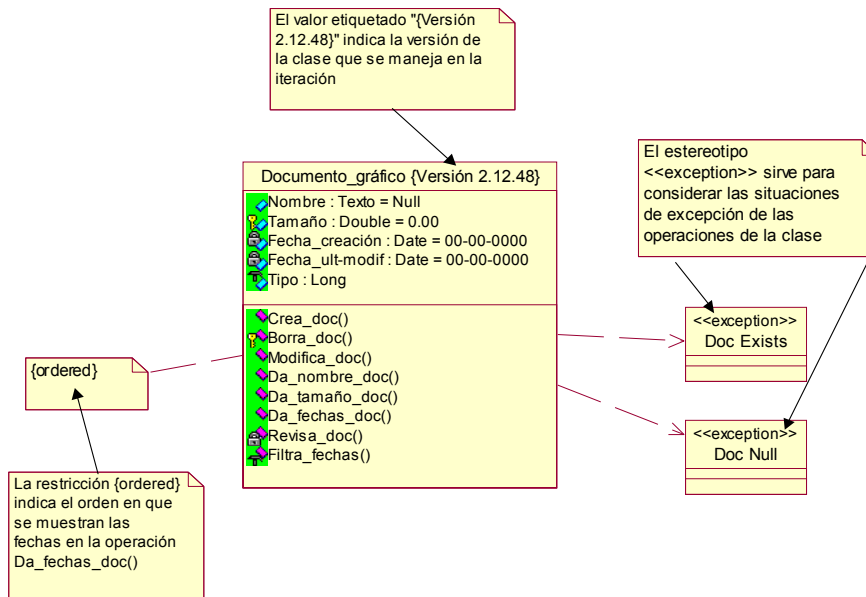


Figura 12.15. Estereotipo como mecanismo común en UML

## 12.3 El Proceso Unificado de Rational

El Proceso Unificado de Rational es un proceso iterativo. Un enfoque iterativo propone una comprensión incremental del problema a través de refinamientos sucesivos y un crecimiento incremental de una solución efectiva a través de varias versiones. Como parte del enfoque iterativo se encuentra la flexibilidad para acomodarse a nuevos requisitos o a cambios tácticos en los objetivos del negocio. También permite que el proyecto identifique y resuelva los riesgos más bien pronto que tarde.

### 11.3.1 Caracterización del Proceso Unificado

Las actividades del Proceso Unificado de Rational destacan en la creación y el mantenimiento de modelos más que documentos sobre papel. Estos modelos proporcionan representaciones semánticas del sistema software que se está desarrollando. Además, estos modelos se basan en los conceptos de objeto y clase y las relaciones entre ellos, y utilizan UML como la notación común. La razón



subyacente al interés que pone el Proceso Unificado en los modelos, antes que en los documentos sobre papel, es minimizar la sobrecarga asociada con la generación y el mantenimiento de los documentos y maximizar el contenido de información relevante.

El desarrollo bajo el Proceso Unificado está *centrado en la arquitectura*. El proceso se centra en establecer al principio una arquitectura software que guía el desarrollo del sistema. Con ello se facilita el desarrollo en paralelo, se minimiza la repetición de trabajos y se incrementa la probabilidad de reutilización de componentes y el mantenimiento posterior del sistema. Este diseño arquitectónico sirve como una sólida base sobre la cual se puede planificar y manejar el desarrollo de software basado en componentes.

Las actividades de desarrollo bajo el Proceso Unificado están *dirigidas por los casos de uso*. El Proceso Unificado pone un gran énfasis en la construcción de sistemas basada en una amplia comprensión de cómo se utilizará el sistema que se entregue. Las nociones de los casos de uso y los escenarios se utilizan para guiar el flujo de procesos desde la captura de los requisitos hasta las pruebas, y para proporcionar caminos que se pueden reproducir durante el desarrollo del sistema.

El Proceso Unificado es un *proceso configurable*. Aunque un único proceso no es adecuado para todas las organizaciones de desarrollo de software, el Proceso Unificado es adaptable y puede configurarse para cubrir las necesidades de proyectos que van desde pequeños equipos de desarrollo de software hasta grandes empresas de desarrollo. También se basa en una arquitectura de proceso simple y clara, que proporciona un marco común a toda una familia de procesos y que, además, puede variarse para acomodarse a distintas situaciones. Dentro del propio Proceso Unificado se encuentran las guías sobre cómo configurar el proceso para adaptarse a las necesidades de una organización.

El Proceso Unificado soporta las *técnicas orientadas a objetos*. Cada modelo es orientado a objetos. Los modelos del Proceso Unificado se basan en los conceptos de objeto y clase y las relaciones entre ellos, y utilizan UML como la notación común.

El Proceso Unificado impulsa un *control de calidad* y una *gestión del riesgo* objetivos y continuos. La evaluación de la calidad va contenida en el proceso, en todas las actividades, e implicando a todos los participantes, mediante medidas y criterios objetivos. No se trata como algo a posteriori o una actividad separada. La gestión del riesgo va contenida en el proceso, de manera que los riesgos para el éxito del proyecto se identifican y se acometen al principio del proceso de desarrollo, cuando todavía hay tiempo de reaccionar.

El Proceso Unificado tiene una estructura matricial donde se relacionan esfuerzos y tiempos. Los tiempos están definidos por las fases y las iteraciones. Los esfuerzos están definidos por los flujos de trabajo del proceso y de soporte. En las secciones siguientes se describirán estos conceptos.

### 11.3.2 Fases e iteraciones

Una *fase* es el intervalo de tiempo entre dos hitos importantes del proceso durante el que se cumple un conjunto bien definido de objetivos, se completan artefactos y se toman decisiones sobre si pasar o no a la siguiente fase. El Proceso Unificado de Rational consta de las cuatro fases siguientes (Figura 12.16): iniciación, elaboración, construcción y transición. Las fases de iniciación y elaboración incluyen las actividades de diseño del ciclo de vida del desarrollo. Las fases de construcción y transición constituyen su producción.

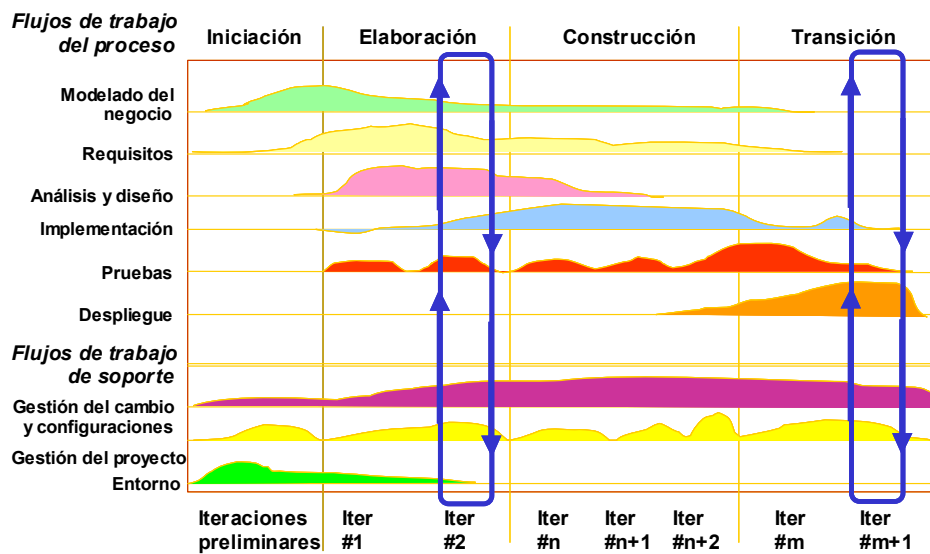


Figura 12.16. El ciclo de vida del desarrollo del software

Dentro de cada fase hay varias iteraciones. Una *iteración* representa un ciclo de desarrollo completo, desde la captura de requisitos en el análisis hasta la implementación y pruebas, que produce como resultado la entrega al cliente o la salida al mercado de un proyecto ejecutable. Cada iteración pasa a través de varios flujos de trabajo del proceso, aunque con un énfasis diferente en cada uno de ellos, dependiendo de la fase en que se encuentre. Durante la iniciación, el interés se

orienta hacia el análisis y el diseño. Durante la construcción, la actividad central es la implementación, y la transición se centra en despliegue.

El paso a través de las cuatro fases principales constituye un *ciclo de vida del desarrollo*, y produce una generación de software. La primera pasada a través de las cuatro fases se denomina ciclo de desarrollo inicial. A menos que acabe la vida del producto, un producto existente evolucionará a la siguiente generación repitiendo la misma secuencia de inicio, elaboración, construcción y transición. Esta es la evolución del sistema, así que los ciclos de desarrollo después del ciclo inicial son los ciclos de evolución (Figura 12.17).

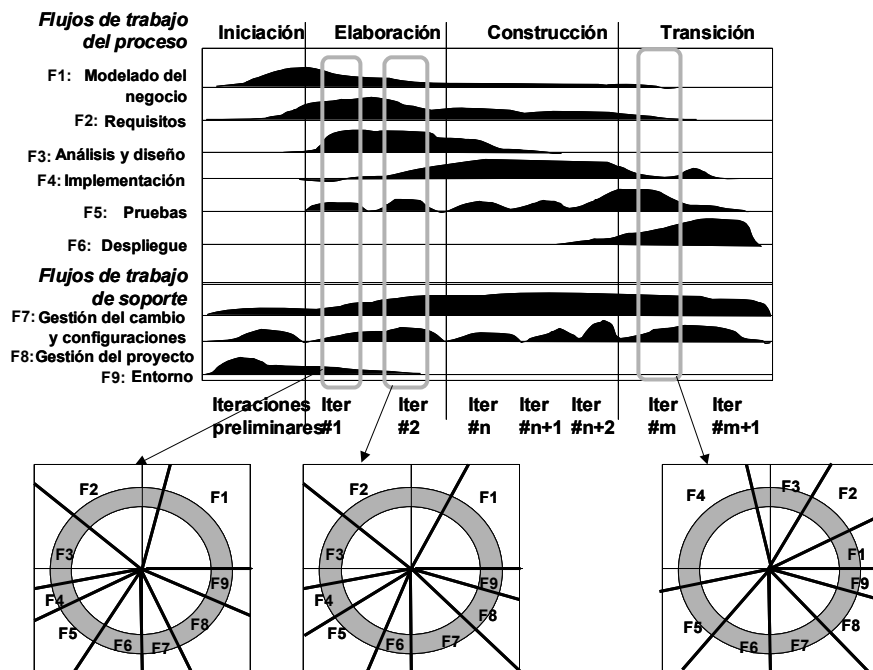


Figura 12.17. Las iteraciones son distintas en el ciclo de vida

Cada fase e iteración se centra en disminuir algún riesgo y concluye con un hito bien definido. La revisión de hitos es el momento adecuado para evaluar cómo se están satisfaciendo los objetivos y si el proyecto necesita ser reestructurado de alguna forma para continuar. A continuación se describe cada una de las fases:

1. **Iniciación.** Durante la fase de iniciación, se establece la planificación del proyecto y se delimita su alcance. La planificación del proyecto incluye los criterios de éxito, la evaluación del riesgo, estimaciones de

recursos que se necesitarán y un plan de fases que muestre la planificación de los hitos principales. Durante la iniciación, es frecuente crear un prototipo ejecutable que sirva para probar los conceptos. Al final de la fase de iniciación se examinan los objetivos del ciclo de vida del proyecto y se decide si proceder con el desarrollo del sistema.

2. **Elaboración.** Los objetivos de la fase de elaboración son analizar el dominio del problema, establecer una base arquitectónica sólida, desarrollar el plan del proyecto y eliminar los elementos de más alto riesgo del proyecto. Las decisiones arquitectónicas deben tomarse con una comprensión del sistema global. Esto implica que se deben describir la mayoría de los requisitos del sistema. Para verificar la arquitectura, se implementa un sistema que demuestre las distintas posibilidades de la arquitectura y ejecute los casos de uso significativos. Al final de la fase de elaboración se examinan el alcance y los objetivos del sistema, la elección de la arquitectura y la resolución de los riesgos más grandes, y se decide si se debe pasar a la construcción.
3. **Construcción.** Durante la fase de construcción, se desarrolla de forma iterativa e incremental un producto completo que está preparado para la transición hacia la comunidad de usuarios. Esto implica describir los requisitos restantes y los criterios de aceptación, refinando el diseño y completando la implementación y las pruebas del software. Al final de la fase de construcción se decide si el software, los lugares donde se instalará y los usuarios están todos preparados para empezar a funcionar.
4. **Transición.** Durante la fase de transición, el software se despliega en la comunidad de usuarios. Una vez que el sistema ha sido puesto en manos de los usuarios finales, a menudo aparecen cuestiones que requieren un desarrollo adicional para ajustar el sistema, corregir algunos problemas no detectados o finalizar algunas características que habían sido pospuestas. Esta fase comienza normalmente con una versión beta del sistema, que luego será reemplazada con el sistema de producción. Al final de la fase de transición se decide si se han satisfecho los objetivos del ciclo de vida del proyecto, y se determina si se debería empezar otro ciclo de desarrollo. Este es también un punto en el que se asimilan las lecciones aprendidas en el proyecto para mejorar el proceso de desarrollo, que será aplicado al próximo proyecto.

### 11.3.3 Los flujos de trabajo del proceso

El Proceso Unificado de Rational consta de nueve flujos de trabajo que son los siguientes (Figura 5.1):

1. **Modelado del negocio:** describe la estructura y la dinámica de la organización.
2. **Requisitos:** describe el método basado en casos de uso para extraer los requisitos.
3. **Análisis y diseño:** describe las diferentes vistas arquitectónicas.
4. **Implementación:** tiene en cuenta el desarrollo de software, la prueba de unidades y la integración.
5. **Pruebas:** describe los casos de pruebas, los procedimientos y las métricas para evaluación de defectos.
6. **Despliegue:** cubre la configuración del sistema entregable.
7. **Gestión de configuraciones:** controla los cambios y mantiene la integridad de los artefactos de un proyecto.
8. **Gestión del Proyecto:** describe varias estrategias de trabajo en un proceso iterativo.
9. **Entorno:** cubre la infraestructura necesaria para desarrollar un sistema.

Dentro de cada flujo de trabajo del proceso hay un conjunto de artefactos y actividades relacionados. Un artefacto es algún documento, informe o ejecutable que se produce, se manipula o se consume. Una actividad describe las tareas (pasos de concepción, realización y revisión) que llevan a cabo los trabajadores para crear o modificar los artefactos, junto con las técnicas y guías para ejecutar las tareas, incluyendo quizá el uso de herramientas para ayudar a automatizar algunas de ellas.

### 11.3.4 Artefactos

Cada actividad del Proceso Unificado de Rational lleva algunos artefactos asociados, bien sean requeridos como entradas, bien sean generados como salidas. Algunos artefactos se utilizan como entradas directas en las actividades siguientes, se mantienen como recursos de referencia en el proyecto, o se generan en algún formato específico, en forma de entregas definidas en el contrato.

Estos artefactos son adicionales a los que proporciona el propio UML y fundamentalmente son los modelos y los conjuntos. A continuación se describen ambos con brevedad.

#### 10.3.4.1 Artefactos modelos

Los *modelos* son el tipo de artefacto más importante en el Proceso Unificado de Rational. Hay nueve modelos que en conjunto cubren todas las decisiones importantes implicadas en la visualización, especificación, construcción y documentación de un sistema con gran cantidad de software. Son los siguientes:

10. **Modelo del negocio:** establece una abstracción de la organización.
11. **Modelo del dominio:** establece el contexto del sistema.
12. **Modelo de casos de uso:** establece los requisitos funcionales del sistema.
13. **Modelo de análisis** (opcional): establece un diseño de las ideas.
14. **Modelo de diseño:** establece el vocabulario del problema y su solución.
15. **Modelo del proceso** (opcional): establece los mecanismos de concurrencia y sincronización del sistema.
16. **Modelo de despliegue:** establece la topología hardware sobre la cual se ejecutará el sistema.
17. **Modelo de implementación:** establece las partes que se utilizarán para ensamblar y hacer disponible el sistema físico.
18. **Modelo de pruebas:** establece las formas de validar y verificar el sistema.

En cada uno de los flujos de trabajo del ciclo de vida del desarrollo del software se trabaja con los modelos descritos, pero no con todos al mismo tiempo, sino siguiendo una secuencia lógica determinada por el flujo de trabajo y la naturaleza del modelo. En la Tabla 12.10 se muestra qué modelos se manejan en cada uno de los flujos de trabajo del proceso de desarrollo.

Por otra parte, el Proceso Unificado recupera el concepto de vista que se ha definido previamente en UML. Para el Proceso Unificado una *vista* es una proyección de un modelo. Y la arquitectura de un sistema se captura en forma de cinco vistas que interactúan entre sí (Figura 12.18): la vista de diseño, la vista de procesos, la vista de despliegue, la vista de implementación y la vista de casos de uso.

Tabla 12.10. Modelos y flujos de trabajo del Proceso Unificado

|                          | Modelado del Negocio | Requisitos | Análisis | Diseño | Implementación | Prueba | Despliegue |
|--------------------------|----------------------|------------|----------|--------|----------------|--------|------------|
| Modelo del Negocio       | X                    |            |          |        |                |        |            |
| Modelo del Dominio       | X                    | X          |          |        |                |        |            |
| Modelo de Casos de Uso   |                      | X          |          |        |                |        |            |
| Modelo de Análisis       |                      |            | X        |        |                |        |            |
| Modelo de Diseño         |                      |            |          | X      |                |        |            |
| Modelo de Procesos       |                      |            |          | X      |                |        |            |
| Modelo de Despliegue     |                      |            |          | X      |                |        | X          |
| Modelo de Implementación |                      |            |          |        | X              |        | X          |
| Modelo de Prueba         |                      |            |          |        |                | X      | X          |

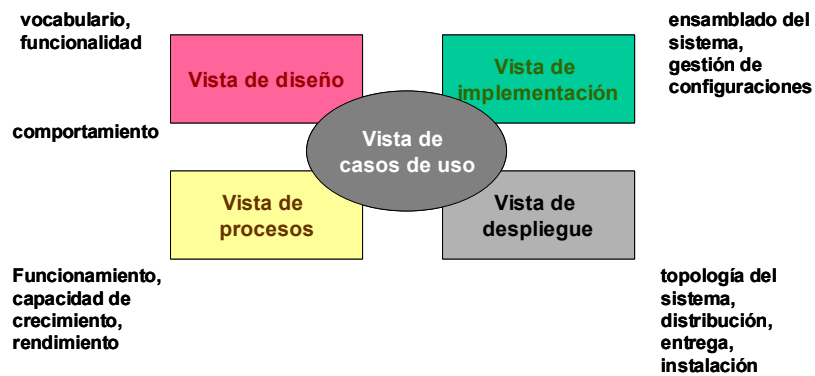


Figura 12.18. Vistas de la arquitectura de un sistema

En la Tabla 12.11 se presenta la correspondencia de los modelos con los flujos de trabajo del proceso del ciclo de vida del software del Proceso Unificado. Pero, para simplificar, se muestran solamente los aspectos relacionados directamente con el desarrollo técnico del proyecto, obviando los flujos de modelado del negocio y despliegue. En consecuencia, se han eliminado los modelos del negocio, del dominio y de procesos, ya que no están los flujos correspondientes.

Tabla 12.11. Modelos y flujos de trabajo del proceso: desarrollo técnico

|                          | Requisitos | Análisis | Diseño | Implementación | Prueba |
|--------------------------|------------|----------|--------|----------------|--------|
| Modelo de Casos de Uso   | X          |          |        |                |        |
| Modelo de Análisis       |            | X        |        |                |        |
| Modelo de Diseño         |            |          | X      |                |        |
| Modelo de Despliegue     |            |          | X      |                |        |
| Modelo de Implementación |            |          |        | X              |        |
| Modelo de Prueba         |            |          |        |                | X      |

El Proceso Unificado necesita por lo menos una herramienta CASE-OO de soporte para poder aplicarse con eficiencia, sea en el modelado, sea en la generación de código y en el diseño de las pruebas. En concreto, la herramienta utilizada en este libro es *Rational Rose Enterprise 2002 Edition* de *Rational Corp.* Donde se ha elegido el *framework* RUP.

#### 10.3.4.2 Otros Artefactos

Los artefactos del Proceso Unificado de Rational se clasifican en artefactos de gestión y artefactos técnicos. Los artefactos técnicos pueden dividirse en cuatro conjuntos principales:

1. *Conjunto de requisitos*: agrupa toda la información que describe lo que debe hacer el sistema. Esto puede comprender un modelo de casos de uso, un modelo de requisitos no funcionales, un modelo del dominio, un modelo de análisis y otras formas de expresión de las necesidades del usuario, incluyendo pero no limitándose a maquetas, prototipos de la interfaz, restricciones legales, etc.
2. *Conjunto de diseño*: agrupa información que describe cómo se va a construir el sistema y captura las decisiones acerca de cómo se va a realizar, teniendo en cuenta las restricciones de tiempo, presupuesto, aplicaciones existentes, reutilización, objetivos de calidad y demás consideraciones. Esto puede implicar un modelo de diseño, un modelo de pruebas y otras formas de expresión de la naturaleza del sistema, incluyendo, pero no limitándose, a prototipos y arquitecturas ejecutables.
3. *Conjunto de implementación*: agrupa toda la información acerca de los elementos software que comprende el sistema, incluyendo, pero no limitándose, a código fuente en varios lenguajes de programación,



archivos de configuración, archivos de datos, componentes software, etcétera, junto con la información que describe cómo ensamblar el sistema.

4. *Conjunto de despliegue*: agrupa toda la información acerca de la forma en que se empaqueta actualmente el software, se distribuye, se instala y se ejecuta en el entorno destino.

## 12.4 Descripciones completas de los genes

En este anexo se incluyen las descripciones completas de los genes, según el formato GenBank que se puede obtener desde la base Nucleotide en la URL <http://www.ncbi.nlm.nih.gov/entrez>.

### 12.4.1 Descripción de CETP en formato GenBank

```

LOCUS      BC066354                993 bp    mRNA    linear    PRI 24-FEB-
2004
DEFINITION Homo sapiens chromosome 20 open reading frame 71, mRNA (cDNA
clone
           MGC:87618 IMAGE:4826101), complete cds.
ACCESSION  BC066354
VERSION    BC066354.1  GI:42491000
KEYWORDS   MGC.
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 993)
  AUTHORS  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M.,
Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F.,
Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

```

Myers, R.M., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED 12477932

REFERENCE 2 (bases 1 to 993)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (03-FEB-2004) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) &

Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA

94305 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and

Myers, R. M.

Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 167 Row: g Column: 7  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis.

FEATURES Location/Qualifiers

source 1..993  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:87618 IMAGE:4826101"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

gene 1..993  
 /gene="C20orf71"  
 /note="synonyms: bA49G10.4, SPLUNC3"  
 /db\_xref="LocusID:128861"

CDS 186..842  
 /gene="C20orf71"  
 /codon\_start=1  
 /product="C20orf71 protein"  
 /protein\_id="AAH66354.1"  
 /db\_xref="GI:42491001"

```

/db_xref="LocusID:128861"

/translation="MMCPLWRLLIFLGLLALPLAPHKQPWPGLAQHRDNKSTLARI
AQGLIKHNAESRIQNIHFGRDLNLSAQVAPGLVWGLISGRKHQQQESRSFDNNIVKM
CAHMSIVVEFWLEKDEFGRRLDLVIGKCDAEPPSSVHVAILTEAIPPKMNQFLYNLKENL
QKVLPHMVESQVCPLIGEILQGLDVKLLKSLIEQEAHEPHTHETSQPSACQAGESPS
"
misc_feature      459..713
                  /gene="C20orf71"
                  /note="LBP_BPI_CETP; Region: LBP / BPI / CETP family,
                  N-terminal domain. The N and C terminal domains of the
                  LBP/BPI/CETP family are structurally similar"
                  /db_xref="CDD:pfam01273"
ORIGIN
1 aatgtgagca agccctggtg gcagcgccag ggtccagtgc agcccctccc cacagcatgc
61 tgggggctaa ttctgatgtc atctttctgc agaaaacat tagaccatcc ctccagactg
121 ccaccctcaa agccgtctgc ccaggcccca tctgacactc ttgacatctg caggteccag
181 accctatgat gtgtccactc tggaggctcc tcactttcct cgggttgctg gccttgccct
241 tggcaccaca caagcagcct tggcctggcc tggcccaagc ccacagagac aacaaatcca
301 ccctggcaag aattattgct cagggcctca taaagcaca cgcagaaagc cgaattcaga
361 acatccactt tggggacaga ctgaatgcct cagcacaagt ggcgccaggg ctggtgggct
421 ggctaatacag cggcaggaaa caccagcagc agcaagagag caggtccttc gataacaaca
481 tcgtaaagat gtgtgcacat atgagcatcg ttgtggagtt ctggctggag aaagacgagt
541 ttggccggag ggatctggtg ataggcaaat gcgatgcaga gccagcagc gtccatgtgg
601 ccatacctcac tgaggctatc ccaccaaaaga tgaatcagtt tctctacaac ctcaaagaga
661 atctgcaaaa agttctccca cacatggtag aaagtcaggt atgtcctctg atcggtgaaa
721 tcctcgggca gctggatgtg aaactggtga aaagcctcat agaacaggag gctgctcatg
781 aaccaacca ccatgaaacc agccaaccct ctgcatgcca ggctggagag tccccagct
841 gacttctgct gatcagaagg aaagtccaca tcttgcaacc ttaagtctcc cttagagtgg
901 ggcttctgct accctaaaaa ctttacccca ggctctgtgg acataccatc ctctctaca
961 ataaactcta gctctggaag aaaaaaaaaa aaa
//

```

## 12.4.2 Descripción de APOE en formato GenBank

```

LOCUS      BC072022                1550 bp    mRNA     linear   PRI 02-JUN-
2004
DEFINITION Homo sapiens apolipoprotein E, mRNA (cDNA clone MGC:88735
IMAGE:6263100), complete cds.
ACCESSION  BC072022
VERSION    BC072022.1  GI:47938399
KEYWORDS   MGC.
SOURCE     Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1550)
  AUTHORS  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M.,
            Schuler,G.D.,

```

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F.,  
 Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,  
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J.,  
 Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED 12477932

REFERENCE 2 (bases 1 to 1550)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-2004) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA

94305

Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and  
 Myers,  
 R. M.

Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 58 Row: j Column: 6  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi:  
 4557324.

FEATURES Location/Qualifiers  
 source 1..1550  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:88735 IMAGE:6263100"  
 /tissue\_type="Skin, melanoma, melanotic"  
 /clone\_lib="NIH\_MGC\_112"

```

/lab_host="DH10B-R"
/note="Vector: pOTB7"
gene      1..1550
          /gene="APOE"
          /db_xref="LocusID:348"
          /db_xref="MIM:107741"
CDS       414..1367
          /gene="APOE"
          /codon_start=1
          /product="apolipoprotein E"
          /protein_id="AAH72022.1"
          /db_xref="GI:47938400"
          /db_xref="LocusID:348"
          /db_xref="MIM:107741"

/translation="MKVLWAALLVTFFLAGCQAKVEQAVETEPEPELRQQTEWQSGQRW
ELALGRFWDYLRWVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTP
VAEETRARLSKELQAAQARLGADMEDVCGRVLVQYRGEVQAMLGQSTEELRVRLASHLR
KLRKRLLRDADDLQKRLAVYQAGAREGAERGLSAIRERLGLPVEQGRVRAATVGS LAG
QPLQERAQAWGERLRARMEEMGSRTDRDLDEVKEQVAEVRAKLEEQAQQIRLQAEAFQ
ARLKSWFEP LVEDMQRQWAGLVEKVQAAVGTSAAPVPSDNH"
ORIGIN
  1  cagccttaga aaagccctca atgtaaacct tatgtttggt actaggggtg cctctctccc
 61  atttcccgct gtggctgagc caaacttcat tactttcctt aggccctctg cccctccccct
121  agcatatggc tctttctcct attgaacaga aaattgagac tatgaagagg tcaacttgta
181  tttaaccaact ttactcattt tcccatttta gagaaaaga ggttgctgctg ttcctatcaa
241  aggcgactct gtgctacatg ttggtttcat cttcgctgct tctctcattt cccacaaaag
301  tgttgcttca ctttctgcct accttttagg tattgatcct cactcctcgt gccgaattcg
361  gcacgagggg acgtccttcc ccaggagccg actggccaat cacaggcagg aagatgaagg
421  ttctgtgggc tgcgttgctg gtcacattcc tggcaggatg ccaggccaag gtggagcaag
481  cggtgagac agagccggag cccgagctgc gccagcagac cgagtggcag agcggccagc
541  gctgggaact ggcactgggt cgcttttggg attacctgcg ctgggtgcag acactgtctg
601  agcaggtgca ggaggagctg ctcagctccc aggtcaccca ggaactgagg gcgctgatgg
661  acgagaccat gaaggagtgt aaggcctaca aatcggaact ggaggaacaa ctgacccccg
721  tggcggagga gacggggca cggtgtcca aggagctgca gggggcgcag gcccggtgg
781  gcgcggacat ggaggacgtg tgcggccgcc tgggtcagta ccgcggcgag gtgcaggcca
841  tgctcggcca gagcaccgag gagctgctgg tgcgcctcgc ctcccacctg cgcaagctgc
901  gtaagcggct cctcccgcat gccgatgacc tgcagaagcg cctggcagtg taccaggccg
961  gggcccgcga gggcgccgag cgcggcctca gcgccatccg cgagcgctg gggcccctgg
1021  tggaacaggg ccgctgctgg gccgccactg tgggctccct ggccggccag ccgctacagg
1081  agcgggceca ggccctggggc gagcggctgc gcgcgcggat ggaggagatg ggcagccgga
1141  cccgcgaccg cctggacgag gtgaaggagc aggtgcgga ggtgcgcgcc aagctggagg
1201  agcaggceca gcagatacgc ctgcaggccg aggccttcca ggcccgcctc aagagctggt
1261  tcgagcccct ggtggaagac atgcagcgc agtggcccg gctggtggag aaggtgcagg
1321  ctgccgtggg caccagcgc gccctgtgct ccagcgacaa tcaactgaac ccgaagcctg
1381  cagccatgcg accccacgcc acccctgccc tctgctcctc gcgcagcctg cagcgggaga
1441  ccctgtcccc gccccagccg tcctcctggg gtggacccta gtttaataaa gattcaccaa
1501  gtttcacgca aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa
//

```

### 12.4.3 Descripción de APOA1 en formato GenBank

LOCUS BC005380 926 bp mRNA linear PRI 03-OCT-2003

DEFINITION Homo sapiens apolipoprotein A-I, mRNA (cDNA clone MGC:12499 IMAGE:3934992), complete cds.

ACCESSION BC005380

VERSION BC005380.1 GI:13529241

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 926)

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 926)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA

94305

Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and  
Myers,  
R. M.

Clone distribution: MGC clone distribution information can be  
found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 16 Row: 1 Column: 23  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi:  
4557320.  
Differences found between this sequence and the human genome  
(build  
34) are described in misc\_difference features below.

FEATURES  
source Location/Qualifiers  
1..926  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:12499 IMAGE:3934992"  
/tissue\_type="Skeletal Muscle"  
/clone\_lib="NIH\_MGC\_81"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
gene 1..926  
/gene="APOA1"  
/db\_xref="LocusID:335"  
/db\_xref="MIM:107680"  
CDS 39..842  
/gene="APOA1"  
/codon\_start=1  
/product="apolipoprotein A-I precursor"  
/protein\_id="AAH05380.1"  
/db\_xref="GI:13529242"  
/db\_xref="LocusID:335"  
/db\_xref="MIM:107680"

/translation="MKA AVLTLAVLFLTGSQARHFWQQDEPPQSPWDRVKDLATVYVD  
VLKDSGRDYVSQFEGSALGKQLNLKLLDNWDSVTSTFSKLRQLGPVTQEFWDNLEKE  
TEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHE  
LQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELQRQLAARLEALKENGGARLAEYHA  
KATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKLNTQ"  
misc\_feature 81..833  
/gene="APOA1"  
/note="Apolipoprotein; Region: Apolipoprotein A1/A4/E  
family. These proteins contain several 22 residue  
repeats  
which form a pair of alpha helices. This family  
includes:  
Apolipoprotein A-I. Apolipoprotein A-IV. Apolipoprotein  
E"  
misc\_difference /db\_xref="CDD:pfam01442"  
896  
/gene="APOA1"

```

                                /note="'C' in cDNA is 'G' in genome."
misc_difference 900..926
                                /gene="APOA1"
                                /note="polyA tail: 27 bases do not align to the
genome."
ORIGIN
  1 agagactgcg agaaggaggt cccccacggc ctttcaggat gaaagctgcg gtgctgacct
  61 tggccgtgct cttcctgacg gggagccagg ctcggcattt ctggcagcaa gatgaacccc
 121 cccagagccc ctgggatcga gtgaaggacc tggcactgt gtacgtggat gtgctcaaag
 181 acagcggcag agactatgtg tcccagtttg aaggctccgc ctgggaaaaa cagctaaacc
 241 taaagctcct tgacaactgg gacagcgtga cctccacctt cagcaagctg cggaacagc
 301 tcggccctgt gaccaggag tcttgggata acctggaaaa ggagacagag ggcctgaggc
 361 aggagatgag caaggatctg gaggaggtga aggccaaggt gcagccctac ctggacgact
 421 tccagaagaa gtggcaggag gagatggagc tctaccgcc aaggtggag cgcctgctgcg
 481 cagagctcca agagggcgcg cgcagaagc tgcacgagct gcaagagaag ctgagccac
 541 tgggcgagga gatgcgcgac cgcgcgcgcg cccatgtgga cgcgctgctg acgcatctgg
 601 cccctacag cgacgagctg cgcagcgcct tggcgcgcg ccttgaggct ctcaaggaga
 661 acgcgggcgc cagactggcc cagtaccacg ccaaggccac cgagcatctg agcacgctca
 721 gcgagaaggc caagcccgcg ctcgaggacc tccgccaagg cctgctgccc gtgctggaga
 781 gcttcaaggt cagcttctcg agcgtctctg aggagtacac taagaagctc aacaccagt
 841 gaggcgcccg ccgcccggc ccttcccggg gctcagaata aacgtttcca aagtgcgaaa
 901 aaaaaaaaaa aaaaaaaaaa aaaaaa
//

```

## 12.4.4 Descripción de LIPC en formato GenBank

```

LOCUS       NM_000236                1603 bp    mRNA    linear    PRI 20-DEC-
2003
DEFINITION  Homo sapiens lipase, hepatic (LIPC), mRNA.
ACCESSION   NM_000236
VERSION     NM_000236.1  GI:4557722
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1603)
  AUTHORS   Chen,W., Srinivasan,S.R., Boerwinkle,E. and Berenson,G.S.
  TITLE     Hepatic lipase promoter C-514T polymorphism influences serial
            changes in HDL cholesterol levels since childhood: the Bogalusa
            Heart Study
  JOURNAL   Atherosclerosis 169 (1), 175-182 (2003)
  PUBMED   12860265
  REMARK    GeneRIF: The HL gene may play an important role in the
regulation
            of HDL-C levels from childhood to adulthood, especially in white
            males.
REFERENCE   2  (bases 1 to 1603)
  AUTHORS   Andersen,R.V., Wittrup,H.H., Tybjaerg-Hansen,A., Steffensen,R.,
            Schnohr,P. and Nordestgaard,B.G.
  TITLE     Hepatic lipase mutations,elevated high-density lipoprotein
            cholesterol, and increased risk of ischemic heart disease: the

```



- Copenhagen City Heart Study  
 JOURNAL J. Am. Coll. Cardiol. 41 (11), 1972-1982 (2003)  
 PUBMED 12798568  
 REMARK GenerIF: Hepatic lipase promoter SNPs are associated with  
 increased HDL cholesterol and, paradoxically, an increased risk of IHD  
 after adjustment for HDL cholesterol, and particularly in individuals  
 with apolipoprotein E epsilon43 genotype.  
 REFERENCE 3 (bases 1 to 1603)  
 AUTHORS Zambon,A., Deeb,S.S., Pauletto,P., Crepaldi,G. and Brunzell,J.D.  
 TITLE Hepatic lipase: a marker for cardiovascular disease risk and  
 response to therapy  
 JOURNAL Curr. Opin. Lipidol. 14 (2), 179-189 (2003)  
 PUBMED 12642787  
 REMARK GenerIF: REVIEW: potential impact of genetic determinants of  
 hepatic lipase activity in modulating both the development of  
 coronary and carotid atherosclerosis will be discussed based on  
 hepatic lipase proposed roles in lipoprotein metabolism  
 REFERENCE 4 (bases 1 to 1603)  
 AUTHORS Su,Z.G., Zhang,S.Z., Hou,Y.P., Zhang,L., Huang,D.J., Liao,L.C.  
 and Xiao,C.Y.  
 TITLE Relationship between a novel polymorphism of hepatic lipase gene  
 and coronary artery disease  
 JOURNAL Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao 34 (6), 780-785  
 (2002)  
 PUBMED 12417924  
 REMARK GenerIF: Results suggest that a T right curved arrow C  
 substitution at -2 of the HL promoter may be associated with the variation  
 of HDL-cholesterol concentration and therefore affect the risk of  
 CAD in Chinese.  
 REFERENCE 5 (bases 1 to 1603)  
 AUTHORS Su,Z., Zhang,S., Nebert,D.W., Zhang,L., Huang,D., Hou,Y.,  
 Liao,L. and Xiao,C.  
 TITLE A novel allele in the promoter of the hepatic lipase is  
 associated with increased concentration of HDL-C and decreased promoter  
 activity  
 JOURNAL J. Lipid Res. 43 (10), 1595-1601 (2002)  
 PUBMED 12364543  
 REMARK GenerIF: LIPC promoter is associated with a lowered HL activity  
 and that this variation may contribute to the increased plasma HDL-C  
 concentration  
 REFERENCE 6 (bases 1 to 1603)  
 AUTHORS Jansen,H., Verhoeven,A.J. and Sijbrands,E.J.  
 TITLE Hepatic lipase: a pro- or anti-atherogenic protein?  
 JOURNAL J. Lipid Res. 43 (9), 1352-1362 (2002)  
 PUBMED 12235167  
 REMARK GenerIF: REVIEW: role in coronary artery disease  
 REFERENCE 7 (bases 1 to 1603)  
 AUTHORS Gonzalez-Navarro,H., Nong,Z., Freeman,L., Bensadoun,A.,  
 Peterson,K. and Santamarina-Fojo,S.

- TITLE Identification of mouse and human macrophages as a site of synthesis of hepatic lipase  
 JOURNAL J. Lipid Res. 43 (5), 671-675 (2002)  
 PUBMED 11971936  
 REMARK GeneRIF: synthesized in peritoneal macrophages  
 REFERENCE 8 (bases 1 to 1603)  
 AUTHORS Yamakawa-Kobayashi,K., Somekawa,Y., Fujimura,M., Tomura,S., Arinami,T. and Hamaguchi,H.
- TITLE Relation of the -514C/T polymorphism in the hepatic lipase gene to serum HDL and LDL cholesterol levels in postmenopausal women under hormone replacement therapy  
 JOURNAL Atherosclerosis 162 (1), 17-21 (2002)  
 PUBMED 11947893  
 REMARK GeneRIF: The -514C/T polymorphism of the HL gene was found to be associated with variations in hepatic lipase activity and serum HDL-C levels.  
 REFERENCE 9 (bases 1 to 1603)  
 AUTHORS Ji,J., Herbison,C.E., Mamotte,C.D., Burke,V., Taylor,R.R. and van Bockxmeer,F.M.
- TITLE Hepatic lipase gene -514 C/T polymorphism and premature coronary heart disease  
 JOURNAL J Cardiovasc Risk 9 (2), 105-113 (2002)  
 PUBMED 12006918  
 REMARK GeneRIF: The T allele of the hepatic lipase-514 C/T polymorphism is related to changes in plasma lipids. The superficially paradoxical predisposition to CHD in males is attributable to impairment of TG rich lipoprotein metabolism and reverse cholesterol transport.  
 REFERENCE 10 (bases 1 to 1603)  
 AUTHORS Santamarina-Fojo,S., Haudenschild,C. and Amar,M.  
 TITLE The role of hepatic lipase in lipoprotein metabolism and atherosclerosis  
 JOURNAL Curr. Opin. Lipidol. 9 (3), 211-219 (1998)  
 PUBMED 9645503  
 REFERENCE 11 (bases 1 to 1603)  
 AUTHORS Tiebel,O., Gehrlich,S., Pietzsch,J., Gromeier,S. and Jaross,W.  
 TITLE 18 bp insertion/duplication with internal missense mutation in human hepatic lipase gene exon 3. Mutations in brief no. 181. Online  
 JOURNAL Hum. Mutat. 12 (3), 216 (1998)  
 PUBMED 10660332  
 REFERENCE 12 (bases 1 to 1603)  
 AUTHORS Takagi,A., Ikeda,Y., Mori,A., Ashida,Y. and Yamamoto,A.  
 TITLE Identification of a BstNI polymorphism in exon 9 of the human hepatic triglyceride lipase gene  
 JOURNAL Mol. Cell. Probes 10 (4), 313-314 (1996)  
 PUBMED 8865180  
 REFERENCE 13 (bases 1 to 1603)  
 AUTHORS Mori,A., Takagi,A., Ikeda,Y., Ashida,Y. and Yamamoto,A.  
 TITLE An AvaII polymorphism in exon 5 of the human hepatic triglyceride lipase gene  
 JOURNAL Mol. Cell. Probes 10 (4), 309-311 (1996)  
 PUBMED 8865179

REFERENCE 14 (bases 1 to 1603)  
 AUTHORS Cai,S.J., Wong,D.M., Chen,S.H. and Chan,L.  
 TITLE Structure of the human hepatic triglyceride lipase gene  
 JOURNAL Biochemistry 28 (23), 8966-8971 (1989)  
 PUBMED 2605236

REFERENCE 15 (bases 1 to 1603)  
 AUTHORS Martin,G.A., Busch,S.J., Meredith,G.D., Cardin,A.D.,  
 Blankenship,D.T., Mao,S.J., Rechten,A.E., Woods,C.W.,  
 Racke,M.M.,  
 Schafer,M.P. et al.  
 TITLE Isolation and cDNA sequence of human postheparin plasma hepatic  
 triglyceride lipase  
 JOURNAL J. Biol. Chem. 263 (22), 10907-10914 (1988)  
 PUBMED 2839510

REFERENCE 16 (bases 1 to 1603)  
 AUTHORS Datta,S., Luo,C.C., Li,W.H., VanTuinen,P., Ledbetter,D.H.,  
 Brown,M.A., Chen,S.H., Liu,S.W. and Chan,L.  
 TITLE Human hepatic lipase. Cloned cDNA sequence, restriction fragment  
 length polymorphisms, chromosomal localization, and evolutionary  
 relationships with lipoprotein lipase and pancreatic lipase  
 JOURNAL J. Biol. Chem. 263 (3), 1107-1110 (1988)  
 PUBMED 2447084

REFERENCE 17 (bases 1 to 1603)  
 AUTHORS Sparkes,R.S., Zollman,S., Klisak,I., Kirchgessner,T.G.,  
 Komaromy,M.C., Mohandas,T., Schotz,M.C. and Lusic,A.J.  
 TITLE Human genes involved in lipolysis of plasma lipoproteins:  
 mapping  
 of loci for lipoprotein lipase to 8p22 and hepatic lipase to  
 15q21  
 JOURNAL Genomics 1 (2), 138-144 (1987)  
 PUBMED 3692485

REFERENCE 18 (bases 1 to 1603)  
 AUTHORS Stahnke,G., Sprengel,R., Augustin,J. and Will,H.  
 TITLE Human hepatic triglyceride lipase: cDNA cloning, amino acid  
 sequence and expression in a cultured cell line  
 JOURNAL Differentiation 35 (1), 45-52 (1987)  
 PUBMED 2828141

COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The  
 reference sequence was derived from X07228.1.  
  
 Summary: LIPC encodes hepatic triglyceride lipase, which is  
 expressed in liver. LIPC has the dual functions of triglyceride  
 hydrolase and ligand/bridging factor for receptor-mediated  
 lipoprotein uptake.

FEATURES Location/Qualifiers  
 source 1..1603  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15q21-q23"  
 gene 1..1603  
 /gene="LIPC"  
 /note="synonyms: HL, HTGL, LIPH"  
 /db\_xref="GeneID:3990"  
 /db\_xref="LocusID:3990"  
 /db\_xref="MIM:151670"  
 CDS 58..1557  
 /gene="LIPC"

```

/EC_number="3.1.1.3"
/note="go_function: triacylglycerol lipase activity
[goid
0004806] [evidence IEA];
go_function: heparin binding [goid 0008201] [evidence
IEA];
go_function: lipid transporter activity [goid 0005319]
[evidence IEA];
go_function: hydrolase activity [goid 0016787]
[evidence
IEA];
go_process: N-linked glycosylation [goid 0006487]
[evidence TAS] [pmid 2447084];
go_process: lipid catabolism [goid 0016042] [evidence
IEA]"
/codon_start=1
/product="lipase C precursor"
/protein_id="NP_000227.1"
/db_xref="GI:4557723"
/db_xref="GeneID:3990"
/db_xref="LocusID:3990"
/db_xref="MIM:151670"

/translation="MDTSPLCFSELLVLCIFIQSSALGQSLKPEPFGRRAQAVETNKT
LHEMKTRFLFLFGETNQGCQIRINHPDTLQECGFNSSLPLVMI IHGWSVDGVLENWIWQ
MVAALKSQPAQPVNVGLVDWITLAHDHYTIAVRNTRLVGKEVAALLRWLEESVQLSRS
HVHLIGYSLGAHVSGFAGSSIGGTHKIGRITGLDAAGPLFEGSAPSNRLSPDDANFVD
AIHTFTREHMGLSVGIKQPIGHYDFYPNGGSFQPGCHFLELYRHIAQHGFNAITQTIK
CSHERSVHLFIDSLLHAGTQSMAYPCGDMNSFSQGLCLSCCKKRCNTLGYHVRQEPRS
KSKRLFLVTRAQSPFKVYHYQLKIQFINQTETPIQTTFTMSLLGTKEKMQKIPITLTK
GIASNKTYSLITLDVDIGELIMIKFKWENSAVWANVWDTVQTIIPWSTGPRHSGLVL
KTIRVKAGETQQRMTFCSENTDDLLLRPTQEKIFVKCEIKSKTSKRKIR"
sig_peptide 58..123
mat_peptide 124..1554
misc_feature 127..1104
misc_feature 1111..1518
barrel)
polyA_signal 1583..1588
ORIGIN
1 ggtctctttg gcttcagaaa ttaccaagaa agcctggacc cggggtgaaa cggagaaatg
61 gacacaagtc cctgtgtttt ctccattctg ttggttttat gcatctttat ccaatcaagt

```

```

121 gcccttgac aaagcctgaa accagagcca tttggaagaa gagctcaagc tgttgaaca
181 aacaaaacgc tgcattgat gaagaccaga ttctgtctct ttggagaaac caatcagggc
241 tgtcagattc gaatcaatca tccggacacg ttacaggagt gcggcttcaa ctccctcctg
301 cctctggtga tgataatcca cgggtggtcg gtggacggcg tgctagaaaa ctggatctgg
361 cagatggtgg ccgctgtaa gtctcagccg gccagccag tgaacgtggg gctggtggac
421 tggatcacc tggcccacga ccaatacacc atcgccgtcc gcaacacccg ccttgtgggc
481 aaggaggtcg cggctcttct cgggtggtcg gaggaatctg ttcaactctc tcgaagccat
541 gttcacctaa ttgggtacag cctgggtgca cacgtgtcag gatttgccgg cagttccatc
601 ggtggaacgc acaagattgg gagaatcaca gggctggatg ccgctggacc tttgtttgag
661 ggaagtgcgc ccagcaatcg tcttttcca gatgatgcca attttgtgga tgccattcat
721 acctttacgc gggagcacat gggcctgagc gtgggcatca aacagcccat aggacactat
781 gacttctatc ccaacggggg ctcttccag cctggctgcc acttcctaga gctctacaga
841 cataattgcc agcagggctt caatgccatc accagacca taaaatgctc ccacgagcga
901 tcggtgcacc tttcatcga ctcttctg cagccggca cgcagagcat ggcctaccgg
961 tgtggtgaca tgaacagctt cagccagggc ctgtgcctga gctgcaagaa gggccgctgc
1021 aacacgctgg gctaccacgt ccgcccaggag ccgcccaggca agagcaagag gctcttctc
1081 gtaacgctgg cccagtcctc cttcaaagtt tatcattacc agttaagat ccagttctc
1141 aaccaaactg agcagccaat acaacaact tttaccatgt cactactcgg aacaaagag
1201 aaaatgcaga aaattcccat cactctgggc aaaggaattg ctagtaataa aacgtattcc
1261 tttcttatca cgctggatgt ggatctggc gagctgatca tgatcaagtt caagtgggaa
1321 aacagtgcag tgtgggcca tgtctgggac acggtccaga ccatcatccc atggagcaca
1381 gggccgccc actcaggcct cgttctgaag acgatcagag tcaaagcagg agaaaaccag
1441 caaagaatga cattttgttc agaaaacaca gatgacctac tacttcgccc aaccaggaa
1501 aaaatcttgc tgaatgtga aataaagtct aaaacatcaa agcgaagat cagatgagat
1561 ttaatgaaga cccagtgtaa agaataaatg aatcttactc ctt
//

```

## 12.4.5 Descripción de SR-BI en formato GenBank

```

LOCUS       NM_005505                2566 bp    mRNA    linear    PRI 24-DEC-
2003
DEFINITION Homo sapiens scavenger receptor class B, member 1 (SCARB1),
mRNA.
ACCESSION   NM_005505
VERSION     NM_005505.3  GI:33620766
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2566)
  AUTHORS   Vishnyakova,T.G., Bocharov,A.V., Baranova,I.N., Chen,Z.,
            Remaley,A.T., Csako,G., Eggerman,T.L. and Patterson,A.P.
  TITLE     Binding and internalization of lipopolysaccharide by Cla-1, a
human
            orthologue of rodent scavenger receptor B1
  JOURNAL   J. Biol. Chem. 278 (25), 22771-22780 (2003)
  PUBMED   12651854
  REMARK    GeneRIF: binds and internalizes lipopolysaccharides and
lipoproteins and is a orthologue of rodent SR-BI
REFERENCE   2 (bases 1 to 2566)
  AUTHORS   Malerod,L., Sporstol,M., Juvet,L.K., Mousavi,A., Gjoen,T. and

```

- TITLE Berg, T.  
 Hepatic scavenger receptor class B, type I is stimulated by peroxisome proliferator-activated receptor gamma and hepatocyte nuclear factor 4alpha  
 JOURNAL Biochem. Biophys. Res. Commun. 305 (3), 557-565 (2003)  
 PUBMED 12763030  
 REMARK GenerIF: Transcriptional activation of SR-BI is stimulated by peroxisome proliferator-activated receptor gamma and hepatocyte nuclear factor 4alpha.
- REFERENCE 3 (bases 1 to 2566)  
 AUTHORS McCarthy, J.J., Lehner, T., Reeves, C., Moliterno, D.J., Newby, L.K., Rogers, W.J. and Topol, E.J.  
 TITLE Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease  
 JOURNAL J. Med. Genet. 40 (6), 453-458 (2003)  
 PUBMED 12807968  
 REMARK GenerIF: Genetic variants in the HDL receptor, SR-B1, may be an important determinant of abnormal lipoproteins in women and may confer particular susceptibility to coronary artery disease.
- REFERENCE 4 (bases 1 to 2566)  
 AUTHORS Osgood, D., Corella, D., Demissie, S., Cupples, L.A., Wilson, P.W., Meigs, J.B., Schaefer, E.J., Coltell, O. and Ordovas, J.M.  
 TITLE Genetic variation at the scavenger receptor class B type I gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: the framingham study  
 JOURNAL J. Clin. Endocrinol. Metab. 88 (6), 2869-2879 (2003)  
 PUBMED 12788901  
 REMARK GenerIF: SR-BI gene variation modulates the lipid profile, particularly in type 2 diabetes, contributing to the metabolic abnormalities.
- REFERENCE 5 (bases 1 to 2566)  
 AUTHORS Perez-Martinez, P., Ordovas, J.M., Lopez-Miranda, J., Gomez, P., Marin, C., Moreno, J., Fuentes, F., Fernandez de la Puebla, R.A. and Perez-Jimenez, F.  
 TITLE Polymorphism exon 1 variant at the locus of the scavenger receptor  
 class B type I gene: influence on plasma LDL cholesterol in healthy subjects during the consumption of diets with different fat contents  
 JOURNAL Am. J. Clin. Nutr. 77 (4), 809-813 (2003)  
 PUBMED 12663276  
 REMARK GenerIF: carriers of the minority allele, 1/2, of the scavenger receptor class B type I gene are more susceptible to the presence of saturated fatty acids in the diet because of a greater increase in LDL cholesterol
- REFERENCE 6 (bases 1 to 2566)  
 AUTHORS Strauss, J.G., Zimmermann, R., Hrzenjak, A., Zhou, Y., Kratky, D., Levak-Frank, S., Kostner, G.M., Zechner, R. and Frank, S.  
 TITLE Endothelial cell-derived lipase mediates uptake and binding of high-density lipoprotein (HDL) particles and the selective uptake of HDL-associated cholesterol esters independent of its enzymic activity  
 JOURNAL Biochem. J. 368 (Pt 1), 69-79 (2002)  
 PUBMED 12164779  
 REMARK GenerIF: EDL mediates both HDL binding and uptake, and the

- selective uptake of HDL-CE, independently of lipolysis and CLA-
1.  
 REFERENCE 7 (bases 1 to 2566)  
 AUTHORS Bultel-Brienne,S., Lestavel,S., Pilon,A., Laffont,I.,  
 Tailleux,A.,  
 Fruchart,J.C., Siest,G. and Clavey,V.  
 TITLE Lipid free apolipoprotein E binds to the class B Type I  
 scavenger  
 receptor I (SR-BI) and enhances cholesteryl ester uptake from  
 lipoproteins  
 JOURNAL J. Biol. Chem. 277 (39), 36092-36099 (2002)  
 PUBMED 12138091  
 REMARK GenerIF: SR-BI regulation of cholesteryl ester uptake is  
 affected  
 by binding to APOE in human cells
- REFERENCE 8 (bases 1 to 2566)  
 AUTHORS Silver,D.L.  
 TITLE A carboxyl-terminal PDZ-interacting domain of scavenger receptor  
 B,  
 type I is essential for cell surface expression in liver  
 JOURNAL J. Biol. Chem. 277 (37), 34042-34047 (2002)  
 PUBMED 12119305  
 REMARK GenerIF: PDZK1 or other PDZ domain proteins may play an  
 important  
 role in regulating SR-BI cell surface expression and hence  
 reverse  
 cholesterol transport.
- REFERENCE 9 (bases 1 to 2566)  
 AUTHORS Langer,C., Gansz,B., Goepfert,C., Engel,T., Uehara,Y., von  
 Dehn,G.,  
 Jansen,H., Assmann,G. and von Eckardstein,A.  
 TITLE Testosterone up-regulates scavenger receptor BI and stimulates  
 cholesterol efflux from macrophages  
 JOURNAL Biochem. Biophys. Res. Commun. 296 (5), 1051-1057 (2002)  
 PUBMED 12207878  
 REMARK GenerIF: Testosterone up-regulates scavenger receptor BI mRNA in  
 macrophages
- REFERENCE 10 (bases 1 to 2566)  
 AUTHORS Hong,S.H., Kim,Y.R., Yoon,Y.M., Min,W.K., Chun,S.I. and Kim,J.Q.  
 TITLE Association between HaeIII polymorphism of scavenger receptor  
 class  
 B type I gene and plasma HDL-cholesterol concentration  
 JOURNAL Ann Clin Biochem 39 (Pt 5), 478-481 (2002)  
 PUBMED 12227853  
 REMARK GenerIF: There is an association between HaeIII polymorphism of  
 scavenger receptor class B type I gene and plasma HDL-  
 cholesterol  
 concentration in patients with CAD.
- REFERENCE 11 (bases 1 to 2566)  
 AUTHORS Kawasaki,Y., Nakagawa,A., Nagaosa,K., Shiratsuchi,A. and  
 Nakanishi,Y.  
 TITLE Phosphatidylserine binding of class B scavenger receptor type I,  
 a  
 phagocytosis receptor of testicular sertoli cells  
 JOURNAL J. Biol. Chem. 277 (30), 27559-27566 (2002)  
 PUBMED 12016218  
 REMARK GenerIF: SR-BI is a phagocytosis-inducing PS receptor of Sertoli  
 cells.
- REFERENCE 12 (bases 1 to 2566)

AUTHORS Li,X.A., Titlow,W.B., Jackson,B.A., Giltiy,N.,  
Nikolova-Karakashian,M., Uittenbogaard,A. and Smart,E.J.

TITLE High density lipoprotein binding to scavenger receptor, Class B,  
type I activates endothelial nitric-oxide synthase in a  
ceramide-dependent manner

JOURNAL J. Biol. Chem. 277 (13), 11058-11063 (2002)

PUBMED 11792700

REMARK GeneRIF: binding to high density lipoprotein activates  
endothelial  
nitric-oxide synthase in a ceramide-dependent manner

REFERENCE 13 (bases 1 to 2566)

AUTHORS Murao,K., Terpstra,V., Green,S.R., Kondratenko,N., Steinberg,D.  
and  
Quehenberger,O.

TITLE Characterization of CLA-1, a human homologue of rodent scavenger  
receptor BI, as a receptor for high density lipoprotein and  
apoptotic thymocytes

JOURNAL J. Biol. Chem. 272 (28), 17551-17557 (1997)

PUBMED 9211901

REFERENCE 14 (bases 1 to 2566)

AUTHORS Calvo,D., Dopazo,J. and Vega,M.A.

TITLE The CD36, CLA-1 (CD36L1), and LIMP2 (CD36L2) gene family:  
cellular  
distribution, chromosomal location, and genetic evolution

JOURNAL Genomics 25 (1), 100-106 (1995)

PUBMED 7539776

REFERENCE 15 (bases 1 to 2566)

AUTHORS Calvo,D. and Vega,M.A.

TITLE Identification, primary structure, and distribution of CLA-1, a  
novel member of the CD36/LIMP2 gene family

JOURNAL J. Biol. Chem. 268 (25), 18929-18935 (1993)

PUBMED 7689561

REFERENCE 16 (bases 1 to 2566)

AUTHORS Skre,H. and Berg,K.

TITLE Cerebellar ataxia and total albinism: a kindred suggesting  
pleiotropism or linkage

JOURNAL Clin. Genet. 5 (3), 196-204 (1974)

PUBMED 4838888

COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to  
final  
NCBI review. The reference sequence was derived from Z22555.1.  
On Aug 13, 2003 this sequence version replaced gi:21361199.

FEATURES Location/Qualifiers

source 1..2566  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/map="12q24.31"

gene 1..2566  
/gene="SCARB1"  
/note="synonyms: CLA1, SRB1, CLA-1, SR-BI, CD36L1"  
/db\_xref="GeneID:949"  
/db\_xref="LocusID:949"  
/db\_xref="MIM:601040"

CDS 70..1599  
/gene="SCARB1"  
/note="CD36 antigen-like 1; scavenger receptor class B  
type 1; CD36 antigen (collagen type I receptor,



```

thrombospondin receptor)-like 1;
go_component: membrane [goid 0016020] [evidence IEA];
go_component: microfibril [goid 0001527] [evidence
IEA];
go_component: membrane fraction [goid 0005624]
[evidence
NR];
go_component: plasma membrane [goid 0005886] [evidence
E]
[pmid 7689561];
go_component: integral to membrane [goid 0016021]
[evidence IEA];
go_function: transporter activity [goid 0005215]
[evidence
E];
go_function: receptor activity [goid 0004872] [evidence
IEA];
go_function: cell adhesion molecule activity [goid
0005194] [evidence IEA];
go_process: cholesterol metabolism [goid 0008203]
[evidence E];
go_process: small molecule transport [goid 0006832]
[evidence E];
go_process: apoptosis [goid 0006915] [evidence NR];
go_process: cell adhesion [goid 0007155] [evidence
IEA]"
/codon_start=1
/product="scavenger receptor class B, member 1"
/protein_id="NP_005496.3"
/db_xref="GI:33620767"
/db_xref="GeneID:949"
/db_xref="LocusID:949"
/db_xref="MIM:601040"

/translation="MGCSAKARWAAGALGVAGLLCAVLGAVMIVMVPVSLIKQQVLKNV
RIDPSSLSFNMWKEIPIPFYLSVYFFDVMNPSEILKGEKPQVRERGPYVYRESRHKSN
ITFNNDTVSFLEYRTFQFQPSKSHGSESDYIVMPNILVLGAAVMMENKPMTLKLIMT
LAFTTLGERAFMNRVTGGEIMWGYKDPLVNLINKYFPGMFPFKDKFGLFAELNNSDGL
FTVFTGVQNISRIHLVDKWNGLSKVDFWHSQCNMINTSGQMWPFPMTPESSLEFYS
PEACRSMKLMYKESGVFEGIPTYRFVAPKTLFANGSIYPPNEGFCPCLESQNVSTC
RFSAPLFLSHPHFLNADPVLAEAVTGLHPNQEAHSLFLDIHPVTGIPMNCVSKLQLSL
YMKSVAGIGQTKIEPVVLP LLWFAESGAMEGETLHTFYTQLVLMPKVMHYAQYVLLA
LGCVLLLVPVICQIRSQEKCYLFWSSSKKGSKDKEAIQAYSESLM TSAPKGSVLQEA
L"
misc_feature 151..1404
/gene="SCARB1"
/note="CD36; Region: CD36 family"
/db_xref="CDD:23154"
polyA_site 2532..2537
/gene="SCARB1"
ORIGIN

```

```

1  cgtcgccgtc  cccgtctcct  gccaggcgcg  gagccctgcg  agccgcgggt  gggccccagg
61  cgcgcagaca  tgggctgctc  cgccaaagcg  cgctgggctg  ccggggcgct  gggcgctcg
121  gggctactgt  gcgctgtgct  gggcgctgct  atgatcgtga  tggtgccgct  gctcatcaag
181  cagcaggctc  ttaagaacgt  gcgcatcgac  cccagtagcc  tgtccttcaa  catgtggaag
241  gagatcccta  tccccttcta  tctctccgct  tacttctttg  acgtcatgaa  ccccagcgag
301  atcctgaaag  gcgagaagcc  gcagggtcgg  gagcgcgggc  cctacgtgta  caggagctcc
361  aggcacaaaa  gcaacatcac  cttcaacaac  aacgacaccg  tgctcttctc  cgagtaccgc
421  accttccagt  tccagccctc  caagtcccac  ggctcggaga  gcgactacat  cgtcatgccc
481  aacatcctgg  tcttgggtgc  gccgggtgat  atggagaata  agcccatgac  cctgaagctc
541  atcatgacct  tggcattcac  caccctcggc  gaacgtgcct  tcatgaaccg  cactgtgggt
601  gagatcatgt  ggggctacaa  ggacccctt  gtgaatctca  tcaacaagta  ctttccaggc
661  atgttcccc  tcaaggacaa  gttcggatta  tttgctgagc  tcaacaactc  cgactctggg
721  ctcttcacgg  tgttcacggg  ggtccagaac  atcagcagga  tccacctcgt  ggacaagtgg
781  aacgggctga  gcaaggttga  cttctggcat  tccgatcagt  gcaacatgat  caatggaact
841  tctgggcaaa  tgtggccgcc  ctcatgact  cctgagtctc  cgctggagtt  ctacagccc
901  gaggcctgcc  gatccatgaa  gctaattgac  aaggagtcag  ggggtgttga  aggcattccc
961  acctatcgct  tcgtggctcc  caaaaccctg  tttgccaacg  ggtccatcta  cccaccaaac
1021  aaaggcttct  gccggtcct  ggagttcggg  attcagaacg  tcagcacctg  caggttcagt
1081  gccctctgt  ttctctccca  tcctcacttc  ctcaacgcg  acccggttct  ggcagaagcg
1141  gtgactggcc  tgaccctaa  ccaggaggca  cactcctgt  tcctggacat  ccaccgggtc
1201  acgggaatcc  ccatgaactg  ctctgtgaaa  ctgcagctga  gcctctacat  gaaatctgtc
1261  gcaggcattg  gacaaactgg  gaagattgag  cctgtggtcc  tgccgctgct  ctggtttgca
1321  gagagcgggg  ccatggaggg  ggagactctt  cacacattct  aactcagct  ggtgttgatg
1381  ccaaggtga  tgactatgc  ccagtagctc  ctctggcgc  tgggctgcgt  cctgctgctg
1441  gtccctgtca  tctgcaaat  ccggagccaa  gagaaatgct  atttattttg  gagttagtag
1501  aaaaaggct  caaaggataa  ggaggccatt  caggcctatt  ctgaatccct  gatgacatca
1561  gctcccaagg  gctctgtgct  gcaggaagca  aaactgtagg  gtccctgagga  caccgtgagc
1621  cagccaggcc  tggccgctgg  gcctgaccgg  cccccagcc  cctacacccc  gcttctccc
1681  gactctccca  gcagacagcc  cccagcccc  acagcctgag  cctcccagct  gccatgtgcc
1741  tgttgcaaac  ctgcacacac  gccctggcac  acatacacac  atgcgtgcag  gcttgtgcag
1801  aactcaggg  atggagctgc  tgctgaaggg  acttgtaggg  agaggctcgt  caacaagcac
1861  tgttctgaa  ccttctctcc  acgtggccca  caggctgacc  acaggggctg  tgggtcctgc
1921  gtccccttcc  tcgggtgagc  ctggcctgtc  ccgttcagcc  gttgggccag  gcttctccc
1981  ctccaagtg  aaacactgca  gtcccgtgt  ggtggctccc  catgcaggac  gggccaggct
2041  gggagtggcc  ccttctgtg  ccaaattcag  tggggactca  gtgccaggc  cctggcacga
2101  gctttggcct  tggctacct  gccaggccag  gcaaagcgcc  tttacacagg  cctcgaaaaa
2161  caatggagt  agcacaagat  gccctgtgca  gctgcccgag  ggtctccgcc  caccgccgcc
2221  ggactttgat  cccccgaag  tcttcacagg  cactgcatcg  ggttgtctgg  cgcccttttc
2281  ctccagccta  aactgacatc  atcctatgga  ctgagccggc  cactctctgg  ccgaagtggc
2341  gcaggctgtg  cccccagct  gcccccacc  cctcacagg  tccctcagat  tataggtgcc
2401  caggctgagg  tgaagaggcc  tgggggccct  gccttccggg  cgctcctgga  ccctggggca
2461  aacctgtgac  cttttctac  tggaaatagaa  atgagtttta  tcatctttga  aaaataattc
2521  actcttgaag  taataaacgt  ttaaaaaaat  ggaaaaaaa  aaaaaa
//

```

## 12.4.6 Descripción de PLIN en formato GenBank

```

LOCUS      BC031084                2979 bp    mRNA     linear   PRI 06-OCT-
2003
DEFINITION Homo sapiens perilipin, mRNA (cDNA clone MGC:33812
IMAGE:5284402),
           complete cds.
ACCESSION BC031084

```

VERSION BC031084.1 GI:21410322  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;  
Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2979)  
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M.,  
Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F.,  
Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J.,  
Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
TITLE Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
MEDLINE 22388257  
PUBMED 12477932  
REFERENCE 2 (bases 1 to 2979)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  
Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA  
94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and  
Myers,  
R. M.

found Clone distribution: MGC clone distribution information can be  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 48 Row: i Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi:  
4505884.  
Differences found between this sequence and the human genome  
(build 34) are described in misc\_difference features below.

FEATURES Location/Qualifiers  
source 1..2979  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:33812 IMAGE:5284402"  
/tissue\_type="Brain, hippocampus"  
/clone\_lib="NIH\_MGC\_95"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
gene 1..2979  
/gene="PLIN"  
/db\_xref="LocusID:5346"  
/db\_xref="MIM:170290"  
misc\_difference 3  
/gene="PLIN"  
/note="'T' in cDNA is 'C' in genome."  
CDS 125..1693  
/gene="PLIN"  
/codon\_start=1  
/product="perilipin"  
/protein\_id="AAH31084.1"  
/db\_xref="GI:21410323"  
/db\_xref="LocusID:5346"  
/db\_xref="MIM:170290"  
  
/translation="MAVNKGLTLLDGDLPQENVLQRVLQLPVVSGTCECFQKTYTST  
KEAHPLVASVCNAYEKGVQSASSLAAWSMEPVVRLSTQFTAANELACRGLDHLEEKI  
PALQYPPEKIASSELKDTISTRRLSARNISVPIASTSDKVLGAALAGCELAWGVARDT  
AEFAANTRAGRLASGGADLALGSIEKVVEYLLPADKEESAPAPGHQQAQESPKAKPSL  
LSRVGALTNTLSRYTVQTMARALEQGHTVAMWIPGVVPLSSLAQWGASVAMQAVSRRR  
SEVRVPWLHSLAAQEEDEDQTDTEGEDTEEEEELETEENKFSEVAALPGPRGLLGG  
VAHTLQKTLQTTISAVTWAPAAVLGMAGRVLHLLTPAPAVSSTKGRAMSLSDALKGVTD  
NVVDTVVHVYVPLPRLSLMEPESEFRDIDNPPAEVERREAERRASGAPSAGPEPAPRLA  
QPRRSLRSAQSPGAPPGGLEDEVATPAAPRPGFPAVPREKPKRRVSDSFFRPSVMEP  
ILGRTHYSQLRKKS"  
misc\_feature 143..1318  
/gene="PLIN"  
/note="perilipin; Region: Perilipin family. The  
perilipin family includes lipid droplet-associated protein

```

(perilipin) and adipose differentiation-related protein
(adipophilin)"
/db_xref="CDD:pfam03036"
misc_difference 271
/gene="PLIN"
/note="'A' in cDNA is 'C' in genome; no amino acid
change."
misc_difference 704
/gene="PLIN"
/note="'G' in cDNA is 'C' in genome; amino acid
difference: 'A' in cDNA, 'P' in genome."
misc_difference 752
/gene="PLIN"
/note="'G' in cDNA is 'A' in genome; amino acid
difference: 'E' in cDNA, 'K' in genome."
misc_difference 862
/gene="PLIN"
/note="'A' in cDNA is 'C' in genome; no amino acid
change."
misc_difference 1237
/gene="PLIN"
/note="'C' in cDNA is 'T' in genome; no amino acid
change."
misc_difference 2144
/gene="PLIN"
/note="'A' in cDNA is 'C' in genome."
misc_difference 2695^2696
/gene="PLIN"
/note="2 bases in genome, GC, are not found in cDNA."
misc_difference 2908..2979
/gene="PLIN"
/note="polyA tail: 72 bases do not align to the
genome."
ORIGIN
1 agctctgggct ctgtgagact gaggtggcgg tcagccggag tgagtgttgg ggtcctgggg
61 cacctgcctt acatggcttg tttatgaaca ttaaaggaa gaagttgaag cttgaggagc
121 gaggatggca gtcaacaagg gcctcacctt gctggatgga gacctccctg agcaggagaa
181 tgtgctgcag cgggtcctgc agctgccggt ggtgagtggc acctgcgaat gcttccagaa
241 gacctacacc agcactaagg aagcccacc actggtggcc tctgtgtgca atgcctatga
301 gaagggcgtg cagagcgcca gtagcttggc tgcctggagc atggagccgg tggccgcgag
361 gctgtccacc cagttcacag ctgccaatga gctggcctgc cgaggcttgg accacctgga
421 ggaaaagatc cccgcctcc agtaccctcc tgaaaagatt gcttctgagc tgaaggacac
481 catctccacc cgcctccgca gtgccagaaa cagcatcagc gttcccatcg cgagcacttc
541 agacaaggtc ctgggggccc ctttggccgg gtgagcagct gcctgggggg tggccagaga
601 cactgaggaa tttgctgcca aactcgagc tggccgactg gcttctggag gggccgactt
661 ggccttgggc agcattgaga aggtgtgga gtacctctc cctgcagaca aggaagagtc
721 agccctgct cctggacacc agcaagccca ggagtctccc aaggccaagc caagcctctt
781 gagcagggtt ggggctctga ccaacacct ctctcgatac accgtgcaga ccatggcccg
841 ggccctggag cagggccaca cagtggccat gtggatccca ggctgtgtgc ccctgagcag
901 cctggcccag tgggtgcct cagtggccat gcaggcgggtg tcccggcggg ggagcgaagt
961 gcgggtaccc tggctgcaca gcctcgagc cgcccaggag gaggatcatg aggaccagac
1021 agacacggag ggagaggaca cggaggagga ggaagaattg gagactgagg agaacaagtt
1081 cagtgaggta gcagccctgc caggccctcg aggcctcctg ggtggtgtgg cacataacct
1141 gcagaagacc ctccagacca ccatctcggc tgtgacatgg gcacctgcag ctgtgctggg
1201 catggcaggg aggggtgctg acctcacacc agccccgct gtctctctca ccaaggggag
1261 ggcatgtcc ctatcagatg cctgaaggg cgttactgac aacgtggtgg acacagtggg
1321 gcattacgtg ccgctcccca ggctgtcgt gatggagccc gagagcgaat tccgggacat
1381 cgacaacca ccagccgag tcgagcggc ggaggcggag cgcagagcgt ctggggcgcc
1441 gtccgcggc ccggagccc cccgcgtct cgcacagccc cgccgagcc tgcgagcggc

```

```

1501 gcagagcccc ggcgcgcccc ccggccccgg cctggaggac gaagtcgcca cgccccgacg
1561 gccgcgcccc ggcttccccg ccgtgccccg cgagaagcca aagcgaggg tcagcgacag
1621 cttcttccgg ccacagctca tggagcccat cctggggcgc acgcattaca gccagctgcg
1681 caagaagagc tgagtcgccg caccagccgc cgcgccccgg gccggcgggg ttctctaaca
1741 aataaacaga acccgcactg cccaggcgag cgttgccact ttcaaatggg tcccctgggg
1801 agctcagcct catcctgatg atgctgcaa ggcgcacttt ttatttttat tttattttta
1861 tttttttttt agcatccttt tggggcttca ctctcagagc cagtttttaa gggacaccag
1921 agcgcagacc tgctctgatt ctatggcttg gttgttacta taagagtaat tgcctaactt
1981 gatttttcat ctcttaacc aaacttgtgg ccaaaagata tttgaccgtt tccaaaattc
2041 agattctgcc tctgcggata aatatttgcc acgaatgagt aactcctgtc accactctga
2101 aggtccagac agaaggtttt gacacattct tagcactgaa ctcatctgtg atctaggatg
2161 atctgttccc cctctgatga acatcctctg atgatctagg ctcccagcag gctactttga
2221 aggaacaat cagatgcaa agctcttggg tgtttattha aaatactagt gtcactttct
2281 gagtaccgc cgcttcacag gctgagcca ggcctgtgtg cttttagtag ccagctgctt
2341 gctcacagcc acatttccat ttgcatcatt actgccttca cctgcatagt cactcttttg
2401 atgctgggga accaaaatgg tgatgatata tagactttat gtatagccac agttcatccc
2461 caaccctagt ctctgaaatg ttaatatattg ataaatctag aaaatgcatt catacaatta
2521 cagaattcaa atattgcaaa aggatgtgtg tctttctccc cgagctcccc tgttccccct
2581 cattgaaaa caccacggtg ccactctctg tgtatgcagg gctatgcacc tgcaggcacg
2641 tgtgatgca ctccccgctt gtgtttacac aagctgtggg gtgttacgca tgcctttttt
2701 tcaacttaata atacagcttg gagagathtt tgtatcacat tataaatccc actcgctctt
2761 tttgatggcc acataataac tactgcataa tatggatacg ccttatttga ttttaactagt
2821 tccctaatga tggactttta agttgtttcc tttttttttc ttttttgcta ctgcaaacga
2881 tgctataata aatgtcctta tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2941 aaaaaaaaaa aaaaaaaaaa aaaggaaaaa aaaaaaaaaa
//

```

## 12.3 Tabla de conversión de unidades

Las medidas de las concentraciones de lípidos y energía se suelen dar en el sistema tradicional que no es el métrico. En esta tesis se ha hecho un esfuerzo para convertir todas las unidades al Sistema Métrico (SI). Sin embargo, para quién no domina las constantes de conversión, porque no trabaja habitualmente con lípidos, es útil poder acceder a ellas por si quiere obtener alguna de las mediciones en cualquiera de los dos sistemas. En este anexo se incluye una tabla de conversión de unidades entre el sistema no métrico y el sistema métrico. También se incluye otra tabla de los prefijos para los múltiplos y divisores de las unidades en el SI.

**Tabla 12.12. Tabla de conversión de unidades entre el sistema no métrico y el sistema métrico**

| Sistema no métrico             | Sistema no métrico | Sistema métrico        |
|--------------------------------|--------------------|------------------------|
| Concentración de colesterol    | 38,67 mg/dL        | 1,00 mmol/L            |
| Concentración de triglicéridos | 38,67 mg/dL        | 1,00 mmol/L            |
| Concentración de fosfolípidos  | 38,67 mg/dL        | 1,00 mmol/L            |
| Radioactividad                 | 1,00 uCi           | 37,00 kBecquerel (Bq)§ |
| Energía                        | 1,00 cal           | 4,184 J                |

§: 1 Becquerel = 1 desintegración por segundo.

**Tabla 12.13. Prefijos para los múltiplos y divisores de las unidades en el SI**

| Prefijo | Símbolo | Factor    | Prefijo | Símbolo | Factor     |
|---------|---------|-----------|---------|---------|------------|
| Exa     | E       | $10^{18}$ | mili    | m       | $10^{-3}$  |
| Penta   | P       | $10^{15}$ | micro   | $\mu$   | $10^{-6}$  |
| Tera    | T       | $10^{12}$ | nano    | n       | $10^{-9}$  |
| Giga    | G       | $10^9$    | pico    | p       | $10^{-12}$ |
| Mega    | M       | $10^6$    | femto   | f       | $10^{-15}$ |
| Kilo    | k       | $10^3$    | atto    | A       | $10^{-18}$ |

## 12.4 Proceso de identificación de genes y polimorfismos asociados a las ECV. Tabla completa

En este anexo se incluye la Tabla 4.3 completada con todos los genes y polimorfismos encontrados de en el proceso de identificación y caracterización de la sección “4.2.3 Búsqueda y detección de polimorfismos (PREDICARD)”.

**Tabla 12.14. Genes y Polimorfismos Identificados. Proceso completado**

| Gen   | Cromosoma | SNP            | Diana                         | Evidencia  | Referencias   |
|-------|-----------|----------------|-------------------------------|--|---|
| APOA1 | 11q21     | -75G/A         | Respuesta a terapia dietética | Determina la respuesta de los niveles de C-HDL a cambios en el consumo de ácidos grasos poliinsaturados.                 | Ordovas JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PW, Schaefer EJ, Tucker K. Polyunsaturated fatty acids modulate the effects of the APOA1 G-A polymorphism on HDL cholesterol concentrations in a sex-specific manner: the Framingham Study. <i>Am J Clin Nutr.</i> 2002; 75:38-46.   |
| APOA4 | 11q21     | G360H Ser347T  | Respuesta a terapia dietética | Determina la respuesta de C-LDL a cambios en el consumo de colesterol. También afecta el estado de lipemia postprandial. | Mata P, Ordovas JM, Lopez-Miranda J, Lichtenstein AH, Clevidence B, Judd JT, Schaefer EJ. ApoA-IV phenotype affects diet-induced plasma LDL cholesterol lowering. <i>Arterioscler Thromb.</i> 1994; 14:884-91.  |
| APOC3 | 11q21     | SstI           | Riesgo                        | Niveles de triglicéridos, potencialmente asociado con un aumento del riesgo cardiovascular                               | Olivieri O, Stranieri C, Bassi A, Zaia B, Girelli D, Pizzolo F, Trabetti E, Cheng S, Grow MA, Pignatti PF, Corrocher R. ApoC-III gene polymorphisms and risk of coronary artery disease. <i>J Lipid Res.</i> 2002;43:1450-7.  |
| APOE  | 19        | E2/E3/E4       | Riesgo y respuesta            | Riesgo Cardiovascular. Respuesta a terapias (Dieta, estatinas).  | Ordovas JM, Mooser V. The APOE locus and the pharmacogenetics of lipid response. <i>Curr Opin Lipidol.</i> 2002; 13:113-7. Review.  |
| APOA5 | 11q21     | 56C/G -1131T/C | Riesgo                        | Hipertrigliceridemia   | Martin S, Nicaud V, Humphries SE, Talmud PJ, EARS group. Contribution of APOA5 gene variants to plasma triglyceride determination and to the response to both fat and glucose tolerance challenges. <i>Biochim Biophys Acta.</i> 2003;1637:217-25.<br>Talmud PJ, Hawe E, Martin S, Olivier M, Miller GJ, Rubin EM, Pennacchio LA, Humphries SE. Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. <i>Hum Mol Genet.</i> 2002; 11:3039-46.   |
| LIPC  | 15q21-q23 | C-514T         | Riesgo y respuesta            | Niveles de C-HDL, Riesgo cardiovascular y respuesta a terapia  | Zambon A, Deeb SS, Pauletto P, Crepaldi G, Brunzell JD. Hepatic lipase: a marker for cardiovascular disease risk and response to therapy. <i>Curr Opin Lipidol.</i> 2003;14:179-89.<br>Deeb SS, Zambon A, Carr MC, Ayyobi AF, Brunzell JD. Hepatic lipase and dyslipidemia: interactions among genetic variants, obesity, gender, and diet. <i>J Lipid Res.</i> 2003; 44:1279-86.<br>Ordovas JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PW, Schaefer EJ, Tucker KL. Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene-nutrient interaction in the Framingham Study. <i>Circulation.</i> 2002; 106:2315-21.<br>Andersen RV, Wittup HH, Tybjaerg-Hansen A, Steffensen R, Schnohr P, Nordestgaard BG. Hepatic lipase mutations, elevated high-density lipoprotein cholesterol, and increased risk of ischemic heart disease: the Copenhagen City Heart Study. <i>J Am Coll Cardiol.</i> 2003; 41:1972-82. |



| Gen  | Cromosoma | SNP                    | Diana                 | Evidencia   | Referencias  |
|------|-----------|------------------------|-----------------------|---|--|
| LPL  | 8p22      | S447X<br>D9N,<br>N291S | Riesgo y<br>respuesta | Hipertrigliceridemia,<br>c-HDL, Riesgo cardiovascular,<br>Respuesta a terapia | <p>Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Folch J, Ordovas JM. Associations of LPL and APOC3 gene polymorphisms on plasma lipids in a Mediterranean population: interaction with tobacco smoking and the APOE locus. <i>J Lipid Res.</i> 2002; 43:416-27.</p> <p>Kastelein JJ, Ordovas JM, Wittekoek ME, Pimstone SN, Wilson WF, Gagne SE, Larson MG, Schaefer EJ, Boer JM, Gerdes C, Hayden MR. Two common mutations (D9N, N291S) in lipoprotein lipase: a cumulative analysis of their influence on plasma lipids and lipoproteins in men and women. <i>Clin Genet.</i> 1999; 56:297-305.</p> <p>Gagne SE, Larson MG, Pimstone SN, Schaefer EJ, Kastelein JJ, Wilson PW, Ordovas JM, Hayden MR. A common truncation variant of lipoprotein lipase (Ser447X) confers protection against coronary heart disease: the Framingham Offspring Study. <i>Clin Genet.</i> 1999; 55:450-4.</p> <p>Ordovas JM, Cupples LA, Corella D, Otvos JD, Osgood D, Martinez A, Lahoz C, Coltell O, Wilson PW, Schaefer EJ. Association of cholesteryl ester transfer protein-TaqIB polymorphism with variations in lipoprotein subclasses and coronary heart disease risk: the Framingham study. <i>Arterioscler Thromb Vasc Biol.</i> 2000; 20:1323-9.</p> <p>Agerholm-Larsen B, Tybjaerg-Hansen A, Schnohr P, Steffensen R, Nordestgaard BG. Common cholesteryl ester transfer protein mutations, decreased HDL cholesterol, and possible decreased risk of ischemic heart disease: The Copenhagen City Heart Study. <i>Circulation.</i> 2000; 102:2197-203.</p> <p>Blankenberg S, Rupprecht HJ, Bickel C, Jiang XC, Poirier O, Lackner KJ, Meyer J, Cambien F, Tiret L; AtheroGene Investigators. Common genetic variation of the cholesteryl ester transfer protein gene strongly predicts future cardiovascular death in patients with coronary artery disease. <i>J Am Coll Cardiol.</i> 2003; 41:1983-9.</p> |
| CETP | 16q21     | Taq1B                  | Riesgo                | c-HDL, Riesgo cardiovascular  |  |

| Gen    | Cromosoma   | SNP                   | Diana              | Evidencia  | Referencias  |
|--------|-------------|-----------------------|--------------------|--|--|
| SCARB1 | 12q         | Exon1, Intron5, exon8 | Riesgo             | c-LDL, c-HDL   | McCarthy JJ, Lehner T, Reeves C, Moliterno DJ, Newby LK, Rogers WJ, Topol EJ. GeneQuest investigators. Association of genetic variants in the HDL receptor, SR-B1, with abnormal lipids in women with coronary artery disease. <i>J Med Genet.</i> 2003; 40:453-8.<br><br>Osgood D, Corella D, Demissie S, Cupples LA, Wilson PW, Meigs JB, Schaefer EJ, Coltell O, Ordovas JM. Genetic variation at the scavenger receptor class B type I gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: the Framingham study. <i>J Clin Endocrinol Metab.</i> 2003; 88:2869-79.<br><br>Acton S, Osgood D, Donoghue M, Corella D, Pocioli M, Cenarro A, Mozas P, Keilly J, Squazzo S, Woolf EA, Ordovas JM. Association of polymorphisms at the SR-B1 gene locus with plasma lipid levels and body mass index in a white population. <i>Arterioscler Thromb Vasc Biol.</i> 1999; 19:1734-43.<br><br>Tai ES, Demissie S, Cupples LA, Corella D, Wilson PW, Schaefer EJ, Ordovas JM. Association between the PPARA L162V polymorphism and plasma lipid levels: the Framingham Offspring Study. <i>Arterioscler Thromb Vasc Biol.</i> 2002; 22:805-10.<br><br>Flavell DM, Jamshidi Y, Hawe E, Pineda Torra I, Taskinen MR, Frick MH, Nieminen MS, Kesaniemi YA, Pasternack A, Staels B, Miller G, Humphries SE, Talmud PJ, Svanne M. Peroxisome proliferator-activated receptor alpha gene variants influence progression of coronary atherosclerosis and risk of coronary artery disease. <i>Circulation.</i> 2002; 105:1440-5.<br><br>Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. <i>Nat Genet.</i> 1998; 20:284-7.<br><br>Clee SM, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, Jukema JW, van Wijland M, van Dam M, Hudson TJ, Brooks-Wilson A, Genest J Jr, Kastelein JJ, Hayden MR. Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. <i>Circulation.</i> 2001; 103:1198-205. |
| PPARA  | 22q12-q13.1 | L162V                 | Riesgo y respuesta | c-LDL, c-HDL, respuesta a niveles de ácidos grasos polinsaturados en la dieta. |  |
| PPARG  | 3p25        | Pro12Ala              | Riesgo y respuesta | Riesgo, c-HDL  |  |
| ABCA1  | 9q22-q31    | R219K                 | Riesgo             | Riesgo, c-HDL  |  |

| Gen  | Cromosoma     | SNP               | Diana                | Evidencia  | Referencias   |
|------|---------------|-------------------|----------------------|--|---|
| ESR1 | 6q25.1        | IVS1-401          | Respuesta en Mujeres | HDL, y respuesta a terapia hormonal sustitutiva.   | Herrington DM, Howard TD, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Brosnihan KB, Meyers DA, Blecker ER. Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease.[comment]. [Clinical Trial. Journal Article] <i>New England Journal of Medicine</i> . 2002; 346:967-74. 2002  |
| ADH3 | 4q22          | Gamma1/<br>gamma2 | Riesgo y respuesta   | c-HDL, riesgo cardiovascular y respuesta al consumo de alcohol.  | Hines L. M., Stampfer M. J., Ma J., Gaziano J. M., Ridker P. M., Hankinson S. E., Sacks F., Rimm E. B., Hunter D. J. N. Genetic Variation in Alcohol Dehydrogenase and the Beneficial Effect of Moderate Alcohol Consumption on Myocardial Infarction <i>Engl J Med</i> 2001; 344:549-555.  |
| GCLM | 1p22.1        | -588C/T           | Riesgo y respuesta   | Riesgo cardiovascular y respuesta al beneficio de antioxidantes.   | Nakamura S, Kugiyama K, Sugiyama S, Miyamoto S, Koide S, Fukushima H, Honda O, Yoshimura M, Ogawa H. Polymorphism in the 5'-flanking region of human glutamate-cysteine ligase modifier subunit gene is associated with myocardial infarction. <i>Circulation</i> . 2002; 105:2968-73.  |
| MMP9 | 20q11.2-q13.1 | C-1562T<br>R279Q  | Riesgo               | Riesgo cardiovascular (niveles elevados de metaloproteinasa 9).  | Morgan AR, Zhang B, Tapper W, Collins A, Ye S. Haplotypic analysis of the MMP-9 gene in relation to coronary artery disease. <i>J Mol Med</i> . 2003; 81:321-6.   |
| IL6  | 7p21          | -174(G/C)         | Riesgo y respuesta   | Concentraciones elevadas de interleukina 6 y aumento del riesgo cardiovascular. Aumentado en presencia de alto consumo de alcohol. | Blankenberg S, Rupprecht HJ, Poirier O, Bickel C, Smieja M, Hafner G, Meyer J, Cambien F, Tiret L; AtheroGene Investigators. Plasma concentrations and genetic variation of matrix metalloproteinase 9 and prognosis of patients with cardiovascular disease. <i>Circulation</i> . 2003; 107:1579-85.<br>Jerrard-Dunne P, Sitzer M, Risley P, Steckel DA, Buehler A, von Kegler S, Markus HS; Carotid Atherosclerosis Progression Study. Interleukin-6 promoter polymorphism modulates the effects of heavy alcohol consumption on early carotid artery atherosclerosis: the Carotid Atherosclerosis Progression Study (CAPS). <i>Stroke</i> . 2003; 34:402-7.<br>Basso F, Lowe GD, Rumley A, McMahon AD, Humphries SE. Interleukin-6 -174G>C polymorphism and risk of coronary heart disease in West of Scotland coronary prevention study (WOSCOPS). <i>Arterioscler Thromb Vasc Biol</i> . 2002; 22:599-604.<br>Jenny NS, Tracy RP, Ogg MS, Luongle A, Kuller LH, Arnold AM, Sharrett AR, Humphries SE. In the elderly, interleukin-6 plasma levels and the -174G>C polymorphism are associated with the development of cardiovascular disease. <i>Arterioscler Thromb Vasc Biol</i> . 2002; 22:2066-71. |

| Gen  | Cromosoma | SNP                         | Diana              | Evidencia   | Referencias   |
|------|-----------|-----------------------------|--------------------|---|---|
| CD14 | 5q31.1    | -159C/T                     | Riesgo y respuesta | Niveles elevados de marcadores de inflamación y mayor riesgo de aterosclerosis que se incrementa especialmente en fumadores | Risley P, Jerrard-Dunne P, Sitzer M, Buehler A, von Kegler S, Markus HS; Carotid Atherosclerosis Progression Study. Promoter polymorphism in the endotoxin receptor (CD14) is associated with increased carotid atherosclerosis only in smokers: the Carotid Atherosclerosis Progression Study (CAPS). Stroke. 2003; 34:600-4.  |
| TLR4 | 9q32-q33  | Asp299Gly<br>y<br>Thr399Ile | Riesgo y respuesta | Mayor riesgo cardiovascular que se revierte tras tratamiento con estatinas.   | Boekholdt SM, Agema WR, Peters RJ, Zwinderman AH, van der Wall EE, Reitsma PH, Kastelein JJ, Jukema JW; REGression GRowth Evaluation Statin Study Group. Variants of toll-like receptor 4 modify the efficacy of statin therapy and the risk of cardiovascular events. Circulation. 2003; 107:2416-21.  |
| NOS3 | 7q36      | Glu298Asp<br>p<br>-786T>C   | Riesgo             | Riesgo de enfermedad cardiovascular   | Rossi GP, Cesari M, Zanchetta M, Colonna S, Maiolino G, Pedon L, Cavallin M, Maiolino P, Pessina AC. The T-786C endothelial nitric oxide synthase genotype is a novel risk factor for coronary artery disease in Caucasian patients of the GENICA study. J Am Coll Cardiol. 2003; 41:930-7.<br><br>Gomma AH, Elrayess MA, Knight C.J, Hawe E, Fox KM, Humphries SE. The endothelial nitric oxide synthase (Glu298Asp and -786T>C) gene polymorphisms are associated with coronary in-stent restenosis. Eur Heart J. 2002; 23:1955-62. |



## 12.5 Algoritmo principal de la herramienta Squeezer2B

```

/**
 * @authors Oscar Coltell and Miguel Arregui
 * Created on April 27th 2004
 */

class SqueezerEngine2B {
/**
 * SQUEEZE
 *
 * "str" representa un array de bytes, donde cada byte es posiblemente
 * un nucleotido de una cadena de ADN o ARN. "length" representa el numero
 * de caracteres de "str".
 *
 * Son caracteres validos:
 *
 *      'a', 'c', 'g', 't', 'u',
 *      'A', 'C', 'G', 'T', 'U',
 *
 * es posible que "str" contenga caracteres no validos y en consecuencia
 * {longitudValida("str") < "length"}.
 * Este metodo estatico crea un BitArray con tantos bits como
 * longitudValida("str") * 2, ya que emplea 2 bits para represetar cada
 * uno de los caracteres validos. Los caracteres no validos son eliminados
 * del BitArray. La siguiente tabla muestra los codigos para los caracteres
 * validos:
 *
 *      b1  b2
 *      A   0  0
 *      C   1  0
 *      G   0  1
 *      T/U  1  1
 *
 * El metodo detecta la presencia de "Ts" en cadenas de ARN y la presencia
de
 * "Us" en cadenas de ADN, lanzando una excepcion en cualquiera de estos
casos.
 */
    public static BitArray squeeze (byte [] str, int init, int end) throws
Exception {
        int nucleotideCount=0;
        for (int i=init; i<end; i++){
            if (isValid(str[i])) nucleotideCount++;
        }
        BitArray squeezeSequence = new BitArray(nucleotideCount * 2);
        int validCharacterCount=0, str_idx=init, bitArray_idx=0, timineCount=0;
        while (validCharacterCount < nucleotideCount){
            switch (str[str_idx++){
                case 'A': case 'a': break;
                case 'C': case 'c': squeezeSequence.set(bitArray_idx); break;
                case 'G': case 'g': squeezeSequence.set(bitArray_idx+1); break;
                case 'T': case 't': if (!squeezeSequence.isDNA()){

```

```

        throw new Exception("RNA cannot have
\"Ts\".");
    }
    timineCount++; // para detectar si hay "Ts" en
cadenas de ARN
    squeezeSequence.set(bitArray_idx);
    squeezeSequence.set(bitArray_idx+1); break;
    case 'U': case 'u': squeezeSequence.setIsDNA(false);
    if (timineCount > 0){
        throw new Exception("DNA cannot have
\"Us\".");
    }
    squeezeSequence.set(bitArray_idx);
    squeezeSequence.set(bitArray_idx+1); break;
    default: validCharacterCount --;
}
validCharacterCount ++;
bitArray_idx = validCharacterCount * 2;
}
return squeezeSequence;
}

/**
 * Este metodo convierte un objeto de la clase BitArray en un array
 * de bytes, de acuerdo con los codigos especificados en la tabla de
 * arriba.
 */
public static byte [] stretch (BitArray bset){
    byte [] originalSequence = new byte[bset.size() / 2];
    boolean b1=false, b2=false, isDNA = bset.isDNA();
    for (int i=0,j=0; i<bset.size(); j++, i=j*2){
        b1 = bset.get(i); b2 = bset.get(i+1);
        if (!b1 && !b2) originalSequence[j] = 'A';
        else if ( b1 && !b2) originalSequence[j] = 'C';
        else if (!b1 && b2) originalSequence[j] = 'G';
        else /* (!b1 && !b2)*/{
            if (isDNA){
                originalSequence[j] = 'T';
            }
            else{
                originalSequence[j] = 'U';
            }
        }
    }
    return originalSequence;
}

/**
 * Dado un byte indica si es un caracter valido o no.
 */
public static boolean isValid (byte b){
    return (b=='A' || b=='C' || b=='G' || b=='T' || b=='U' ||
        b=='a' || b=='c' || b=='g' || b=='t' || b=='u');
}
}
// Eof

```

## 12.6 Algoritmo principal de la herramienta Squeezer3B

```

/**
 * @authors Oscar Coltell and Miguel Arregui
 * Created on April 27th 2004
 */

class SqueezerEngine3B {
/**
 * SQUEEZE
 *
 * "str" representa un array de bytes, donde cada byte es posiblemente
 * un nucleotido de una cadena de ADN o ARN. "length" representa el numero
 * de caracteres de "str".
 *
 * Son caracteres validos:
 *
 *      'a', 'c', 'g', 't', 'u',
 *      'A', 'C', 'G', 'T', 'U',
 *      'n', 'N', 'k', 'K',
 *      ' ', '\t', '\r' '\n'
 *
 * es posible que "str" contenga caracteres no validos y en consecuencia
 * {longitudValida("str") < "length"}.
 * Este metodo estatico crea un BitArray con tantos bits como
 * longitudValida("str") * 3, ya que emplea 3 bits para represetar cada
 * uno de los caracteres validos. Los caracteres no validos son eliminados
 * del BitArray. La siguiente tabla muestra los codigos para los caracteres
 * validos:
 *
 *      b1  b2  b3
 *      A   0   0   0
 *      C   1   0   0
 *      G   0   1   0
 *      ' '\t' 1   1   0
 *      T/U   0   0   1
 *      '\r'  1   0   1
 *      K/N   0   1   1
 *      '\n'  1   1   1
 *
 * El metodo detecta la presencia de "Ts" en cadenas de ARN y la presencia
 * de
 * "Us" en cadenas de ADN, lanzando una excepcion en cualquiera de estos
 * casos.
 */
    public static BitArray squeeze (byte [] str, int init, int end) throws
    Exception {
        int nucleotideCount=0;
        for (int i=init; i<end; i++){
            if (isValid(str[i])) nucleotideCount++;
        }
        BitArray squeezeSequence = new BitArray(nucleotideCount * 3);
        int validCharacterCount=0, str_idx=init, bitArray_idx=0, timineCount=0;
        while (validCharacterCount < nucleotideCount){
            switch (str[str_idx++){
                case 'A': case 'a': break;

```



```

        case 'C': case 'c': squeezeSequence.set(bitArray_idx); break;
        case 'G': case 'g': squeezeSequence.set(bitArray_idx+1); break;
        case 'T': case 't': if (!squeezeSequence.isDNA()){
                                throw new Exception("RNA cannot have
\"Ts\".");
                                }
                                timineCount++; // para detectar si hay "Ts" en
cadenas de ARN
                                squeezeSequence.set(bitArray_idx+2); break;
        case 'U': case 'u': squeezeSequence.setIsDNA(false);
                                if (timineCount > 0){
                                    throw new Exception("DNA cannot have
\"Us\".");
                                }
                                squeezeSequence.set(bitArray_idx+2); break;
        case 'K': case 'k':
        case 'N': case 'n': squeezeSequence.set(bitArray_idx+1);
                                squeezeSequence.set(bitArray_idx+2); break;
        case '\r':
                                squeezeSequence.set(bitArray_idx);
                                squeezeSequence.set(bitArray_idx+2); break;
        case '\n':
                                squeezeSequence.set(bitArray_idx);
                                squeezeSequence.set(bitArray_idx+1);
                                squeezeSequence.set(bitArray_idx+2); break;
        case ' ': case '\t': squeezeSequence.set(bitArray_idx);
                                squeezeSequence.set(bitArray_idx+1); break;
        default: validCharacterCount --;
    }
    validCharacterCount ++;
    bitArray_idx = validCharacterCount * 3;
}
return squeezeSequence;
}

/**
 * Este metodo convierte un objeto de la clase BitArray en un array
 * de bytes, de acuerdo con los codigos especificados en la tabla de
 * arriba.
 */
public static byte [] stretch (BitArray bset){
    byte [] originalSequence = new byte[bset.size() / 3];
    boolean b1=false, b2=false, b3= false, isDNA = bset.isDNA();
    for (int i=0,j=0; i<bset.size(); j++, i=j*3){
        b1 = bset.get(i); b2 = bset.get(i+1); b3 = bset.get(i+2);
        if (!b1 && !b2 && !b3) originalSequence[j] = 'A';
        else if ( b1 && !b2 && !b3) originalSequence[j] = 'C';
        else if (!b1 && b2 && !b3) originalSequence[j] = 'G';
        else if (!b1 && !b2 && b3){
            if (isDNA){
                originalSequence[j] = 'T';
            }
            else{
                originalSequence[j] = 'U';
            }
        }
        else if (!b1 && b2 && b3) originalSequence[j] = 'N';
        else if (b1 && !b2 && b3) originalSequence[j] = '\r';
        else if (b1 && b2 && b3) originalSequence[j] = '\n';
        else /*(!b1 && b2 && !b3)*/ originalSequence[j] = ' ';
    }
}

```

```
    }
    return originalSequence;
}

/**
 * Dado un byte indica si es un caracter valido o no.
 */
public static boolean isValid (byte b){
    return (b=='A' || b=='C' || b=='G' || b=='T' || b=='U' ||
           b=='a' || b=='c' || b=='g' || b=='t' || b=='u' ||
           b=='K' || b=='k' || b=='N' || b=='n' ||
           b==' ' || b=='\t' || b=='\r' || b=='\n');
}

public static void main (String [] args){
    byte [] orig =
{'A','u','u','C','G','U','k','u','n','u','j','p','u','u','u','c'};
    BitArray barray = null;
    byte [] res = null;
    try {
        barray = squeeze(orig, 0, orig.length);
        barray.show();
        res = stretch(barray);
    }
    catch (Exception e){
        System.err.println("Error: " + e.getMessage());
        System.exit(-1);
    }
    System.out.print("Orig: ");
    for (int i=0; i<orig.length; i++){
        System.out.print(" " + (char)orig[i]);
    }
    System.out.println("");

    System.out.print("Res: ");
    for (int i=0; i<res.length; i++){
        System.out.print(" " + (char)res[i]);
    }
    System.out.println("");
}
}
// Eof
```

## 12.7 Estructura de datos principal usada por las herramientas Squeezer2B y Squeezer3B

```

/**
 * @authors Oscar Coltell and Miguel Arregui
 * Created on April 27th 2004
 */

/**
 * Esta clase representa un conjunto de bits. Es serializable
 * porque de esta manera se puede guardar en un fichero de datos
 * o enviar por la red. Los bits son almacenados en un array de
 * octetos, tal que un objeto de esta clase contara con tantos
 * octetos como (numBits/8), mas uno si (numBits%8) > 0.
 * Inicialmente todos los octetos toman 0 como valor, lo que
 * equivale a inicializar cada bit a 0.
 *
 * El metodo "set" permite poner a 1 un bit cualquiera dentro del
 * objeto. El metodo "get" permite ver que valor tiene un bit dado
 * dentro del objeto. El metodo "size" devuelve la cantidad total
 * de bits del objeto. El metodo "show" muestra por la salida std
 * una representacion del objeto, donde los bits salen de izqda a
 * derecha de menor peso a mayor.
 *
 * El booleano "isDNA" permite diferenciar si el contenido del
 * objeto es ADN o ARN.
 */
class BitArray implements java.io.Serializable {
    private byte [] bits;
    private int bitCount;
    private boolean isDNA;

    public BitArray (int bitCount){
        this.isDNA = true;
        this.bitCount = bitCount;
        int byteCount = (bitCount/8) + ((bitCount%8)>0)?1:0;
        this.bits = new byte[byteCount];
        for (int i=0; i<byteCount; i++) bits[i]=0;
    }

    public boolean isDNA () {return isDNA;}

    public void setIsDNA (boolean val){isDNA=val;}

    public int size () {return bitCount;}

    public void set (int i){bits[i/8] |= (0x01<<(i%8));}

    public boolean get (int i){return ((bits[i/8] & (0x01<<(i%8))) >
(byte)0);}

    public void show () {

```

---

```
        for (int i=0; i<bitCount; i++){
            if (get(i)) System.out.print(" 1"); else System.out.print(" 0");
        }
        System.out.println("");
    }
}
// Eof
```



