

DEPARTAMENT DE GENÈTICA

GEN *FOXP2*: ESQUIZOFRENIA, ALUCINACIONES
AUDITIVAS Y LENGUAJE.

M^a AMPARO TOLOSA MONTERO

UNIVERSITAT DE VALÈNCIA
Servei de Publicacions
2009

Aquesta Tesi Doctoral va ser presentada a València el dia 7 de juliol de 2009 davant un tribunal format per:

- Dra. Roser González Duarte
- Dr. Jordi Obiols Llandrich
- Dr. Javier Costas Costas
- Dr. Antonio Benítez Burraco
- Dra. M^a Dolores Moltó Ruiz

Va ser dirigida per:

Dra. Rosa de Frutos Illán

Dr. Julio Sanjuán Arias

©Copyright: Servei de Publicacions
M^a Amparo Tolosa Montero

Dipòsit legal: V-4185-2010

I.S.B.N.: 978-84-370-7708-6

Edita: Universitat de València

Servei de Publicacions

C/ Arts Gràfiques, 13 baix

46010 València

Spain

Telèfon:(0034)963864115



VNIVERSITATIS VALÈNCIA

Facultad de Ciencias Biológicas
Departamento de Genética

Gen *FOXP2* : esquizofrenia,
alucinaciones auditivas y lenguaje

Tesis Doctoral

M^a Amparo Tolosa Montero

2009

La Dra Rosa de Frutos Illán, Catedrática de Genética del Departamento de Genética, Facultad de Biología, Universidad de Valencia y el Dr. Julio Sanjuán Arias, Profesor Titular del Departamento de Medicina y Psiquiatría, Facultad de Medicina, Universidad de Valencia

CERTIFICAN:

Que la memoria titulada "Gen *FOXP2*: esquizofrenia, alucinaciones auditivas y lenguaje" ha sido realizada bajo su codirección en el Departamento de Genética, Facultad de Biología por la Licenciada M^a Amparo Tolosa Montero.

Y para que así conste, en cumplimiento de la legislación vigente, firman el presente certificado en Valencia, a 27 de Abril del 2009.

Dra. Rosa de Frutos Illán

Dr. Julio Sanjuán Arias

Para todos los que vieron el largo camino y encontraron fuerzas para acabarlo, y especialmente para mi padre, mi madre y mi hermana.

El desarrollo de esta tesis ha constituido una importante parte de mi vida. Han sido años de alegrías y tristezas científicas, trabajo, y sobre todo, de aprendizaje constante. Durante este tiempo, muchas personas se han cruzado en mi camino o han recorrido parte conmigo y a muchas de ellas tengo que agradecer el poder llevar a cabo esta tesis.

En primer lugar, quiero hacer una mención especial a todas las personas que voluntaria y desinteresadamente han participado en los diferentes estudios. Todo el trabajo realizado no hubiera sido posible sin ellos.

Me gustaría agradecer a mis directores de tesis, Rosa de Frutos y Julio Sanjuán la oportunidad y confianza depositadas en mí, así como el hecho de iniciarme en el terreno de la investigación. Son algunos ya los años desde que empecé y aunque siempre jefes y directores, hace tiempo que descubrí que además de grandes profesionales son grandes personas.

También quiero agradecer al resto de profesores de la línea, Loli, M^aJosé, Carmen, Lluís, su contribución, bajo la forma de consejos o sugerencias, a este trabajo. Así como a todo el personal del Departament de Medicina de la Universitat de València, por su participación en la parte clínica del estudio.

El trabajo diario en el laboratorio nunca hubiera sido posible sin la inestimable presencia y ayuda, a veces de la forma menos seria posible, de los Moleculares. Ellos han hecho de éste mi segundo hogar. Tres personas marcaron el inicio de mi trayectoria en el laboratorio: Isabel, mi gran maestra, Josep que siempre ha estado ahí para cualquier cosa y no dudo que seguirá estando y Juan Antonio, que me mandó la primera digestión, a la cual se me olvidó añadir DNA...A Olga, Ivette y José Luís que han sido sufridores de línea conmigo tengo que agradecerles la infinita paciencia, los comentarios y los consejos. Olga, empezamos al mismo tiempo y parece que acabaremos también casi a la vez. Muchas gracias por ser el magnífico apoyo que siempre has sido. José Luís, la estadística no sería lo mismo sin ti. A Sirena, gracias por su ayuda en trabajo y por los momentos de esparcimiento. A Laura, por las tertulias. A Javi, por sus monólogos. Y a todos los colaboradores que han intervenido en algún paso de la investigación (Ana, Ángels, Jero, Noe...). A todos los moleculares, presentes y pasados, agradezco la ayuda, la amistad y los karaokes improvisados. Ha sido un placer trabajar con ellos.

No puedo olvidarme del resto de personas con las que día a día he convivido en el departamento: los GMDses, los bioquímicos...A todos ellos muchas gracias por los consejos y por las sobremesas.

Agradezco también la ayuda de todo el personal PAS del departamento, Carmen, José Carlos, Fede, Javi, Nuria y Begoña por su ayuda en mis dudas de protocolos y paciencia en tareas administrativas.

Quiero agradecer también la ayuda prestada por otros laboratorios bien de la Universitat de Valencia, o bien del Hospital la Fe de Valencia por permitirme utilizar determinados aparatos o ayudarme con experimentos.

Sin duda una de las experiencias que más ha enriquecido la etapa de realización de mi tesis fue la estancia que realicé en el *Prince of Wales International Centre* de Oxford, bajo la dirección de Tim Crow. A este ilustre psiquiatra tengo que agradecer el gran ejemplo de profesionalidad, trabajo y amabilidad que mostró en todo momento.

Agradezco también al resto de investigadores y colaboradores que me ayudaron a sentirme como en casa: Norman, Steven, James, Enrica, Manaán, Pillerin, Rosana y especialmente, a Tom, por su continua asistencia, Pablo por sus ánimos y a Adam, por, cómo él dice, enseñarme tantas cosas de la ciencia y la vida.

Un grupo muy especial al que tengo que agradecer gran parte de las fuerzas para seguir adelante es el de las *Superbiologistas*: Almu, Ana, Encarna, Fran, Marián, MJ, Montse, Olga, Rosario. Gracias a las comidas de los martes y jueves, las excursiones, cenas, meriendas, comidas y situaciones de crisis en las que he podido contar con ellos para animarme o animarlos. Es un gusto saber que están y comprenden, sin importar la distancia.

También tengo que agradecer especialmente a Anna y Beatriz, por tener la paciencia suficiente como para escucharme en los malos días y aguantarme los días buenos. Beatriz, siento mucho hablar tanto de la tesis, ya va a pasar... y Anna, gracias por todo.

Y por último, agradecer a los luchadores de todos los días. Los que siempre han estado conmigo, me han ofrecido su asistencia, me han empujado cuando lo necesitaba, y me han dado su apoyo incondicional y absoluto, aunque no supieran muchas veces de lo que estaba hablando. A mi padre, mi madre y mi hermana les agradezco y dedico esta tesis.

Índice



INTRODUCCIÓN	1
I. Esquizofrenia	1
I.1 Bases genéticas de la esquizofrenia	4
I.1.1 Estudios de ligamiento	6
I.1.2 Estudios de asociación caso-control	7
I.1.3 Epigenética y esquizofrenia	10
II. Alucinaciones auditivas como fenotipo alternativo de la esquizofrenia	12
II.1 Endofenotipos y fenotipos alternativos en el estudio de la esquizofrenia	12
II.2 Alucinaciones auditivas	13
II.2.1 Estudios de neuroimagen en alucinaciones auditivas	15
II.2.2 Estudios genéticos sobre las alucinaciones auditivas	15
II.2.3 Modelos de alucinaciones auditivas	16
III. Lenguaje	18
IV. El gen <i>FOXP2</i>	22
IV.1 Descubrimiento	22
IV.2 Estructura del gen <i>FOXP2</i>	24
IV.3 Variabilidad en el gen <i>FOXP2</i>	29
IV.4 Genes diana de <i>FOXP2</i>	30
IV.5 Expresión del gen <i>FOXP2</i>	31
IV.5.1 Expresión en mamíferos	31
IV.5.2 Expresión en no mamíferos	33
IV.5.3 Animales <i>knockout</i>	34
IV.6 Estudios de neuroimagen de la familia KE	34
IV.7 Función de la proteína FOXP2	35
V. Selección positiva en el genoma humano	36
V.1 Detección de la selección positiva	36
V.2 Selección positiva en regiones codificantes	37
V.3 Evolución en regiones no codificantes	38
V.4 Estudios comparativos de expresión	38
V.5 Selección positiva y esquizofrenia	39
V.6 Evolución del gen <i>FOXP2</i> como ejemplo de selección positiva en el linaje humano	40
V.7 RNA no codificante: <i>HAR1A</i>	43
V.7.1 Expresión de <i>HAR1A</i> y <i>HAR1R</i>	44
V.7.2 Función del gen <i>HAR1A</i>	44

HIPÓTESIS Y OBJETIVOS	45
MATERIAL Y MÉTODOS	47
I. Material biológico	47
I.1 Material biológico humano	47
I.1.1 Muestras utilizadas en el estudio de asociación	47
I.1.1.1 Pacientes	47
I.1.1.1.1 Diagnóstico	47
I.1.1.1.2 Características demográficas	48
I.1.1.1.3 Variables clínicas analizadas	48
I.1.1.1.4 Escalas clínicas Manchester y PSYRATs	49
I.1.1.2 Controles	50
I.1.2 Muestra con Trastorno Específico del Lenguaje (TEL)	51
I.1.3 Muestras de tejido <i>postmortem</i> humano	51
I.2 Muestras de Primates	51
I.3 Cepas bacterianas	52
I.4 Células eucarióticas	52
II. Métodos moleculares	53
II.1 Técnicas básicas comunes	53
II.1.1 Extracción de ácidos nucleicos	53
II.1.2 Cuantificación de ácidos nucleicos	53
II.1.3 PCR	54
II.1.3.1 PCR de colonia	56
II.1.4 Electroforesis	57
II.1.4.1 Electroforesis en gel de agarosa	57
II.1.4.2 Electroforesis en gel de acrilamida	57
II.1.5 Purificación de ácidos nucleicos	58
II.1.5.1 Productos de PCR	58
II.1.5.2 Productos de la digestión con enzimas de restricción	58
II.1.5.3 Purificación a partir de geles de agarosa	59
II.1.6 Secuenciación	59
II.1.7 Clonación	60
II.1.7.1 Ligación	61
II.1.7.2 Transformación	62
II.2 Genotipación	64
II.2.1 Selección de SNPs	64
II.2.2 Métodos de genotipación utilizados	64

II.2.3 Genotipación por RFLPs	66
II.2.3.1 Diseño de cebadores	66
II.2.3.2 Establecimiento de las condiciones de amplificación para cada polimorfismo	67
II.2.3.3 Digestión de los productos de PCR	68
II.3 Búsqueda de mutaciones descritas en región codificante	70
II.4 Búsqueda de variación en el número de repeticiones de trinucleótidos mediante electroforesis capilar	71
II.5 Análisis de la secuencia de la región promotora del gen <i>FOXP2</i> en niños con TEL	72
II.6 Análisis de la región promotora del gen <i>FOXP2</i> en primates	74
II.7 Análisis funcional de la región promotora	75
II.7.1 Regiones reguladoras analizadas	76
II.7.2 Amplificación de los fragmentos de interés de la región reguladora	76
II.7.3 Clonación en el vector pCR®2.1TOPO®	77
II.7.4 Subclonación en pSEAP2-Basic	78
II.7.5 Vectores de expresión utilizados	78
II.7.6 Transfección en células eucarióticas	79
II.7.7 Medida de la actividad de la Fosfatasa Alcalina Secretada y la β -galactosidasa	80
II.7.7.1 Preparación de extractos para la medida de actividades de la Fosfatasa Alcalina Secretada y la β -galactosidasa	80
II.7.7.2 Medida de la actividad de la Fosfatasa Alcalina Secretada y la β -galactosidasa	80
II.8 Análisis de metilación	80
II.8.1 Análisis de metilación mediante la utilización de enzimas de restricción sensibles a metilación	81
II.8.2 Análisis del estado de metilación mediante la técnica del bisulfito	82
II.8.2.1 Muestras analizadas por medio de la modificación con bisulfito	83
II.8.2.2 Diseño de cebadores	84
II.8.2.3 Tratamiento con bisulfito	85
II.8.2.4 Amplificación por PCR del DNA convertido	87
II.8.2.5 Clonación de los productos de PCR	88
II.8.2.6 Obtención de la secuencia de cada clon	88
II.9 PCR cuantitativa	88
II.9.1 Muestras analizadas	88
II.9.2 RT-PCR	89
II.9.3 Diseño de cebadores para la PCR cuantitativa	89
II.9.4 Condiciones de la amplificación por PCR para comprobar conversión del RNA a cDNA	90
II.9.5 qRT-PCR	90
II.9.6 Análisis y cuantificación de resultados	92

III. Herramientas informáticas y bases de datos	93
III.1 Programas informáticos	93
III.1.1 Programas utilizados en el estudio de asociación	93
III.1.2 Programas de análisis de secuencias	94
III.1.3 Programas de análisis evolutivo	95
III.1.3.1 Paquete informático Genomatix	95
III.1.3.2 Otros programas de análisis evolutivo	96
III.1.4 Otros programas	96
III.2 Bases de datos y direcciones de Internet	97
RESULTADOS	99
I. Variaciones estructurales en los genes <i>FOXP2</i> y <i>HAR1A</i>	99
I.1 Estudio de asociación caso-control con polimorfismos del gen <i>FOXP2</i>	99
I.1.1 Análisis del desequilibrio de ligamiento	102
I.1.2 Distribución de las frecuencias alélicas y genotípicas en pacientes y controles	104
I.1.3 Análisis de haplotipos en pacientes y controles	111
I.1.4 Medida del poder estadístico	116
I.1.5 Análisis de variables clínicas	118
I.1.5.1 Variables clínicas con dos estados	118
I.1.5.2 Análisis de haplotipos de variables clínicas con dos estados	123
I.1.5.3 Variables clínicas con más de dos estados	124
I.2 Estudio de asociación caso-control con polimorfismos del gen <i>HAR1A</i>	127
I.2.1 Análisis del desequilibrio de ligamiento	128
I.2.2 Distribución de las frecuencias alélicas y genotípicas de polimorfismos del gen <i>HAR1A</i> en pacientes y controles	129
I.2.3 Análisis de haplotipos	132
I.2.4 Medida del poder estadístico	134
I.2.5 Análisis de variables clínicas	135
I.2.5.1 Variables clínicas con dos estados	136
I.2.5.2 Análisis de haplotipos de variables clínicas con dos estados	137
I.2.5.3 Variables clínicas con más de dos estados	138
I.3 Cambio en el número de repeticiones de trinucleótidos en el gen <i>FOXP2</i>	138
I.4 Análisis del gen <i>FOXP2</i> en niños con TEL	140
I.4.1 Búsqueda de mutaciones en región codificante	140
I.4.2 Búsqueda de mutaciones en la región promotora	141

II. Análisis evolutivo de la región promotora del gen <i>FOXP2</i> en diferentes especies de primates	143
II.1 Obtención de secuencias	143
II.2 Análisis de los estados de los polimorfismos humanos en las secuencias de primates analizadas	145
II.3 Búsqueda de regiones conservadas	145
II.4 Delimitación de la región promotora	151
II.4.1 Sitios de unión de factores de transcripción en la región promotora	153
II.5 Análisis filogenético	158
III. Delimitación funcional del promotor	164
III.1 Obtención de las construcciones	165
III.2 Transfección	165
III.3 Medidas de la actividad de la Fosfatasa Alcalina Secretada y la β-galactosidasa	166
IV. Variaciones funcionales	170
IV.1 Análisis de metilación en la región promotora	170
IV.1.1 Análisis de metilación mediante la utilización de enzimas de restricción sensibles a metilación	171
IV.1.2 Análisis de metilación mediante la técnica del bisulfito	173
IV.2 Estudio de los niveles de mRNA en diferentes áreas cerebrales en pacientes con esquizofrenia y controles	179
 DISCUSIÓN	 187
 I. Variaciones estructurales de genes sometidos a selección positiva y esquizofrenia	 187
I.1 Estudio de asociación con variantes de los genes <i>FOXP2</i> y <i>HAR1A</i>	187
I.2 Búsqueda de expansiones de trinucleótidos en el gen <i>FOXP2</i>	196
I.3 Búsqueda de mutaciones en la región promotora	199
II. Análisis del promotor del gen <i>FOXP2</i>	200
II.1 Análisis evolutivo de la región promotora	200
II.2 Caracterización funcional del promotor del gen <i>FOXP2</i>	204
III Variaciones funcionales	206
 CONCLUSIONES	 213
 BIBLIOGRAFÍA	 215

Introducción



I. Esquizofrenia

La esquizofrenia es una de las enfermedades mentales más graves y afecta aproximadamente al 1% de la población. Aparece habitualmente en gente joven, (los hombres tienen una edad de inicio más temprana que las mujeres) y se distribuye en toda la población mundial con una asombrosa homogeneidad respecto a los síntomas.

La esquizofrenia fue definida inicialmente por E. Kraepelin como *Dementia Praecox*. Kraepelin puso el acento en el curso hacia el deterioro que sufrían estos pacientes en comparación con las Psicosis Maniaco-Depresivas (ahora llamadas T. Bipolar) que no tenían asociado un deterioro cognitivo (traducción en Kraepelin, 1970). Posteriormente, Eugen Bleuler cambió el nombre de *Dementia Praecox* por el de esquizofrenia, que es la denominación que aparece hoy en día en todas las clasificaciones internacionales (Bleuler, 1911). Para Bleuler habría cuatro síntomas fundamentales: alteraciones asociativas (del pensamiento), alteraciones afectivas, autismo y ambivalencia. Sin embargo, los síntomas descritos por Bleuler no estaban bien operativizados y no ofrecían una buena fiabilidad diagnóstica.

En un acercamiento más práctico, K. Schneider postuló lo que él denominó síntomas de primer orden, los cuales, aunque no eran totalmente específicos de esquizofrenia sí eran muy fáciles de identificar y muy fiables (Schneider, 1959). Estos síntomas son: pensamientos audibles, voces que comentan, voces que discuten, experiencias de pasividad, robo de pensamiento, trasmisión de pensamiento y percepciones delirantes. Éstos son los síntomas que se han adoptado actualmente como diagnóstico en las clasificaciones internacionales (DSM, CIE).

Así, el diagnóstico de esta enfermedad siguiendo los criterios DSM-IV (APA, 2002) de diagnóstico incluidos en el apartado de Esquizofrenia y otros Trastornos Psicóticos o Trastorno Bipolar, está basado en la aparición concomitante de al menos dos de los siguientes síntomas durante una fracción de tiempo significativa a lo largo de un período de seis meses: delirios, alucinaciones, lenguaje desorganizado, comportamiento catatónico o exageradamente desorganizado y síntomas negativos.

Debido a su impacto social-económico, el estudio de la esquizofrenia ha dado lugar a la aparición en la actualidad de más de 4000 publicaciones al año (datos obtenidos de www.pubmed.org). Sin embargo, tanto su etiología como patofisiología todavía se mantienen sin definir y los tratamientos disponibles tienen un efecto modesto. Así, a pesar de tener un gran número de resultados discretos, la mayoría de ellos no han sido replicados y sólo unos pocos se han corroborado total o parcialmente.

En 1988, Wyatt y colaboradores iniciaron, con motivo del aniversario de la creación de la revista *Schizophrenia Research*, la tradición de reunir en una misma publicación los hechos más consistentes sobre la esquizofrenia (Wyatt *et al.*, 1988). Veinte años después, en la revisión de 2008, Tandon y colaboradores, resumieron los 77 descubrimientos más representativos del estudio de esta enfermedad, agrupándolos en términos de su relevancia para la esquizofrenia: etiología, patofisiología, manifestaciones clínicas y tratamientos. Para esto, puntuaron cada hecho en una escala de 0 a 3 para las medidas de reproducibilidad, carácter fundamental para la esquizofrenia y durabilidad en el tiempo.

En la caja I1 se presentan los hechos más representativos tomados de este artículo.

Introducción

Caja I1			
Hechos	Reproducibilidad	Fundamental en esquizofrenia	Durabilidad en el tiempo
<i>Epidemiología (etiología)</i>			
Incidencia anual de 8-40/100.000/año	+++	++	+++
Mayor incidencia en habitantes de núcleos urbanos	++	++	++
Mayor incidencia en inmigrantes	++	++	++
Riesgo a lo largo de la vida de 0.7%	++	++	++
Prevalencia de 2-10/1000, con núcleos de elevada y baja prevalencia	+++	+	+++
Mayor prevalencia en clases socio-económicas bajas	+++	+	+++
Heredable. Factores genéticos contribuyen al 80% de propensión a la enfermedad.	+++	++	+++
Heterogeneidad genética	+++	+	+
Múltiples factores medioambientales de pequeño efecto	+++	++	++
<i>Neurobiología (patofisiología)</i>			
Volumen total cerebral reducido y espacios ventriculares lateral y tercero aumentados.	+++	++	+++
Volumen de materia gris reducido en regiones específicas del cerebro, como estructuras de los lóbulos temporales medio y superior, córtex prefrontal y tálamo	++	++	++
Presencia de anomalías cerebrales en el momento de aparición de la enfermedad	++	++	++
Agonistas de la dopamina exacerbaban los síntomas de la enfermedad, mientras que antagonistas de la dopamina-2 los alivian	+++	+	+++
<i>Rasgos clínicos (expresión e identificación de la enfermedad)</i>			
Límites nosológicos indistintos entre esquizofrenia y otros trastornos psiquiátricos	+++	+	++
Aunque se han descrito los síntomas característicos, ninguno es patognómico y el diagnóstico está basado en el perfil de síntomas y curso	+++	++	+++
Heterogeneidad en la neurobiología, manifestaciones clínicas, curso y respuesta al tratamiento entre pacientes	+++	+	++
La esquizofrenia es un trastorno crónico y con recaídas, que generalmente presenta una remisión incompleta	++	++	+++
La esquizofrenia se caracteriza por una mezcla de síntomas positivos, negativos, cognitivos y de humor	+++	++	+++
La severidad de los síntomas varía entre pacientes y durante el curso de la enfermedad	+++	++	+++
Existe una discapacidad intelectual generalizada	+++	+	++
Inicio de los síntomas psicóticos durante la adolescencia o edad adulta temprana	+++	++	+++
Edad de inicio más temprana en varones	+++	++	+++
Presencia de defectos premórbidos en una proporción sustancial de pacientes	+++	++	+++
<i>Prevención y tratamiento</i>			
Los antagonistas de la dopamina-2 son los únicos agentes efectivos terapéuticos disponibles	+++	+	+++
Los antipsicóticos tienen eficacia limitada en los síntomas negativos y déficits cognitivos	++	++	++
Amplia variación en el perfil de efectos adversos de los antipsicóticos	+++	0	+++

Aunque existe cierta controversia acerca de la uniformidad en la incidencia de la esquizofrenia en la población mundial, sí se acepta una presencia constante (1-2%) en las diferentes poblaciones humanas, independientemente de la cultura o medio ambiente, cuando se tienen en cuenta los criterios más restrictivos de la enfermedad (Jablensky *et al.*, 1992). Esta presencia constante plantea una importante paradoja, puesto que a pesar de los efectos negativos en la eficacia biológica provocados por la enfermedad, ésta se sigue manteniendo de forma constante en la población (Huxley, 1964).

Además, diferentes datos empíricos sugieren que factores ambientales como vivir en áreas urbanas, ser inmigrante o determinados acontecimientos vitales y ambiente familiar pueden inducir síntomas psicóticos breves o ser decisivos en el curso y en las recaídas de los pacientes (Sundquist *et al.*, 2004; McGrath, 2006; Revisión en Sullivan, 2005).

Otro dato de gran interés es el hecho de que aunque la esquizofrenia aparece normalmente por primera vez durante la adolescencia o temprana edad adulta, puede ser precedida a menudo por meses e incluso años de deterioro de la función antes de que el primer síntoma psicótico sea observado (Hafner, 1992).

Aunque la revisión de Tandon y colaboradores aporta una excelente visión global del estado de la investigación en esquizofrenia, y plantea el debate sobre algunas de las cuestiones, la falta de respuestas esclarecedoras y consistentes en alguno de los campos, como en el de la etiología o genética lleva a autores como DeLisi a cuestionar si todos los hechos reflejados en el resumen son realmente consistentes (DeLisi, 2008). Así, esta autora propone los siguientes 6 hechos como incuestionables:

- La esquizofrenia es una enfermedad que causa consecuencias devastadoras en la vida de las personas que la sufren.
- Tiene una distribución mundial.
- Sus síntomas no se reconocen hasta la adolescencia tardía o edad adulta temprana.
- Se trata de una enfermedad cerebral, tanto estructural como funcionalmente.
- Es familiar.
- No es contagiosa, neoplásica o basada en un daño cerebral traumático

Durante los últimos años, el desarrollo de técnicas estructurales o funcionales dentro del campo de la **neuroimagen** ha aportado un nuevo enfoque al estudio de la neurobiología, ampliando el volumen de información disponible sobre diferentes patologías nerviosas, y proporcionando nuevos interrogantes en muchos casos (Heinz *et al.*, 2003; McGuire y Matsumoto, 2004; Fusar-Poli y Broome, 2006).

Diferentes hallazgos de neuroimagen estructural muestran que hay diferencias morfométricas consistentes entre pacientes con esquizofrenia y controles sanos. Aunque no presentes en todos los pacientes, se ha observado dilatación ventricular, disminución de la densidad en el área del córtex temporal izquierdo y reducción del hipocampo y amígdala (ver revisión en Gur *et al.*, 2007). Existen también evidencias de reducción de la circunvolución temporal superior, tálamo y cuerpo caloso (Metaanálisis en Honea *et al.*, 2005). En lo que respecta a la neuroimagen funcional, se ha puesto de manifiesto que en función de las tareas a las que se les somete, los pacientes con esquizofrenia pueden presentar una hipo o hiperactivación cerebral (Fusar-Poli *et al.*, 2007).

Introducción

Las aportaciones de la neuroimagen en estudios longitudinales con individuos de alto riesgo a padecer esquizofrenia o con pacientes con primeros episodios han sido muy útiles, puesto que indican cambios en la estructura cerebral con el tiempo y la existencia de reducciones en la estructura del lóbulo temporal (reducciones en materia gris de las circunvoluciones fusiforme y del parahipocampo izquierdo) alrededor del momento de transición hacia el diagnóstico de psicosis e incluso antes (Pantelis *et al.*, 2003; Job *et al.*, 2005).

I.1 Bases genéticas de la esquizofrenia

Frente a los factores ambientales, los factores genéticos constituyen la causa conocida más relevante de vulnerabilidad a la esquizofrenia. Estudios familiares, de adopción y de gemelos indican que la esquizofrenia es una enfermedad familiar (Revisión en Sullivan, 2003 y Sullivan, 2005).

Los estudios de agregación familiar muestran que aunque la prevalencia de la esquizofrenia es de alrededor del 1% en la población general (Jablensky *et al.*, 1992), este valor aumenta considerablemente entre los miembros de una misma familia. En la figura I1, obtenida de la revisión de Gottesman que toma datos de más de 40 estudios, se observa que a medida que aumenta el grado de relación familiar (implicando un aumento del número de alelos compartidos) el riesgo a padecer la enfermedad es mayor. Así, por ejemplo, mientras que los hermanos de un individuo esquizofrénico presentan un riesgo promedio de 9%, este porcentaje aumenta significativamente en los hijos de padres esquizofrénicos, obteniéndose un 46% (Gottesman, 1991).

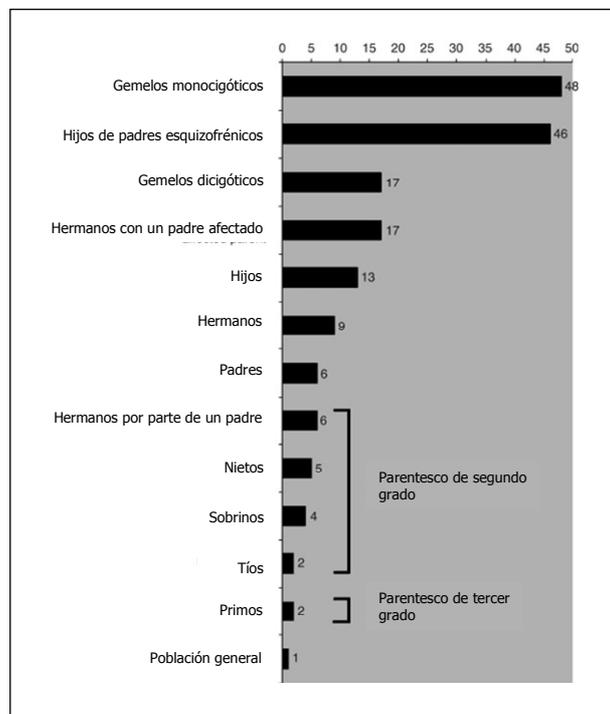


Figura I1. Prevalencia de la esquizofrenia en familias. Figura tomada de Gottesman, 1991 con modificaciones.

Además, la agregación familiar presenta valores mayores en esquizofrenia que en otras enfermedades médicas de origen multifactorial, como la diabetes o la epilepsia (Warner *et al.*, 1995).

Los estudios con gemelos muestran que la concordancia para la esquizofrenia en gemelos monocigóticos varía entre el 15-58% mientras que para los gemelos dicigóticos se sitúa entre el 4-27% (Farmer *et al.*, 1987; Tsuang y Faraone, 1997; Cardno *et al.*, 2002). La heredabilidad obtenida a partir de este tipo de estudios se estima en el 80% (Sullivan *et al.*, 2003).

Los estudios de adopción permiten la evaluación de factores genéticos en esquizofrenia independientemente de la influencia del ambiente familiar (revisión en Ingraham y Kety, 2000). Los primeros estudios de este tipo realizados en esquizofrenia mostraron una mayor prevalencia de la enfermedad en niños con madres esquizofrénicas criados por familias de acogida o instituciones, que en niños control (Heston *et al.*, 1966; Higgins *et al.*, 1997). Posteriormente Tienari, con una amplia muestra procedente de Finlandia, describió tasas de psicosis significativamente más altas entre niños adoptados con padres biológicos esquizofrénicos (Tienari *et al.*, 1985). Del mismo modo, Kety y colaboradores llevaron a cabo un estudio a nivel nacional en Dinamarca, en el que observó que la esquizofrenia crónica se daba con mayor frecuencia entre los familiares biológicos de individuos adoptados con esquizofrenia que entre familiares biológicos de controles (Kety *et al.*, 1975; replicado en Kety *et al.*, 1994).

La esquizofrenia, al igual que la mayoría de enfermedades psiquiátricas, entra dentro del grupo considerado como **enfermedades complejas** (Lander, 1994). Este tipo de enfermedades se puede definir como aquellas en las que la transmisión hereditaria no se ajusta a los modelos de herencia simple mendeliana. Existen diferentes factores que pueden afectar a la alteración en las proporciones mendelianas de la herencia de la esquizofrenia (Schork *et al.*, 2007). El primero de ellos es la penetrancia incompleta, lo que indica que el hecho de que llevar un alelo de riesgo no significa que la enfermedad se manifieste necesariamente en un individuo y son necesarios otros factores, como por ejemplo el ambiente. La existencia de fenocopias, cuando un individuo presenta determinada afección sin llevar el alelo que determina la susceptibilidad a padecerla es otro factor. También lo son la heterogeneidad genética, producida cuando son muchos los genes que intervienen en el mecanismo de acción de una determinada afección, de forma que mutaciones en cualquiera de ellos da lugar al mismo fenotipo y la heterogeneidad clínica. Por último, otro factor a tener en cuenta son los cambios epigenéticos, no codificados en la secuencia del DNA y que también pueden dar lugar a la transmisión no mendeliana de un carácter. De este punto en concreto se hablará en más profundidad en apartados posteriores.

Partiendo del hecho de que debe haber múltiples genes involucrados en el desarrollo de la esquizofrenia, se han propuesto dos modelos principales sobre la estructura alélica de los genes que pueden estar involucrados.

- *Modelo de enfermedades complejas/variantes raras*

Según este modelo, la enfermedad puede estar producida por numerosos alelos, potencialmente de alto riesgo, pertenecientes a numerosos *loci*, cada uno de ellos con una baja frecuencia en la población. Puesto que su frecuencia es baja, se espera que estos alelos no sean compartidos por diferentes subpoblaciones o grupos étnicos y debido a su carácter de alto riesgo se espera que tengan efectos fenotípicos deletéreos y tiendan a ser eliminados por selección purificadora, con lo

Introducción

que estas variantes se mantienen en la población por el equilibrio mutación-selección. (Pritchard, 2001; Wright y Hastie, 2001; McClellan *et al.*, 2007)

- *Modelo de enfermedades comunes/variantes comunes.*

Este modelo, predice que el espectro de alelos que confieren susceptibilidad a una enfermedad está constituido principalmente por alelos comunes, de riesgo moderado y compartidos por múltiples subpoblaciones. Se espera así, que estas variantes comunes no estén sometidas a una fuerte selección purificadora y se mantengan en la población por algún tipo de selección equilibradora (Becker, 2004; Hirschhorn, 2005).

La hipótesis de trabajo asumida en la mayoría de las publicaciones actuales en enfermedades complejas, y especialmente en esquizofrenia sigue el modelo de enfermedades comunes/variantes comunes. Sin embargo, algunos autores empiezan a sugerir que algunas mutaciones que predisponen a padecer este desorden tienen elevada penetrancia, son raras e incluso específicas de casos individuales o familias, de origen reciente y confieren alto riesgo (Walsh *et al.*, 2008).

Una vez confirmado el hecho de que se trata de una enfermedad heredable, las principales aproximaciones realizadas con el objetivo de determinar las bases moleculares y genéticas han sido el análisis de ligamiento y los estudios de asociación.

1.1.1 Estudios de ligamiento

Los estudios de ligamiento permiten identificar regiones cromosómicas que contengan genes involucrados en un desorden determinado a través del estudio de la cotransmisión de marcadores genéticos con la enfermedad en familias con varios individuos afectados. Estos estudios tienen la ventaja de no precisar de conocimientos previos acerca de la patofisiología de la enfermedad y permiten obtener información sobre la localización de genes que puedan predisponer o proteger frente al desarrollo de la enfermedad.

En el caso de la esquizofrenia, los resultados no han sido muy prometedores. En una crítica comparación de algunos de los principales metaanálisis de estudios de ligamiento realizados en esquizofrenia (ver tabla I1), Crow compara tres metaanálisis de búsquedas en el genoma para esquizofrenia y trastorno bipolar y tres de los estudios comparativos entre hermanos para esquizofrenia y desorden esquizoafectivo. Tras la comparación, se concluye la falta de replicación de resultados consistentes entre los diferentes estudios, aún en el caso de utilizar metodologías similares y en algún caso compartir parte de la muestra (Crow, 2007).

Lo que se puede concluir a partir de los resultados del análisis de ligamiento en esquizofrenia es que, aunque se ha encontrado ligamiento para la mayoría de los cromosomas, posteriores metaanálisis fallan en la replicación de los resultados. Aún así, algunos estudios como el de Arinami (Arinami *et al.*, 2005) han conseguido confirmar el ligamiento en 1p, 14q y 20p, utilizando análisis de ligamiento con elevada densidad de marcadores. Los loci más replicados hasta el momento son los siguientes: 1q, 2, 6p, 8p, 10, 13q, 22p (Brzustowicz *et al.*, 2000; Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8, 1996; Gurling *et al.*, 2001; DeLisi *et al.*, 2002).

Fenotipo utilizado	Número de individuos		Regiones cromosómicas	Estudio
	Nº de familias	Nº individuos		
Esquizofrenia, trastorno esquizoafectivo	681	1929	8p, 13q, 22q	Badner and Gershon, 2002 metaanálisis
Trastorno bipolar, trastorno esquizoafectivo y depresión unipolar	353	1228	13q, 22q	
Esquizofrenia y trastorno esquizoafectivo	1208	2945	1pq, 2pq, 2q, 3p, 5q, 6p ter22.3, 6p 22.1-21.1, 8p 22-21.1, 11q, 14pq, 20p, 22p ter-12	Lewis <i>et al.</i> , 2003 metaanálisis
Trastorno bipolar y trastorno esquizoafectivo	347	959	9p, 10q, 14q 24-32, 18q	Segurado <i>et al.</i> , 2003 metaanálisis
	Pares de hermanos			
Esquizofrenia y trastorno esquizoafectivo	382		2, 8, 10p, 14	DeLysi <i>et al.</i> , 2002
Esquizofrenia y trastorno esquizoafectivo	353		10q, 17, 22	Williams <i>et al.</i> , 2003
Esquizofrenia y trastorno esquizoafectivo	409		5, 8, 10, 11	Suarez <i>et al.</i> , 2006

Tabla I1. Estudios de ligamiento comparados (tomado de Crow, 2007).

De este modo, aunque este tipo de aproximación ha sido productiva para el caso de las enfermedades de herencia mendeliana, los resultados han sido limitados cuando se ha aplicado en el estudio de enfermedades complejas, como la esquizofrenia, donde los genes implicados sólo explican una pequeña fracción de la heredabilidad total observada. Los motivos de esta falta de resultados son diferentes: baja heredabilidad de algunos de algunos caracteres complejos, heterogeneidad fenotípica y genética, selección de marcadores no adecuados, imprecisión en la definición de los fenotipos, tamaños muestrales pequeños y bajo poder estadístico para identificar variantes genéticas comunes que tienen efecto bajo en la enfermedad.

1.1.2 Estudios de asociación caso-control

Los estudios de asociación están basados en la comparación de variantes genotípicas en una población para la cual se conoce información fenotípica (como presencia o ausencia de enfermedad en los estudios caso-control, o en un rango de valores de caracteres). En su forma más simple se comparan las frecuencias genotípicas o alélicas de un *locus* polimórfico entre pacientes y controles (caso-control), aunque también pueden utilizarse muestras familiares en las que los individuos no afectados se consideran la muestra control, evitando así el problema de la estratificación de la población. Si se encuentra relación estadísticamente significativa entre genotipo y fenotipo se dice que existe una asociación entre la variante y la enfermedad o rasgo.

A diferencia de los estudios de ligamiento, en los estudios de asociación se precisa de una hipótesis previa que permita seleccionar qué genes se van a estudiar. La selección tradicional de los genes está basada bien en su función (genes candidatos funcionales) o bien en su localización en regiones cromosómicas prometedoras, en su mayor parte definidas por estudios de ligamiento (genes candidatos posicionales).

Introducción

Durante los últimos años, se ha evaluado un elevado número de genes candidatos en relación con el estudio de la psicosis, principalmente genes relacionados con las rutas dopaminérgica y glutamatérgica. En el año 2005, se publicaron muchos trabajos sobre una serie de genes candidatos para la esquizofrenia (Harrison y Weinberger, 2005). Estos trabajos se basaban en los resultados obtenidos en estudios de ligamiento, de asociación o en su posible participación en procesos relacionados con la enfermedad y dieron lugar a una lista de genes que llegaron a denominarse "genes de la esquizofrenia" (Harrison y Weinberger, 2005; Owen *et al.*, 2005). Estos genes que parecían estar consistentemente asociados con la esquizofrenia eran los siguientes: *COMT* (catecol-O-metiltransferasa), *DTNBP1* (disbindina), *NRG1* (Neuregulina 1), *RGS4* (Regulador 4 de la señalización de la proteína G), *GRM3* (Receptor metabotrópico 3 del glutamato), *DISC1* (*Disrupted-in-schizophrenia*), *G72* (Activador de la D-amino oxidasa), *DAAO* (D-amino oxidasa), *PPP3CC* (subunidad gamma catalítica), *CHRNA7* (Receptor $\alpha 7$ nicotínico), *PRODH2* (Prolina deshidrogenasa), *Akt1* (Proteína kinasa B). Desgraciadamente, los prometedores resultados obtenidos inicialmente no han podido ser replicados para la mayoría de estos genes y se ha ido descartando su posible relación con la esquizofrenia.

En la actualidad, existe un nuevo *ranking* de genes potencialmente implicados en esquizofrenia que se actualiza periódicamente en el foro internacional de esta enfermedad (<http://www.schizophreniaforum.org/>). En esta nueva lista únicamente se incluyen genes que contienen al menos una variante que muestre una OR (*odds ratio*) resumen de los estudios realizados significativa y la valoración de los genes situados en las principales posiciones se lleva a cabo por efecto del tamaño de la muestra de los estudios realizados (Allen *et al.*, 2008). En la tabla I2, se muestra la lista de los 10 genes considerados con mayor probabilidad de conferir susceptibilidad a esquizofrenia, según *Sczgene* de *Schizophrenia Forum*, así como el cromosoma en el que se localizan y el número de estudios de asociación caso control realizados en población caucásica y asiática. Estos datos fueron tomados en septiembre de 2008. El primer gen de la lista es el gen *SLC8A1* que codifica para la proteína "Miembro 1 de la Familia 18 de Portadores de Solutos (Monoamina Vesicular)", para el que se han obtenido resultados positivos en estudios caso-control tanto en población caucásica como asiática (Bly *et al.*, 2005; Chen *et al.*, 2007; Lohoff *et al.*, 2008; Talkowski *et al.*, 2008; Richards *et al.*, 2006) aunque también se han encontrado resultados negativos en estudios con familias (Talkowski *et al.*, 2008). En el resto de genes propuestos (tabla I2) se observa que el único que se mantiene de los relacionados en el año 2005 es el gen *DAOA* (también denominado *G27*), que codifica para el Activador de la D-amino Oxidasa.

En esta tabla se observa que aunque se han obtenido resultados positivos para estos genes, considerados como los más probables para estar relacionados con esquizofrenia, también hay estudios negativos, por lo que sigue sin tenerse un gen candidato con resultados lo suficientemente consistentes como para convencer a la comunidad científica.

Algunas de las razones esgrimidas para explicar estos resultados coinciden con las expuestas en el caso de los análisis de ligamiento, como son: el pequeño tamaño muestral, la heterogeneidad genética, la falta de definición del fenotipo o limitaciones estadísticas. Pero, sobre estos argumentos está el hecho de que las variantes utilizadas, en su mayoría SNPs (polimorfismos de un solo nucleótido), tengan un impacto menor en la expresión del gen o en su función y en definitiva un moderado efecto sobre el fenotipo.

Gen	Proteína	Cromosoma	Número de estudios de asociación	
			Población caucásica	Población asiática
<i>SLC8A1</i>	Miembro 1 de la familia 18 de portadores de solutos (monoamina vesicular)	8	3 positivos	2 positivos
<i>GRIN2B</i>	Subunidad 2B del receptor de N-metil-D-aspartato	12	1 positivo 1 tendencia 1 negativo	3 positivos 3 negativos
<i>GABRB2</i>	Receptor GABA beta 2	5	4 positivos 1 negativo	4 positivos 1 negativo
<i>DRD2</i>	Receptor de la dopamina D2	11	13 positivos 26 negativos	6 positivos 1 tendencia 13 negativos
<i>PLXNA2</i>	Plexina 2	1	2 positivos	2 negativos 1 positivo
<i>RPGRIP1L</i>	Proteína RPGR-que interacciona con protein 1-like	16	7 positivos	-
<i>TPH1</i>	Triptófano hidrolasa 1	11	2 positivos 1 negativo	2 positivos 3 negativos
<i>DRD4</i>	Receptor de la dopamina D4	11	1 positivo 1 tendencia 22 negativos	6 positivos 9 negativos
<i>IL1B</i>	Interleukina 1 beta	2	3 positivos 4 negativos	4 negativos
<i>DAOA</i>	Activador de la D-amino oxidasa	13	8 positivos 5 negativos	4 positivos

Tabla 12. Lista de los 10 genes considerados con mayor probabilidad de conferir susceptibilidad a esquizofrenia, según *Sczgene* de *Schizophrenia Forum*.

Un nuevo tipo de aproximación a los estudios de asociación son los *Genome Wide Association Studies* (GWAS). Estos estudios de asociación exploran el genoma en su conjunto con el objetivo de encontrar variantes génicas causales de una enfermedad. Se trata de la genotipación masiva de SNPs (uno de los chips más utilizados, de la casa comercial Illumina® contiene 500.000 SNPs) con el objeto de cubrir una gran parte de las variantes comunes del genoma humano. Puesto que no se necesita un conocimiento de la localización de las variantes causales, esta aproximación puede explotar las ventajas de los estudios de asociación sin tener que adivinar la identidad de los genes causales. Algunas de las limitaciones que presentan los estudios GWAS son: la necesidad de un conocimiento de la variación genética a seleccionar (limitación solucionada con la aparición de bases de datos como la del *hapmap*), la capacidad para genotipar un número suficiente de variantes (actualmente solucionada con las infraestructuras disponibles) y que se necesitan muestras de gran tamaño (problema solventado con la aparición de consorcios entre grupos de investigación) (Hirschhorn y Daly, 2005).

En todo caso, hay que tener en cuenta que los GWAs son sensibles a estructuraciones poblacionales y que si los *loci* detectados tienen un pequeño efecto sobre el fenotipo, el poder para detectarlos es bajo. Los primeros estudios GWAS en esquizofrenia, realizados con muestras de tamaño modesto (500-800 casos) han mostrado ningún y un único resultado significativo, en un gen que codifica para un receptor de citocinas (Sullivan *et al.*, 2008; Lencz *et al.*, 2007, respectivamente).

Introducción

Recientemente, otra aproximación al estudio genético de la esquizofrenia son los estudios en los que se analizan variaciones poco frecuentes en el número de copias (CNVs poco frecuentes). Los dos primeros estudios a gran escala realizados hasta la fecha mostraron que el número de deleciones o duplicaciones de baja frecuencia aumenta en pacientes con esquizofrenia (International Schizophrenia Consortium, 2008; Stefansson *et al.*, 2008). Este aumento aunque modesto, es estadísticamente significativo. Ambos estudios identificaron las mismas dos regiones cromosómicas (1q21.1 y 15q13.3) en las que la variación en el número de copias está asociada con el riesgo a la esquizofrenia. De estas dos regiones, la primera de ellas ya había sido identificada como potencial región con genes candidatos en función de diferentes estudios de ligamiento (Brzustowics *et al.*, 2000; Gurling *et al.*, 2001).

1.1.3 Epigenética y esquizofrenia

La falta de resultados consistentes en estudios de ligamiento y asociación ha llevado a algunos autores a postular que además de la vulnerabilidad que pueda ser debida a variaciones en la secuencia del DNA, también intervienen factores epigenéticos, es decir, mecanismos que regulan la actividad génica sin alterar el código genético o secuencia de DNA (Kato *et al.*, 2002, Petronis, 2004).

Las modificaciones que afectan a la remodelación de la cromatina, como la acetilación, ubiquitinación, metilación o defosforilación de histonas, o la metilación del DNA, constituyen mecanismos conocidos que pueden afectar a la expresión de los genes. Entre estos, la metilación constituye el mecanismo que ha sido objeto de mayor atención. En el genoma humano la metilación del DNA sólo tiene lugar en las citosinas (así como la modificación de las histonas relacionadas) por lo que se puede crear un patrón diferente en regiones ricas en GC respecto de fragmentos ricos en adenina o timina (Gruenbaum *et al.*, 1982). A esto se añade el hecho de que el dinucleótido CpG es tanto la principal diana de metilación en el DNA de mamíferos como un punto caliente de mutaciones. Así, la existencia de regiones ricas en GC cerca de los genes se relaciona directamente con una mayor o menor expresión del gen, en función del estado metilado o no de las mismas (Bird, 1986; Hapgood *et al.*, 2001; Costello y Plass, 2001).

La regulación epigenética se asocia cada vez más a desórdenes psiquiátricos, encontrándose ejemplos en depresión o adicción, entre otros (Tsankova *et al.*, 2007). Se tienen bastantes pruebas de que los procesos epigenéticos afectan al desarrollo neural, y recientemente han surgido evidencias de su papel en la regulación de neuronas maduras diferenciadas (Hsieh y Cage, 2005).

En el caso de la esquizofrenia, los mecanismos epigenéticos podrían explicar la discordancia entre gemelos monozigóticos, la ausencia de diferencias entre casos esporádicos y familiares o el variable curso de la enfermedad, además de contribuir a la heterogeneidad fenotípica (Petronis, 2003; Petronis, 2004). Petronis aboga por la existencia de cambios epigenéticos dinámicos desde el nacimiento que aumenten la vulnerabilidad a sufrir esquizofrenia, aunque únicamente aquellos individuos que alcancen un nivel de desregulación epigenética determinado, debido a diferentes factores incluido el ambiental, mostrarán los signos clínicos de la enfermedad. La severidad de los cambios epigenéticos podría fluctuar en el tiempo, ocasionando los períodos sin síntomas y las recaídas (Figura I2).

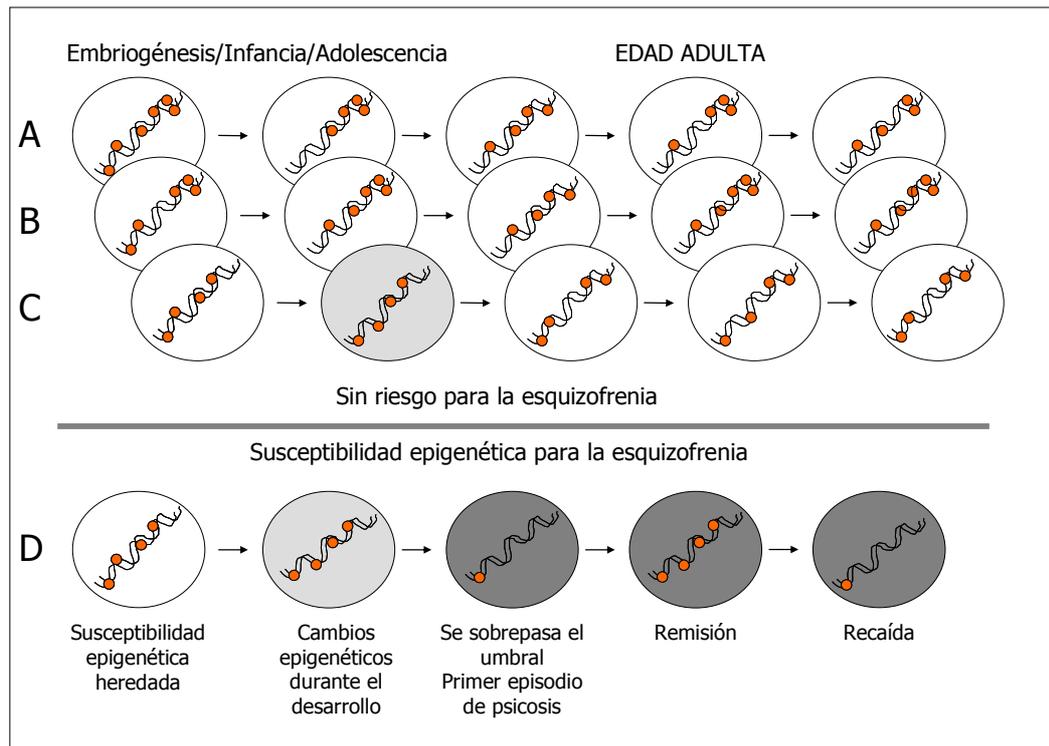


Figura I2. Cambios epigenéticos durante el desarrollo. Los escenarios A, B y C muestran el desarrollo epigenético normal de un gen que podría predisponer a esquizofrenia (círculos blancos representan individuos sanos y círculos grises, individuos límite). El escenario D representa como una pre-epimutación se convierte en un problema epigenético severo que da lugar a esquizofrenia (círculos negros). En el caso de los escenarios C y D el desarrollo epigenético es muy diferente así como el pronóstico clínico, a pesar de comenzar de un punto inicial idéntico. Tomado de Petronis 2004, con modificaciones.

Existen evidencias de que procesos de metilación diferencial participan en la patogénesis de la esquizofrenia. Se ha postulado que la hipermetilación de determinados genes en neuronas GABAérgicas cause disfunción del circuito neuronal GABAérgico (teoría GABAérgica), de forma que las funciones cerebrales se verían alteradas debido a la ruptura de la sincronización con otras redes neurales (Costa *et al.*, 2004). Uno de los casos más estudiados es el del gen que codifica para la reelina, en el que se ha observado que una hipermetilación del promotor está relacionada con niveles más bajos de expresión en cerebros *postmortem* de individuos con esquizofrenia (Chen *et al.*, 2002; Grayson *et al.* 2005), aunque en el último estudio realizado no se replicaron los resultados previos (Tochigi *et al.*, 2008).

Los primeros resultados obtenidos a partir de genes discretos han impulsado el desarrollo de análisis epigenómicos a mayor escala. Algunos de los primeros resultados de estos estudios muestran evidencias de diferencias en la metilación asociadas a psicosis en numerosos *loci* involucrados en procesos de neurotransmisión y desarrollo neuronal (Mill *et al.*, 2008).

Indirectamente relacionada con la teoría epigenética se encuentra la teoría del origen común entre psicosis y lenguaje, propuesta por Crow, (última revisión en Crow, 2008), según la cual un único gen sujeto a control epigenético (gen que codifica para la protocaderina X) es el responsable del origen del lenguaje, apareciendo la esquizofrenia como residuo del proceso de especiación.

En resumen, aunque se ha avanzado mucho en el estudio de la genética de la esquizofrenia, todavía no se han identificado genes que inequívocamente estén

Introducción

relacionados directamente con esta enfermedad y gran parte de los resultados obtenidos son contradictorios.

Una de las posibles causas es que los métodos y herramientas utilizados hasta el momento no resultan lo suficientemente potentes o resolutivos para detectar las variables genéticas implicadas en el desarrollo de la esquizofrenia. La continua dinámica de la investigación ha hecho surgir nuevas aproximaciones destinadas a solventar esta cuestión, como la de los ya mencionados GWAs, los recientes estudios a nivel genómico con CNVs, los análisis de expresión a nivel global o la posible implicación de factores epigenéticos más allá de la secuencia de DNA.

Sin embargo, subyacente a todas las aproximaciones aplicadas a la búsqueda de los genes implicados en la esquizofrenia está el hecho de que la esquizofrenia es una entidad definida por consenso y no tiene porqué tener una validez biológica (Sanjuán, 2007). Esta enfermedad constituye un fenotipo muy amplio y heterogéneo, lo que dificulta su estudio mediante técnicas genéticas que precisan de un fenotipo concreto y bien definido.

II. Alucinaciones auditivas como fenotipo alternativo de la esquizofrenia

II.1 Endofenotipos y fenotipos alternativos en el estudio de la esquizofrenia

La selección de un fenotipo adecuado para llevar a cabo los estudios de asociación constituye un paso fundamental para el éxito de un experimento. La heterogeneidad de la esquizofrenia se ve reflejada en la expresión variable de los síntomas en múltiples dominios (percepción, lenguaje, atención, expresión de emociones...), en la variabilidad de aparición de los síntomas, la respuesta al tratamiento, la evolución de la enfermedad y otros factores que determinan su pronóstico (Weiser *et al.*, 2005). Este hecho hace que la capacidad predictiva del fenotipo desde el genotipo se vea comprometida. Además de considerar la esquizofrenia como un fenotipo en si mismo, son numerosos los estudios que plantean dos posibilidades: ampliar el fenotipo y considerar la existencia de una vulnerabilidad biológica a un continuo entre la esquizofrenia y psicosis maniaco-depresiva, o bien buscar fenotipos más simples para poder obtener grupos más homogéneos con los que comparar variables biológicas (Sanjuán, 2007). Considerando la esquizofrenia como un síndrome demasiado complejo para poder ser afrontado como un todo, al igual que en otras enfermedades heterogéneas, muchos investigadores tratan de abordar el tema mediante la utilización de fenotipos alternativos y endofenotipos (Gottesman y Gould, 2003; Bearden y Freimer, 2006; Cannon y Keller, 2006). Un fenotipo alternativo puede consistir en el síntoma o conjunto de síntomas clínicos que se definen en sustitución de los criterios DSM-IV de esquizofrenia (por ejemplo, las alucinaciones auditivas, déficit cognitivo, algunos fenotipos extremos o subtipos clínicos, entre otros). Un endofenotipo se trata de un componente medible de la ruta patofisiológica que se toma como base para correlacionar con los resultados obtenidos (Weiser *et al.*, 2005). Los endofenotipos en esquizofrenia pueden ser neurofisiológicos, bioquímicos, endocrinos, neuroanatómicos, cognitivos, o rasgos de la personalidad y representan una forma más simple de la base genética o ambiental que el síndrome de la enfermedad por sí mismo (por ejemplo, se

suelen utilizar, entre otros, el déficit en la memoria de trabajo, déficit en memoria episódica, movimientos desordenados de los ojos, o inhibición atenuada del potencial P50 y del P300).

En el caso de este estudio, se abordó el estudio de la esquizofrenia tomando como fenotipo alternativo las alucinaciones auditivas.

II.2 Alucinaciones auditivas

A partir de la definición de alucinación en sentido amplio elaborada por David (David, 2004), enmarcándola dentro del sistema sensorial correspondiente a las alucinaciones auditivas (AA), podemos definir éstas como *"experiencias sensoriales que ocurren sin la estimulación externa correspondiente del oído, que tienen el suficiente sentido de la realidad como para parecer una percepción verídica, sobre la cual el sujeto no siente que tiene control directo y que ocurre estando el sujeto despierto"*

Este tipo de alucinaciones, pueden variar tanto en forma como en contenido, de manera que pueden incluir desde sonidos primitivos como golpes, soplos, disparos, o truenos, hasta lloros risas, susurros o conversaciones (Nayani y David, 1996; Watkins, 1998). En el caso específico de las alucinaciones verbales pueden consistir en una voz que diga los pensamientos del individuo en voz alta, una voz que comente el comportamiento de la persona o voces que dan órdenes o instrucciones.

Las características fenomenológicas de las AA pueden ser tanto físicas (frecuencia, volumen o claridad), como personales (contenido emocional, enunciado en primera, segunda o tercera persona, número de voces o el carácter conocido de la voz).

A menudo este tipo de alucinación tiene una calidad negativa, y consiste en voces que insultan y critican al paciente, le dicen que haga cosas inaceptables o lo amenazan (Close y Garety, 1998; Davies *et al.*, 2001; Delespaul *et al.*, 2002; Johns *et al.*, 2002). La frecuencia con que las alucinaciones tienen un contenido emocional, unido a una posible existencia de alteraciones en los aspectos emocionales del habla en individuos con predisposición a sufrir AA, que aumentan su propensión a experimentarlas (Woodruff, 2004), han llevado a proponer que las AA podrían estar asociadas con déficits en el procesamiento afectivo auditivo (Rossell *et al.*, 2005).

Es importante tener en cuenta que tanto las AA como las alucinaciones en general no están limitadas a un único diagnóstico como la esquizofrenia, sino que se presentan en otros desórdenes psiquiátricos, e incluso en población normal (Brasic, 1998; David, 1999; Johns, 2005).

Dentro de las alucinaciones clínicas pueden distinguirse tres tipos: las observadas en desórdenes psicóticos, las observadas en desórdenes neurológicos y aquellas observadas en un contexto de abuso de drogas. Aunque estos tres tipos compartan mecanismos comunes, ciertas características parecen dominar en cada tipo en función de sus características fenomenológicas y los mecanismos biológicos y psicológicos subyacentes. Por ejemplo, en esquizofrenia y desórdenes afectivos son más frecuentes las AA, mientras que en la enfermedad de Parkinson o Alzheimer, las más frecuentes son las visuales (Aleman y Laroi, 2008).

Introducción

Las similitudes entre los grupos alucinadores clínicos y no clínicos, teniendo en cuenta todos los tipos de alucinaciones (auditivas, visuales, y otros tipos) han hecho que algunos investigadores sugieran que las alucinaciones pueden formar parte de un continuo con experiencias normales. Esta hipótesis, denominada *hipótesis del continuum*, argumenta que la mayor diferencia entre los grupos patológico y normal es cuantitativa más que cualitativa, por lo que quizás no sea la naturaleza de las experiencias alucinatorias lo que cause que las personas se conviertan en pacientes psiquiátricos, sino la manera en la que el individuo reacciona a sus experiencias (Strauss 1969; van Os *et al.*, 2000; Bentall, 2003).

En cualquier caso, las AA cumplen una serie de características que las convierten en un buen fenotipo alternativo para el estudio de la vulnerabilidad genética a la esquizofrenia:

- Con una prevalencia del 56.2-70% (Slade y Bentall, 1988; Andreasen y Flaum, 1991) son el tipo de alucinación más común en esquizofrenia, seguidas de las alucinaciones visuales, táctiles, olfativas y sensoriales.
- Se heredan (Rosenthal y Quinn, 1977).
- Se pueden cuantificar, por medio de diferentes escalas. En este estudio en concreto, se utiliza la escala PSYRATS, escala que evalúa las dimensiones de las alucinaciones auditivas (Haddock *et al.*, 1999; versión en español de González *et al.*, 2003).
- Se pueden observar indirectamente por medio de la neuroimagen (revisiones en Allen *et al.*, 2008 y Aguilar *et al.*, 2008).
- Están estrechamente relacionadas con otro de los síntomas de la esquizofrenia, como son las alteraciones del lenguaje (Hoffman, 1986; Stephane *et al.*, 2001).

Existen evidencias de la existencia de déficits en el dominio del lenguaje en pacientes esquizofrénicos (DeLisi, 2001; Covington, 2005; Stevens *et al.*, 2000; Mckenna y Oh, 2005). Así, distintos autores apuntan a que pacientes esquizofrénicos con alucinaciones se caracterizan por déficits en la percepción del lenguaje. Por ejemplo, McKay y colaboradores encontraron que los principales déficits sensoriales de sus pacientes estaban relacionados con estímulos del habla (McKay *et al.*, 2000) y diferentes estudios han mostrado una disminución en la ventaja del oído derecho durante escuchas dicóticas en pacientes esquizofrénicos con AA respecto a controles. (Bruder *et al.*, 1995, Green *et al.*, 1994; Levitan *et al.*, 1999; Loberg *et al.*, 2004).

Todo esto lleva a pensar que las AA se deben a un aumento de la actividad en las regiones cerebrales del habla, similar a la producida por el habla real, de forma que se reduce la capacidad de procesar los estímulos del oído derecho. Aunque otra posibilidad sería que una alteración cerebral en el hemisferio izquierdo causara la reducción en la ventaja del oído derecho.

II.2.1 Estudios de neuroimagen en alucinaciones auditivas

Al igual que en la esquizofrenia, durante los últimos años, las técnicas de neuroimagen funcional, como la PET y la resonancia magnética funcional (fMRI) han permitido el estudio de las áreas cerebrales involucradas en las alucinaciones.

Aunque no todos los estudios de neuroimagen estructural muestran los mismos resultados, la mayoría de ellos sugiere una asociación de volúmenes de materia gris reducidos en el lóbulo temporal con las AA, concretamente en la circunvolución temporal superior izquierda, incluyendo la corteza auditiva primaria (Barta *et al.*, 1990; Flaum *et al.*, 1994; Rajarethinam *et al.*, 2000). También se ha observado una reducción en los volúmenes de la corteza cerebelar y prefrontal que pueden estar asociados con alteraciones en la monitorización o conciencia y reconocimiento del habla interior.

Los estudios de neuroimagen funcional muestran que, comparado con el cerebro no alucinador, el cerebro alucinador se caracteriza por una entrada más fuerte desde los centros subcorticales (especialmente el tálamo), reducido control del córtex prefrontal dorsolateral, activación aberrante de los centros de atención emocional (amígdala, cíngulo anterior ventral) y activación atenuada del cíngulo anterior dorsal, lo que está relacionado con monitorización de errores (Aleman y Laroi, 2008). Para las AA, el córtex de asociación sensorial debe ser el área auditiva secundaria (área de Broadmann 42) y las áreas de percepción auditiva adyacentes áreas como el surco temporal superior. Para las alucinaciones verbales específicamente se postula un acoplamiento entre áreas de producción del lenguaje (área de Broca) y áreas de recepción del lenguaje (área de Wernicke) (Hoffman *et al.*, 2007).

II.2.2 Estudios genéticos sobre las alucinaciones auditivas

Pese a que las AA constituyen un fenotipo concreto y bien caracterizado, resulta sorprendente la ausencia de estudios acerca de los mecanismos etiológicos que las ocasionan. Hasta el momento se han llevado a cabo pocos estudios genéticos teniendo en cuenta las alucinaciones, y menos todavía con AA.

Los primeros de ellos estaban basados en las hipótesis de que implican a neurotransmisores de las rutas serotoninérgica o dopaminérgica en la aparición de este síntoma (Malhotra *et al.*, 1998; Wei *et al.*, 1999; Sanjuán *et al.*, 2004; Toirac *et al.*, 2007), aunque posteriormente se ha ampliado el estudio hacia genes involucrados en el neurodesarrollo, y más específicamente en el lenguaje, como el gen *FOXP2* (Sanjuán *et al.*, 2006), o en el tamaño cerebral, como el gen *ASPM* (Rivero *et al.*, 2006).

En la tabla I3 se resumen los resultados obtenidos hasta el momento.

Así, el campo del estudio genético de las alucinaciones, está todavía en desarrollo y todavía es necesaria la realización de estudios que confirmen los resultados obtenidos hasta el momento.

Introducción

Autores	Desorden estudiado	Genes estudiados	Resultados
Malhotra <i>et al.</i> , 1998	Alucinaciones Esquizofrenia	<i>5-HTT</i>	Asociación entre alucinaciones y la variante I/I polimorfismo del promotor del gen
Wei <i>et al.</i> , 1999	Alucinaciones auditivas en esquizofrenia	<i>CCK-AR</i>	Asociación positiva
Okubo <i>et al.</i> , 2002	Alucinaciones en <i>delirium tremes</i>	<i>CCK-AR</i>	Asociación positiva
Sanjuán <i>et al.</i> , 2004	Alucinaciones auditivas en esquizofrenia	<i>CCK-AR</i>	Asociación positiva
Chang <i>et al.</i> , 2004	Alucinaciones en Alzheimer	<i>ApoE ε4</i>	Asociación positiva
Ott <i>et al.</i> , 2005	Sugestionabilidad relacionada con alucinaciones	<i>5HT2A</i>	Asociación del rasgo con una variante del polimorfismo T102C. Interacción genotipos T102C con genotipos del gen <i>COMT</i>
Sanjuán <i>et al.</i> , 2005	Alucinaciones auditivas	<i>FOXP2</i>	Asociación negativa
Sanjuán <i>et al.</i> , 2006a	Alucinaciones auditivas	<i>5-HTT</i>	Asociación positiva
Sanjuán <i>et al.</i> , 2006b	Alucinaciones auditivas	<i>FOXP2</i>	Asociación positiva con un polimorfismo de <i>FOXP2</i> y con un haplotipo
Rivero <i>et al.</i> , 2006	Alucinaciones auditivas	<i>ASPM</i>	Asociación negativa
Toirac <i>et al.</i> , 2007	Alucinaciones auditivas	<i>CCK-AR</i>	Asociación positiva
Tolosa <i>et al.</i> , 2008	Alucinaciones auditivas	<i>HAR1A</i>	Tendencia

Tabla I3. Resumen de los estudios genéticos realizados en alucinaciones auditivas.

II.2.3 Modelos de alucinaciones auditivas

Recopilando toda la información disponible sobre las alucinaciones, Aleman y Laroi (Aleman y Laroi, 2008) proponen un modelo multifactorial que incorpora las alteraciones cognitivas y neurales que se conocen.

Estos autores proponen 4 rutas diferentes que pueden resultar en alucinaciones:

- Alteraciones en las vías perceptuales debidas a lesiones en rutas sensoriales o en el sistema de excitación.
- Procesos que actúen sobre centros corticales que integran la información perceptiva.
- Activación no limitada del centro de atención.
- Ruta cognitiva con papeles clave del estado afectivo, factores *top-down* y el sistema de monitorización propio.

Según estos autores, las dos últimas rutas serían las más probables en la mayoría de alucinaciones, Además, proponen que la alta frecuencia de alucinaciones verbales en esquizofrenia y desórdenes psicóticos puede deberse a la desorganización de los sistemas de atención y del lenguaje, y a las entradas a estos sistemas desde el sistema de emoción y motivación.

En la figura I3 se muestra el modelo propuesto para las alucinaciones auditivas.

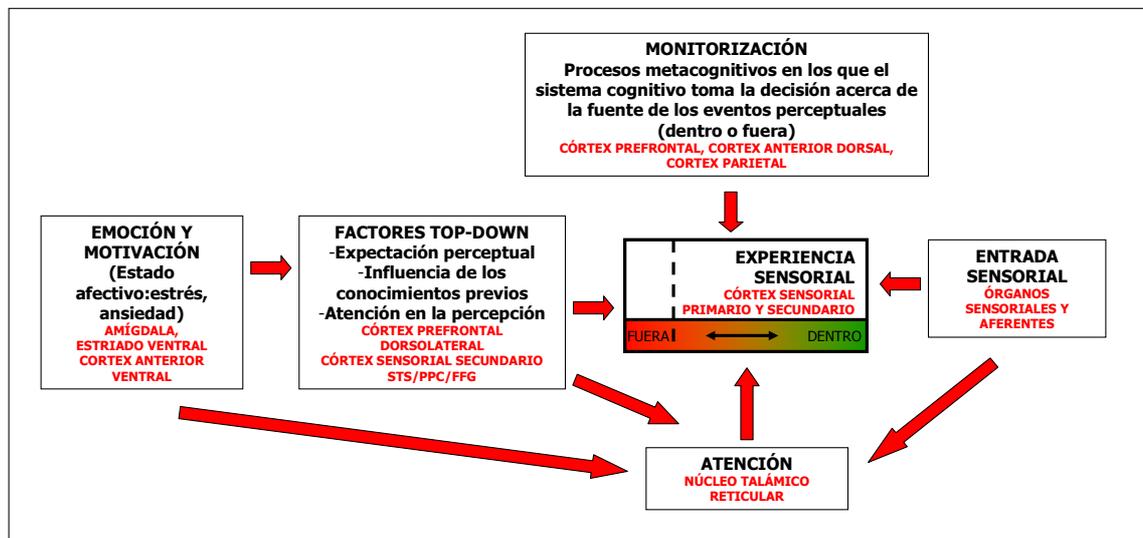


Figura I3. Modelo de las alucinaciones auditivas. Se muestran los diferentes sistemas que pueden afectar desarrollo de alucinaciones auditivas, junto con las áreas cerebrales que participan en cada uno de ellos, así como los factores genéticos pueden estar influyendo en cada caso (Tomado de Aleman y Laroi 2008 con modificaciones).

Nuestro grupo de investigación propone un modelo etiopatogénico alternativo de las alucinaciones, mostrado en la figura I4, en el que factores genéticos y factores ambientales del desarrollo crearían un fondo en el que el individuo sería propenso a tener alucinaciones y que bajo determinados factores como estrés, privación sensorial, consumo de drogas o enfermedades cerebrales desencadenaría la experiencia alucinatoria (Sanjuán *et al.*, 2005).

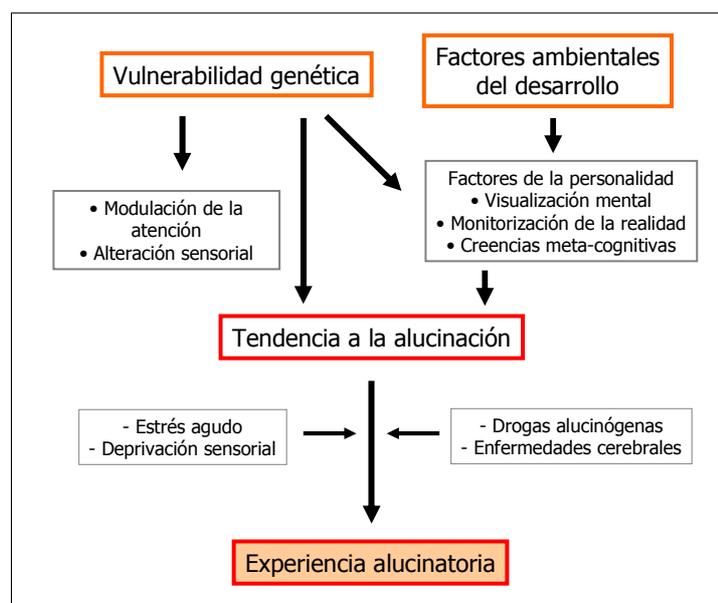


Figura I4. Modelo etiopatogénico de alucinaciones (Sanjuán *et al.*, 2005).

Introducción

Aplicado al caso de las alucinaciones auditivas en psicosis, este modelo, reflejado en la figura I5, propone (Aguilar *et al.*, 2008):

- Participación de factores genéticos a diferentes niveles: ocasionando una vulnerabilidad a los desórdenes del lenguaje, en la que intervendrían genes relacionados con este rasgo, que daría lugar a la experiencia de voces; dando lugar a vulnerabilidad a tener una respuesta emocional anormal, en la que intervendrían genes relacionados con el estado emocional y proporcionando susceptibilidad a experimentar voces y reaccionar de forma anormal a las mismas.
- Participación de factores culturales, que afecten al contenido y adaptación social de las voces.

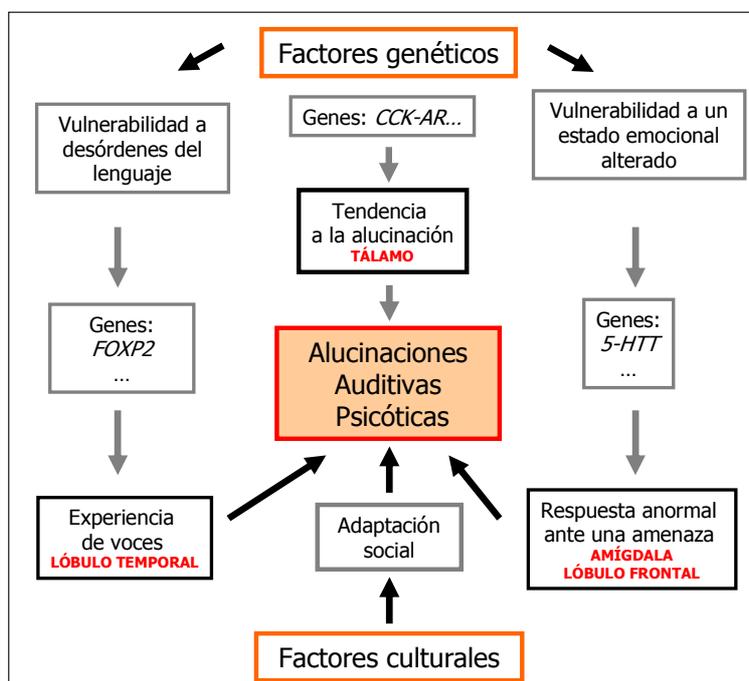


Figura I5. Modelo etiopatogénico de alucinaciones auditivas en psicosis (Tomado de Aguilar *et al.*, 2008 con modificaciones).

III. Lenguaje

El lenguaje es una característica definitoria de la especie humana. Se puede considerar que "el lenguaje humano es un sistema computacional rico que coordina simultáneamente representaciones sintácticas, semánticas, fonológicas y pragmáticas, sistemas sensoriales y motores y el conocimiento del mundo tanto del individuo que habla como del que escucha" (Fisher y Marcus, 2006). En la práctica, una persona normal tiene acceso a decenas de miles de palabras y puede, mediante un complejo juego de reglas estructurales, ensamblarlas en un número potencialmente infinito de sentencias con sentido, para referirse a hechos pasados, presentes o futuros, reales o abstractos. Este sistema se adquiere sin un esfuerzo consciente o una instrucción formal (Pinker, 2003).

Sin embargo, la comunicación no humana está restringida normalmente a simples mensajes con poca complejidad estructural, como las llamadas de alarma y de identificación de señales.

Bases neurológicas del lenguaje

Los primeros estudios realizados en lesiones cerebrales relacionadas con alteraciones del lenguaje dieron lugar a un modelo tradicional en el que el área de Broca (circunvolución frontal inferior) y el área de Wernicke (en la circunvolución temporal superior) constituían sustratos específicos del lenguaje, que servían para las funciones de producción del habla y/o gramática y significado o comprensión, respectivamente (Damasio, 1992).

En la actualidad se considera que estas dos áreas forman parte de un circuito mayor, todavía poco conocido en la que están involucradas diferentes áreas de la corteza cerebral, junto con estructuras subcorticales como el tálamo, ganglios basales, y el cerebelo, estos dos últimos implicados en aprendizaje y ejecución de secuencias de acciones motoras o procesos cognitivos (Poeppel y Hickok, 2004; Demonet *et al.*, 2005). Los estudios de neuroimagen refuerzan la idea de que los diferentes aspectos del procesamiento del lenguaje están relacionados con un amplio rango de estructuras (Demonet *et al.*, 2005; revisión en Bookheimer, 2002).

Origen del lenguaje

El origen del lenguaje es un campo de estudio que ha generado gran controversia durante las últimas décadas, de forma que la mayoría de las cuestiones planteadas siguen abiertas. ¿Puede decirse que la evolución adaptativa ha modelado el lenguaje al igual que sucede en otros rasgos? ¿Surgió como una entidad independiente o como un producto secundario de otras propiedades? ¿Se originó de forma gradual o en un único proceso evolutivo brusco? Estas son sólo algunas de las preguntas planteadas. Hasta hace poco, la mayor parte del estudio del origen del lenguaje se llevaba a cabo mediante registros arqueológicos, reconstrucciones lingüísticas y modelos computacionales (Fisher y Marcus, 2006). Sin embargo, los resultados obtenidos por estos métodos no fueron concluyentes. Por esta razón, surgió como una poderosa herramienta el método comparativo, que utiliza datos empíricos de especies vivas para obtener inferencias de ancestros extintos. Inicialmente este estudio se realizó en primates no humanos, aunque el pensamiento actual procedente de la neurociencia, biología molecular y biología del desarrollo indica que muchos aspectos de la función neural y del desarrollo están altamente conservados, por lo que el método comparativo se ha ampliado a todos los vertebrados.

Esta aproximación comparativa ha llevado entre otros descubrimientos a la detección de las llamadas neuronas espejo, un grupo de neuronas localizadas en la corteza premotora del macaco, que se activan no sólo cuando el animal ejecuta determinadas acciones, sino también cuando observa movimientos similares en otro individuo (Rizzolatti *et al.*, 1996; Ferrari *et al.*, 2003). Este área además, se presenta como homóloga al área de Brodmann 45 en humanos, que forma parte del área de Broca. Así, Rizzolatti propone la hipótesis del Sistema Espejo, según la cual el área de Broca en humanos evolucionó a partir de un mecanismo básico, no relacionado originalmente con la comunicación, el sistema de espejo del ancestro común de macaco y humano y la habilidad de este sistema para generar y reconocer un juego de acciones proporcionó las bases evolutivas de la paridad del lenguaje, en la que una unidad

Introducción

significa lo mismo para el que habla y el que escucha (Rizzolatti y Craighero, 2004). Curiosamente, se ha hipotetizado que este mismo sistema podría estar relacionado con la esquizofrenia y que las alucinaciones auditivas verbales se producen debido a que las rutas auditivas que están activas durante las alucinaciones producen el procesamiento verbal de la voz originada, pero puesto que no se mantiene registro de esa voz creada, se trata como externa (Arbib, 2007).

Alteraciones del lenguaje y estudios genéticos

La adquisición del lenguaje es universal en la especie humana. Sin embargo, una proporción significativa de niños tienen déficits relacionados con el lenguaje que no pueden ser explicados por una causa conocida, como retraso mental, parálisis cerebral o defectos auditivos. Para algunos de estos niños, estas dificultades con la comunicación se resuelven con la edad, aunque en otros casos se mantienen en el estado adulto. Muchas veces algunos de estos desórdenes solapan en sus definiciones o en su ocurrencia en individuos y familias, lo que lleva a plantearse si las causas subyacentes pueden solaparse o si pueden deberse a diferentes manifestaciones de un mismo déficit (Lewis *et al.*, 2006).

La identificación de los genes que afecten a estos desórdenes puede ayudar a definir la patogénesis de cada uno y demostrar sus relaciones en un nivel funcional:

- *Dislexia*

Dificultades en la lectura o deletreo, a pesar de aptitudes verbales normales y oportunidades educacionales adecuadas. A menudo asociada con alteraciones sutiles en el procesamiento del lenguaje.

- *Trastorno específico del lenguaje (TEL)*

Retraso significativo del habla a pesar de no tener daños en el oído, ausencia de problemas neurológicos conocidos e inteligencia normal.

- *Trastorno en la articulación del habla*

Retraso significativo en la adquisición de los sonidos del habla articulada, que constituye un subtipo de TEL.

- *Autismo*

Síndrome caracterizado por déficits en la interacción social recíproca y comunicación, acompañado de comportamientos repetitivos y estereotipados.

- *Dispraxia verbal del desarrollo*

Déficits en el aprendizaje y producción de secuencias de movimientos vocales a menudo acompañados por problemas en gramática y en el lenguaje expresivo o receptivo.

- *Esquizofrenia*

Aunque la esquizofrenia no constituye un desorden del lenguaje *per se*, como se comentó en apartados anteriores, las alteraciones del lenguaje en diferentes dimensiones, son uno de los principales síntomas presentes en este síndrome (DeLisi, 2001; Covington, 2005; Stevens *et al.*, 2000; Mckenna y Oh, 2005).

En la tabla I4 aparecen resumidos los resultados de los estudios genéticos más relevantes llevados con las alteraciones del lenguaje anteriormente mencionadas.

Déficit del lenguaje	Estudios familiares, de gemelos y de adopción	Estudios de ligamiento	Genes candidatos	Referencias
Dislexia	Prevalencia: 5% niños en edad escolar. Heredabilidad: 0.30-0.72 Concordancia en gemelos: MZ= 0.68, DZ=0.36	Las regiones que han sido replicadas de forma más consistente están en: 6p22.2 (DYX2), 15q21 (DYX1), 2p16-p15 (DYX3) y 1p36 (DYX8)	Genes candidatos: <i>EKN1 (DYX1)</i> , <i>ROBO1 (DYX5)</i> , <i>DCDC2 (DYX2)</i> , <i>KIAA019 (DYX2)</i>	Francks <i>et al.</i> , 2002 Bakwin 1973, DeFries <i>et al.</i> , 1987 Smith, 2007 Lewis, 2006
TEL	Prevalencia: 3-10% Heredabilidad: 0.45 Concordancia en gemelos: MZ=0.70-1.00 DZ=0.46-0.50	16q24 19q13 13q21 2p23 7q31	Aunque se han testado numerosos genes que se expresan en cerebro hasta el momento no existen genes relacionados directamente el TEL.	Tomblin <i>et al.</i> , 1997 Tomblin <i>et al.</i> , 1998 International SLI Consortium 2002, 2004 O'Brien, 2003 Bartlett, 2004 Smith, 2007
Trastorno en la articulación del habla (subtipo de TEL)	Prevalencia: 5% en niños de 6 años Heredabilidad: 0.97 Concordancia en gemelos: MZ=0.84-1.00 DZ=0.56-0.62	3q13 15q14	-	Shriberg <i>et al.</i> , 1999 Matheny y Bruggenmann, 1973 Stein <i>et al.</i> , 2004 Stein <i>et al.</i> , 2006 Smith, 2007
Autismo	Prevalencia: 5-10/10000 Concordancia en gemelos: MZ=0.69-0.95 DZ=0-0.24	Diferentes regiones cromosómicas implicadas. 3q, 7q o 13q21 posiblemente implicadas en lenguaje.	Aunque se han testado numerosos genes que se expresan en cerebro hasta el momento no existen genes relacionados directamente con los déficits en el lenguaje presentes en el autismo.	Folstein y Rutter, 1977 Yeargin-Allsopp <i>et al.</i> , 2003 Muhle <i>et al.</i> , 2004 Fombonne, 2003 Fisher y Marcus, 2006
Dispraxia Verbal del Desarrollo	Prevalencia:0.00125	-	Una forma mendeliana es causada por mutaciones en el gen <i>FOXP2</i> en 7q31	Felsenfeld, 2002 Shriberg, 1997

Tabla I4. Resumen de resultados de estudios familiares, ligamiento, asociación y genes candidatos en diferentes desórdenes del lenguaje.

IV. El gen *FOXP2*

Hasta el momento *FOXP2* es el primer gen que se ha relacionado con una alteración del lenguaje. Inicialmente, su descubrimiento fue acogido con grandes expectativas, sin embargo, si bien ha proporcionado valiosa información sobre algunos de los mecanismos que pueden participar en el desarrollo del lenguaje, todavía quedan muchas cuestiones abiertas. Así, aunque la consideración inicial de que *FOXP2* es el "gen del lenguaje", sea errónea, puesto que debe haber otros genes implicados, el estudio de este gen sigue aportando información sobre procesos que participan en plasticidad neuronal y desarrollo cerebral.

IV.1 Descubrimiento

A principios de los años 90, apareció publicado el caso de una familia inglesa de tres generaciones en la que la mitad de los individuos estaban afectados por alteraciones en la adquisición del habla y del lenguaje (Hurst *et al.*, 1990; Gopnik y Crago, 1991). Lo sorprendente del caso era que parecía tratarse de una alteración que seguía un patrón de herencia dominante monogénica (figura I6) de manera que los individuos que no tenían el trastorno no habían heredado la copia mutada del potencial gen de riesgo. El elevado número de individuos de la familia KE, como se denomina a esta familia, junto con el hecho de que los niños habían crecido en un ambiente similar, confirmaban de forma indirecta el carácter monogénico.

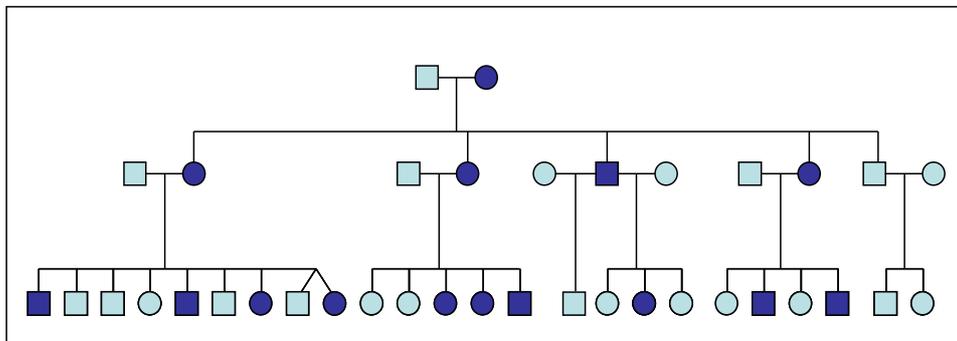


Figura I6. Genealogía de la familia KE (Tomado de Lai *et al.* 2001), con modificaciones).

Dada la importancia del hallazgo, la caracterización del trastorno se convirtió rápidamente en un tema de interés y de cierta controversia acerca del fenotipo exacto de la familia KE, aunque parece haberse llegado a unas conclusiones comunes.

En primer lugar, aunque la patología de la familia KE está limitada al sistema nervioso central, los miembros afectados sufren alteraciones en diferentes aspectos de la función cerebral (Gopnik y Crago, 1991; Vargha-Khadem *et al.*, 1995; Watkins *et al.*, 2002a), como son:

- Dispraxia verbal
Dificultad para controlar los movimientos bucales complejos necesarios para el habla.

- Alteraciones del lenguaje

Déficits en habilidades relacionadas con el lenguaje, que se extienden tanto al dominio expresivo (problemas en tareas de repetición de palabras) como el receptivo (déficits en vocabulario receptivo) y que afectan tanto al lenguaje escrito como al hablado. Estas alteraciones también afectan a la comprensión y producción gramatical.

Un tema que todavía genera cierta controversia dentro de la patología observada de la familia KE es si las dificultades del lenguaje receptivo y expresivo constituyen un resultado secundario de la articulación alterada del habla durante el desarrollo temprano o se deben a un déficit básico del control motor (Alcock *et al.*, 2000a; Alcock *et al.*, 2000b).

En paralelo a la caracterización del fenotipo de la familia KE se inició la tarea de localizar al gen responsable del desorden, hecho que no se conseguiría hasta 11 años después del descubrimiento de dicha familia. Mediante análisis de ligamiento con marcadores de DNA se pudo acotar un intervalo en el brazo largo del cromosoma 7 como portador del gen afectado (Fisher *et al.*, 1998). En la región genómica delimitada, se localizaban 70 genes, por lo que se empezó a explorar algunos de ellos para intentar localizar la mutación (Lai *et al.*, 2000). Sin embargo, el descubrimiento de un individuo (denominado CS), no relacionado con la familia KE, que tenía un diagnóstico similar de dispraxia verbal severa acompañada de una alteración en el lenguaje receptivo y expresivo permitió acelerar la identificación del gen. Esto fue debido a que el fenotipo de CS estaba asociado a una reorganización cromosómica, una traslocación entre los cromosomas 5 y 7, en la que el punto de rotura del cromosoma 7 se encontraba en el mismo intervalo acotado por ligamiento. Se pensó por tanto que el mismo gen era el responsable de los dos casos. Finalmente, el mapeo preciso del punto de rotura de la traslocación de CS indicó que se interrumpía un gen nuevo que no había sido totalmente caracterizado (Lai *et al.*, 2001). El análisis de los exones del gen permitió la identificación de una mutación en el exón 14 que provocaba un cambio de aminoácido arginina-histidina en todos los miembros afectados de la familia KE en heterocigosis y que se encontraba ausente en los individuos no afectados. El punto de rotura de la traslocación del individuo CS se localizó entre los exones 3b y 4 del gen.

Mediante el análisis de la secuencia aminoacídica predicha se observó que el extremo carboxi-terminal contenía un segmento de 84 aminoácidos que mostraban elevada similitud con el dominio de unión al DNA de la familia de factores de transcripción *forkhead*. Así, el gen se denominó gen *FOXP2*, según la nomenclatura estándar. Las reglas convenidas según el comité de Nomenclatura para la familia de genes *forkhead*, establecen la denominación de FOXP2 en género *Homo*, Foxp2 en *Mus* y FoxP2 en el resto de especies, en cursiva para genes y sin cursiva para proteínas (Kaestner *et al.*, 2000).

En este mismo dominio *forkhead* se localizaba la mutación encontrada en la familia KE, concretamente en una posición adyacente a una histidina que tiene contacto directo con el DNA y que permanece invariable en todos los miembros de la familia de proteínas, desde levaduras a humanos (figura I7). Esto sugería que el residuo de arginina es crucial para la función del dominio y que la sustitución observada en la familia KE altera las propiedades de unión al DNA de la proteína FOXP2.

Introducción

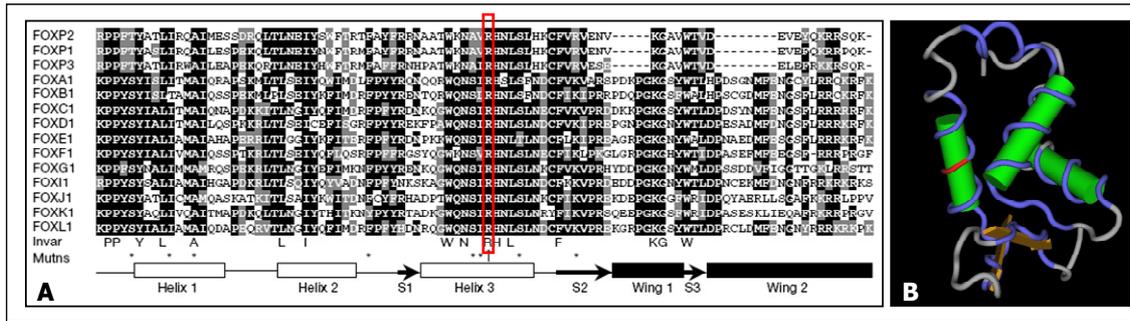


Figura 17. Dominio *forkhead* de la proteína FOXP2. A. Dominio *forkhead* de la proteína FOXP2 alineado con proteínas humanas representativas de diferentes ramas de la familia FOX. Los residuos invariables se muestran en la última fila. Los asteriscos muestran sitios de mutaciones que se han visto implicadas en enfermedades humanas. En rojo aparece resaltada la posición donde se produce el cambio de aminoácido en los miembros afectados de la familia KE. Debajo del alineamiento se muestra la estructura propuesta del dominio *forkhead* (figura extraída de Lai *et al.*, 2001 con modificaciones). B. Estructura tridimensional del dominio *forkhead*, con la posición del cambio de aminoácido de la familia KE en rojo (extraída de <http://www.well.ox.ac.uk/~simon/FOXP2/index.shtml>).

IV.2 Estructura del gen *FOXP2*

En la primera descripción del gen, Lai y colaboradores estimaron que *FOXP2* estaba constituido por 19 exones, 17 de ellos codificantes que daban lugar tras un diferente procesado alternativo de los exones 3a y 3b, a cuatro transcritos diferentes, con dos sitios de inicio de la traducción (Lai *et al.*, 2001). El producto final del gen eran tres isoformas de la proteína FOXP2. Posteriormente, Bruce y Margolis detectaron 6 nuevos exones, tres de ellos en la región no traducida y una versión más larga del exón 10 que da lugar a una proteína truncada y Schroeder y Myers describieron un exón más en el intrón 1 (Bruce y Margolis, 2002; Schroeder y Myers, 2008). Así pues, el gen *FOXP2* comprende 26 exones, 18 de ellos codificantes, distribuidos en un contexto genómico de poco más de 600 kilobases (figura I8).

Se ha estimado la existencia de al menos 4 sitios de inicio de la transcripción alternativos localizados en los exones s1, 1, 1b y 2 (Schroeder y Myers, 2008). A partir de estudios de PCR cuantitativa y ensayos funcionales del promotor estos mismos autores sugieren que los sitios de inicio de la transcripción localizados en s1 y 2 contribuyen a los niveles basales en una amplia variedad de tejidos adultos humanos. La caracterización de la región o regiones promotoras no está totalmente clara. Tomando como sitio de inicio de la transcripción el extremo 5' del exón s1, Bruce y Margolis estudiaron regiones 5' de los exones s1, s2, s3 y 1 y detectaron dos potenciales regiones promotoras cerca del primer exón, una entre -235 y +220 y la otra entre -1531 y -1208, coincidentes con una isla GC, y determinaron que la región promotora tenía que localizarse en esa área cercana al exón s1 (Bruce y Margolis, 2002).

El gen *FOXP2* se caracteriza por la presencia de diferentes dominios directamente relacionados con sus funciones:

- **Dominio *forkhead*/FOX**

La presencia del dominio *forkhead* caracteriza a todos los miembros de la familia de genes *forkhead* o FOX. Este dominio comprende tres hélices- α anfipáticas seguidas de dos largos lazos (figura I7) y se encuentra muy conservado a lo largo de la escala evolutiva, especialmente la tercera hélice- α , región por donde se une específicamente

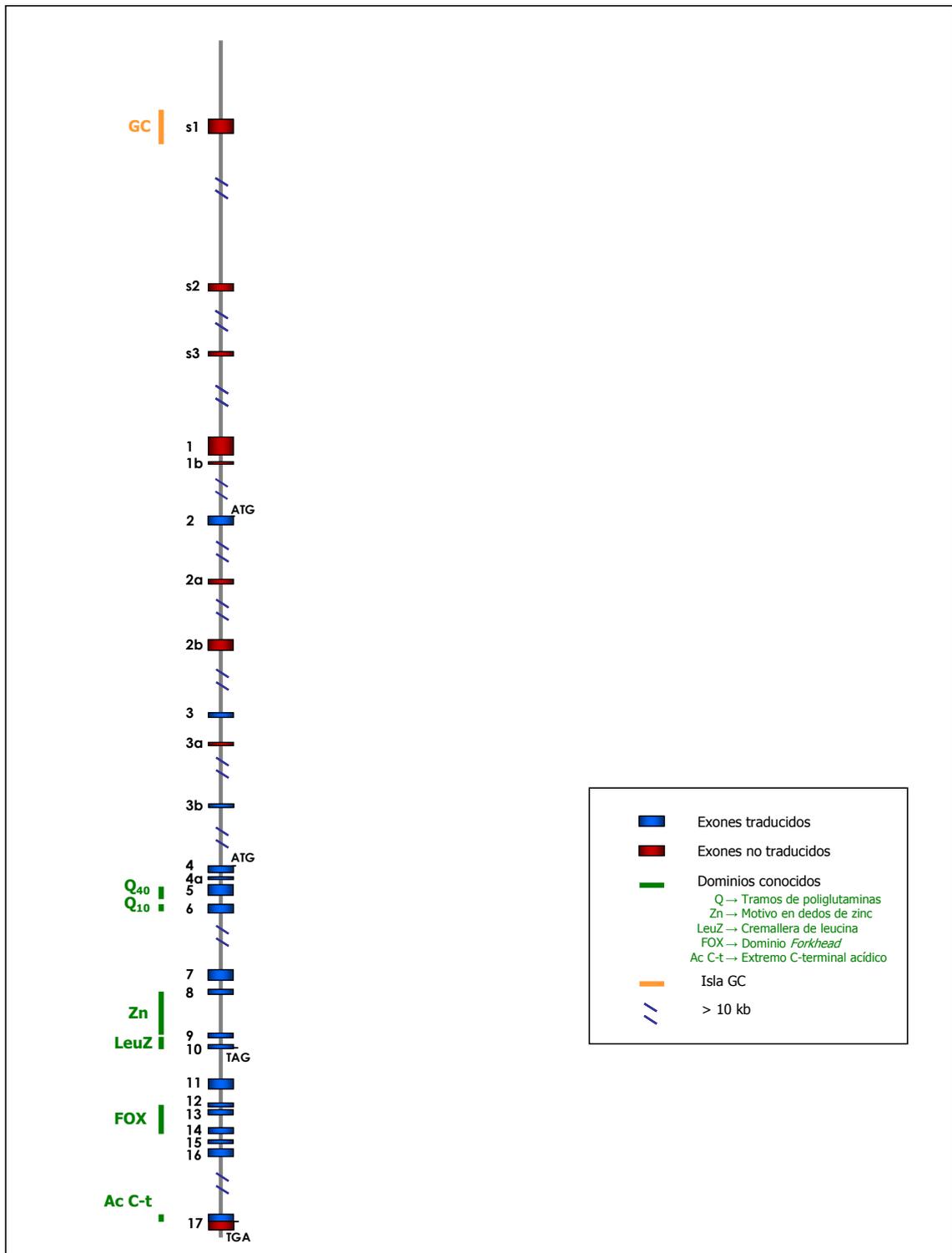


Figura I8. Estructura del gen *FOXP2*.

Introducción

el factor de transcripción al surco mayor del DNA. Cada proteína FOX es capaz de regular la transcripción de un determinado subconjunto de genes. Esta especificidad viene determinada parcialmente por la estructura del dominio y el ambiente celular en el que se encuentra (Katoh y Katoh, 2004).

Dada su función como controladores de la expresión génica, y su función en el desarrollo embrionario, no es extraño que mutaciones en diferentes genes *forkhead* estén ligadas a desórdenes del desarrollo. Por ejemplo, alteraciones en el gen *FOXC1* dan lugar a glaucoma y otros desórdenes relacionados con el ojo, mientras que la disrupción del gen *FOXP3* causa una forma severa de deficiencia inmune (revisado en Carlsson y Mahlapuu, 2002; Lehmann *et al.*, 2003). En el caso de los desórdenes de herencia dominante relacionados con genes *FOX*, donde hay una copia mutada y una copia normal del gen en cuestión, se piensa que las anomalías resultantes se deben a los niveles reducidos de proteína *FOX*. Esta idea de la importancia de la dosis de los genes *FOX* ha sido reforzada por estudios de disrupción de genes en ratón (Shu *et al.*, 2005).

▪ Tramos de poliglutaminas

El gen *FOXP2* contiene dos tramos de poliglutaminas, uno de 40 repeticiones y otro de 10 en los exones 5 y 6 respectivamente (Figura I8). Estos tramos están formados por una mezcla de tripletes CAG y CAA, no siendo repeticiones puras. El número de repeticiones no muestra variaciones en individuos normales (Lai *et al.*, 2001) y el análisis de individuos con trastornos progresivos del movimiento mostró también una elevada estabilidad en este número, puesto que únicamente dos individuos presentaban una delección en un triplete, sin detectarse expansiones (Bruce y Margolis, 2002).

Los tramos de poliglutaminas forman normalmente parte de dominios de regulación de la transcripción y son frecuentes en enfermedades neurodegenerativas (Perutz *et al.*, 1994; Margolis *et al.*, 1997). Se desconoce cual puede ser su función en el gen *FOXP2*, ya que la mayoría de genes de la familia *FOX* no contiene esta estructura. Únicamente el gen *FOXP1*, que presenta alta homología con el *FOXP2* comparte tramos de poliglutaminas, aunque presente expansiones diferentes (Shu *et al.*, 2001)

▪ Otros dominios

Los *dominios de cremallera de leucina* intervienen en procesos de formación de homodímeros y heterodímeros entre las proteínas FOXP1, FOXP2 y FOXP4, necesarios para la función de estas proteínas como reguladores transcripcionales (Li *et al.*, 2004).

El *extremo ácido* contiene una de las secuencias de localización nuclear, necesarias para la correcta localización de la proteína FOXP2 (Mizutani *et al.*, 2007).

El *dominio en dedos de zinc* actúa como dominio regulador de la transcripción (Li *et al.*, 2004).

Tabla 15. Resumen de los estudios de búsqueda de mutaciones y de asociación realizados con el gen *FOXP2*

Autores	Trastorno	Muestras	Polimorfismos estudiados	Polimorfismos descritos	Tipo de estudio	Asociación
Newbury <i>et al.</i> , 2002	Autismo	169 familias multiplex con autismo 43 familias con desórdenes del lenguaje	rs923875 rs17137124 Intrón2 (TAGA) ⁿ Intrón2 (CA) ⁿ Intrón3b (CA) ⁿ Intrón1b (CA) ⁿ	11 cambios en intrones Delección dos poliglutaminas en exón 6 (no asociado a enfermedad)	TDT	Negativa
Meaburn <i>et al.</i> , 2002	Baja puntuación en pruebas de lenguaje	270 niños	Mutación familia KE	-	Identificación mutación de la familia KE	Negativa
Wassink <i>et al.</i> , 2002	Autismo	75 familias (con par de hermanos afectados) 60 probandos autistas independientes	Mutación familia KE Intrón1(GATA) ⁿ Intrón1 (TTTA) ⁿ	Cambio sinónimo C→T en exón 5 (G247G) Delección 18 nt (6 glutaminas) en tramo 40 glutaminas del exón 5 en un probando y su madre. Delección 15 nt (5 glutaminas) en tramo 40 glutaminas del exón 5 en dos probandos y su madre.	Búsqueda de mutaciones en exones por SSCPS y secuenciación Test de ligamiento con los elementos repetitivos	-
Bruce y Margolis, 2002	Desórdenes progresivos del movimiento	142 individuos	Tramos poliglutaminas	Inserción 3 nt en dos individuos, en un alelo.	Búsqueda expansiones de tramos de poliglutaminas	-
Gauthier <i>et al.</i> , 2003	Autismo	72 tríos 98 controles	Cambio sinónimo en exón 5 (Q190Q) 1 cambio intrón 5 2 cambios intrón 13	Cambio sinónimo en exón 5 (Q190Q) 1 cambio intrón 5 2 cambios intrón 13	ETDT (Extended Transmission Disequilibrium Test)	Negativa
O'Brien <i>et al.</i> , 2003	TEL	604 niños (1608 individuos totales)	Intrón1(GATA) ⁿ Intrón1 (TTTA) ⁿ Intrón3 (TTTA) ⁿ Mutación familia KE	1 cambio +24 pb después exón 14 1 cambio +43 pb después exón 14 1 cambio -9 exón 14 Ausentes en 92 controles	Ligamiento Asociación: -Fenotipos discretos: AFBAC, ETDT -Puntuaciones continuas del lenguaje: Q/TDT	Negativa
Kaminen <i>et al.</i> , 2003	Dislexia	6 individuos afectados	-	-	Búsqueda de mutaciones en región codificante	-
Gong <i>et al.</i> , 2004	Autismo	181 tríos	rs1852469 rs2396753 rs1456031	-	Asociación: TDT Haplotipos	Asociación para rs1456031 (transmisión preferente alelo C) Asociación haplotipos rs1456031C con rs2396753C y rs1852469A

Autores	Trastorno	Muestras	Polimorfismos estudiados	Polimorfismos descritos	Tipo de estudio	Asociación
Sanjuán <i>et al.</i> , 2005	Esquizofrenia	149 pacientes 137 controles	Mutación familia KE rs923875 rs17137124		Asociación caso-control	Negativa
Li <i>et al.</i> , 2005	Autismo	53 pacientes 50 controles	-	1 Delección CAA en exón 5 (4 pacientes y 2 controles) Cambio sinónimo en exón 5 A569G 3 cambios en intrones En región codificante: -Cambio no sinónimo en exón 2 Q17L (ausente en 366 cromosomas control) -Inserción de 12 nt en tramo de poliglutaminas (ausente en 228 cromosomas control) -Codón de parada en exón 7 R328X (ausente en 252 cromosomas control). Segrega con alteraciones del lenguaje en una familia.	Rastreo de mutaciones en región codificante.	Frecuencia de TT con GT en intrón15 mayor en autistas.
MacDermot <i>et al.</i> , 2005	Dispraxia verbal	49 probandos	-			-
Marui <i>et al.</i> , 2005	Autismo	170 pacientes 214 controles	rs2106900 rs2061183 rs1456029 rs1005958 rs1058335	-	Asociación	Negativa
Richier <i>et al.</i> , 2006	Autismo	146 Pacientes 124 Controles	Secuenciadas regiones ultraconservadas en 7q	Encontradas variantes raras en pacientes y controles	Secuenciación regiones ultraconservadas	-
Sanjuán <i>et al.</i> , 2006	Esquizofrenia	186 pacientes con AA 160 controles	rs7803667 rs10447760 rs923875 rs1597548 rs2396722 rs1358278 rs1852469 rs2396753 rs17137124 rs1456031	-	Asociación	Positiva para rs2396753 (frecuencias alélicas P=0.02; frecuencias genotípicas P=0.07) Haplotipos: Rs7803667/rs10447760C /rs923875A/ Rs1358278A/rs2396753A

Tabla 15 (continuación) . Resumen de los estudios de búsqueda de mutaciones y de asociación realizados con el gen FOXP2

IV.3 Variabilidad en el gen *FOXP2*

Al convertirse el gen *FOXP2* en el primer gen relacionado directamente con una alteración del lenguaje, no tardaron en surgir diferentes estudios encaminados a la búsqueda de mutaciones o estudios de asociación con diferentes patologías relacionadas con el lenguaje. El hecho de que se hubiera encontrado ligamiento en la región 7q31, donde se encuentra el gen *FOXP2* para diferentes desórdenes del lenguaje, como el TEL, autismo o dislexia prometía resultados interesantes (Ashley-Koch *et al.*, 1999; Warburton *et al.*, 2000; Folstein y Mankoski, 2000; Alarcón *et al.*, 2002; Vorstman *et al.*, 2006).

En cuanto a la **búsqueda de mutaciones**, hasta el momento únicamente se han descrito 3 mutaciones relacionadas con trastornos del lenguaje, independientemente de alguna modificación del tramo de poliglutaminas. La primera de ellas es la mutación R553H en el exón 14 ya descrita, identificada en la familia KE (Lai *et al.*, 2001). Posteriormente, en un estudio de búsqueda de mutaciones en individuos con dispraxia verbal del desarrollo se identificaron dos mutaciones más: un cambio no sinónimo en el exón 2 (Q17L) que se encuentra cerca del extremo N-terminal, en una región de función no determinada, y un codón de parada en el exón 7 (R328X), que segregaba con la alteración del lenguaje en una de las familias estudiadas y da lugar a una proteína truncada y (MacDermot *et al.*, 2005).

Se ha realizado el análisis de los efectos de las mutaciones de *FOXP2* a nivel funcional utilizando diferentes aproximaciones: localización celular, unión al DNA y propiedades de transactivación, todo ello en líneas celulares (Vernes *et al.*, 2006). En concreto, se analizaron los tres cambios codificantes que habían sido identificados en alteraciones del lenguaje. También se exploraron las características de una isoforma que da lugar a la variante truncada *FOXP2-10+*, que carece de extremo C-terminal.

Los resultados obtenidos mostraron que en el caso de la variante encontrada en la familia KE, la proteína mutante obtenida no se localiza exclusivamente en el núcleo, como es el caso de la proteína normal, sino que se observa su presencia también en el citoplasma. Esta proteína mutante además, pierde la capacidad de unirse al DNA, así como la actividad de transactivación (reflejada como capacidad de reprimir el promotor de SV40). Cuando esta variante se expresa en heterocigosis con la proteína *FOXP2* endógena normal, como ocurre en los miembros afectados de la familia KE, es posible que la dimerización necesaria para la correcta actividad de la proteína ocurra entre proteínas mutantes y normales, dejando abierta la cuestión de si los efectos observados en la familia se deben a la pérdida de función de la proteína o al hecho añadido de que la proteína mutante interfiera con la actividad de la normal (Vernes *et al.*, 2006).

En cuanto a la variante que da lugar a un codón de parada en posición 328, los efectos que produce parecen ser debidos a un problema de dosis, puesto que además de perder capacidad de unión al DNA, así como la de transactivación, se localiza en el citoplasma, y da lugar a niveles de expresión más bajos. La variante Q17L no mostró ninguna diferencia con la proteína normal, por lo que probablemente no se trata de ninguna mutación patológica. Los estudios de localización de la proteína truncada *FOXP2-10+*, indicaron su presencia en el citoplasma, probablemente debido a la ausencia del dominio FOX, en forma de agregados que colocalizaron con marcadores de "agresomas" como la ubiquitina. Esto, unido a que esta isoforma no pierde la actividad transactivadora permitió a los autores sugerir un papel en la regulación postraduccional de isoformas de *FOXP2* mayores, así como señalar la importancia que

Introducción

pueden tener las diferentes formas de procesamiento del RNA mensajero del gen en la regulación (Vernes *et al.*, 2006).

Además de la búsqueda de mutaciones, se han realizado diferentes **estudios de asociación** con patologías como el TEL, autismo o alucinaciones auditivas. En la tabla 15 se muestra un resumen de la búsqueda de mutaciones y estudios de asociación publicados hasta la fecha.

IV.4 Genes diana de FOXP2

Los primeros datos sobre los genes que podrían interactuar con *FOXP2* se obtuvieron en pulmón de ratón, donde el gen *Foxp2* juega un papel importante en la diferenciación epitelial. En este tejido se estimó que los genes ortólogos *Foxp2* y *Foxp1*, reprimen la expresión de los genes *CC10* y *T1 alpha* de ratón y *SP-C* humano (Shu *et al.*, 2001; Shu *et al.*, 2007). La inhibición de la transcripción de *SP-C* se ha visto confirmada y ocurre a través de interacción con el homeodominio de Nkx2.1 (Zhou *et al.*, 2008)

Sin embargo, aunque los estudios combinados sobre el fenotipo de la familia KE, neuroimagen, expresión y funcionales del gen sugerían que una dosis reducida de la proteína FOXP2 tiene un impacto en el desarrollo y función de determinados circuitos neurales, incluyendo algunos importantes para la adquisición del habla y del lenguaje, hasta hace poco se desconocía ninguna diana específica de tejido cerebral.

En dos estudios paralelos Spiteri, Vernes y colaboradores (Spiteri *et al.*, 2007; Vernes *et al.*, 2007) identificaron las dianas de la proteína FOXP2 en tejido cerebral fetal embrionario procedente de los ganglios basales y de la porción inferior de la corteza frontal, en pulmón en desarrollo y en células neurales.

En tejido cerebral embrionario se encontraron 285 dianas de las cuales 110 solapaban con dianas encontradas en tejido pulmonar. Respecto a la función de los genes diana identificados, al igual que en el caso de genes diana de otros genes *forkhead*, se encontró relación con funciones como la regulación del crecimiento, desarrollo embrionario y transducción de señales. Además se encontraron diferencias entre las diferentes regiones cerebrales analizadas. En la corteza frontal inferior se observaron genes diana involucrados en el crecimiento y morfogénesis, mientras que en los ganglios basales, principalmente genes relacionados con el desarrollo (Spiteri *et al.*, 2007). El análisis *in silico* de las interacciones directas e indirectas entre las dianas de FOXP2, mostró algunas redes interesantes como la asociada con las rutas de señalización Wnt/ β -catenina y IGF-1. Genes en esta red han sido implicados en la diferenciación de neuronas y neuroglía, muerte celular neuronal, adhesión celular y control de la transcripción (Vernes *et al.*, 2007).

Dada la complejidad genética de un rasgo tan ventajoso como el lenguaje, y puesto que el gen *FOXP2* había sufrido selección positiva, se planteó que ésta actuara también sobre las dianas de la proteína. Se encontró que 14 de las dianas identificadas mostraron evolución acelerada y posiblemente selección positiva. Estas dianas incluyen genes implicados en la formación del patrón del prosencéfalo en vertebrados. Además 47 de los genes diana identificados, que incluyen genes relacionados con la formación del patrón del sistema nervioso central, moléculas guía y transmisión neural y factores

de transcripción involucrados en el desarrollo neural, se expresan de forma diferente entre cerebros humanos y de chimpancé. Estos datos sugieren el papel de genes específicos en la cognición humana, genes con diferente expresión en primates y potencialmente bajo selección positiva en humanos (Spiteri *et al.*, 2007).

Otro dato muy interesante es el hecho de que 15 de las dianas identificadas (*APOD*, *CCK*, *CCK-AR*, *CCND2*, *CD5*, *DISC1*, *DRD2*, *GABBR1*, *MT2A*, *NOS1*, *PMX2B*, *TDO2*, *TIMELESS*, *WNT1*, *ZNF74*) han mostrado alguna evidencia de asociación con esquizofrenia.

En la mayoría de genes diana encontrados se identificaron sitios consenso de unión al DNA basados en la estructura del dominio FOX de FOXP2. También se encontró una sobrerrepresentación de sitios de unión a factores de transcripción en las secuencias diana, que incluyen sitios de unión a factores de transcripción relacionados en procesos del desarrollo, señalización Wnt y plasticidad, así como genes implicados en enfermedades como el Alzheimer. La concurrencia de estos sitios subraya el papel potencial de FOXP2 en rutas clave del desarrollo y función neural. La relación de FOXP2 con la cascada de señalización Wnt se ve reforzada por la identificación de Lef1, factor de transcripción activado por dicha ruta, como regulador de la expresión de *FoxP2* en sistema nervioso central del pez cebra (Bonkowsky *et al.*, 2008).

Aunque inicialmente la actividad de FOXP2 parecía concentrarse en su actuación como represor transcripcional, en los estudios de Vernes y Spiteri (Vernes *et al.*, 2007; Spiteri *et al.*, 2007) se encontraron los primeros ejemplos de activación transcripcional.

Recientemente, se ha descrito que FOXP2 regula la expresión del gen *CNTNAP2*, que codifica para una proteína de la familia de las neurexinas, moléculas de adhesión y receptores del sistema nervioso de vertebrados (Vernes *et al.*, 2008). Este resultado es especialmente relevante, puesto que además de encontrarse asociación de polimorfismos de este gen con el ítem de repetición de palabras sin sentido en niños con SLI y constituir un gen candidato a autismo (Vernes *et al.*, 2008; Arking *et al.*, 2008; Alarcón *et al.*, 2008), se ha sugerido que la haploinsuficiencia de esta proteína esté relacionada con esquizofrenia y epilepsia (Friedman *et al.*, 2008).

IV.5 Expresión del gen FOXP2

Los primeros datos de expresión del gen estaban basados en la identificación del RNA mensajero por hibridación *in situ* o *Northern blot* y detección de la proteína mediante técnicas de inmunohistoquímica. Como se menciona previamente, según la nomenclatura aceptada, *FOXP2* corresponde al gen humano, *Foxp2* al gen de ratón y la denominación *FoxP2* se utiliza para el resto de especies.

IV.5.1 Expresión en mamíferos

El gen *FOXP2* se expresa no sólo en el cerebro, sino en otros órganos, incluyendo pulmón, corazón, hígado e intestino (Lai *et al.*, 2001; Shu *et al.*, 2001; Lu *et al.*, 2002).

Introducción

Su función en los **pulmones** durante el desarrollo embrionario inhibe la expresión de genes asociados con la diferenciación de las células del epitelio pulmonar, teniendo un importante papel en la especificación y diferenciación del tejido epitelial del pulmón (Shu *et al.*, 2001; Shu *et al.*, 2007).

A nivel del **cerebro**, la expresión de *Foxp2* ha sido descrita en rata y ratón en diferentes estados del desarrollo desde embrión a adulto (Shu *et al.*, 2001; Lu *et al.*, 2002; Ferland *et al.*, 2003; Lai *et al.*, 2003; Takahashi *et al.*, 2003), así como la de *FOXP2* en embriones y tejidos fetales humanos entre 6 y 22 semanas (Lai *et al.*, 2003; Teramitsu *et al.*, 2004).

Los patrones de expresión cerebral de ratón y ser humano muestran similitudes en estadios comparables. En concreto, en tejido cerebral, se expresa en:

- *Núcleos sensoriales*

El gen *Foxp2/FOXP2* se expresa en bulbo olfatorio del ratón adulto, colículos superior e inferior y los cuerpos geniculados lateral y medial (estructuras visuales y auditivas del cerebro medio y tálamo) de ratón adulto y feto humano y núcleos talámicos somatosensoriales en embriones de ratón y humano.

- *Núcleos límbicos*

Se encuentra expresión en amígdala y núcleos paraventriculares del hipotálamo y tálamo en roedores y en los núcleos talámicos dorsales mediales en feto humano. En roedores se ha observado también expresión en el compartimento estriosomal, relacionado con funciones límbicas.

- *Córtex cerebral*

Durante el desarrollo embrionario *Foxp2* se expresa en la capa interior de la placa cortical de roedores y primates, con tendencia a mayor expresión en los laterales. Posteriormente, la expresión queda limitada a la capa VI infragranular y en menor extensión a la capa V.

- *Estructuras motoras*

Foxp2 se expresa en diferentes estructuras de los ganglios basales, como el núcleo caudado y putamen de todos los mamíferos en todas las edades investigadas, además del núcleo acumbens en rata, globo pálido y núcleos subtalámicos en feto humano, y sustancia nigra en ratón a todas las edades. En cuanto a estructuras talámicas, se expresa en diferentes subdivisiones relacionadas con los ganglios basales. En romboencéfalo, *Foxp2* se expresa en el cerebelo, concretamente en las células de Purkinje y en el complejo olivar de todas las especies analizadas, así como en otras estructuras cerebelares.

La expresión de *Foxp1* y *Foxp4*, los genes más cercanos a *Foxp2* solapa parcialmente con la de éste, lo que unido al hecho de que las tres proteínas forman heterodímeros a través de los motivos de cremallera de leucina hace pensar que sus funciones represoras puedan depender de funciones sinérgicas y que el equilibrio de los diferentes genes *FOXP* es crucial para el correcto desarrollo cerebral (Lu *et al.*, 2002; Li *et al.*, 2004).

IV.5.2 Expresión en no mamíferos

Expresión en aves con aprendizaje vocal

Las aves cantoras, al igual que los seres humanos aprenden las vocalizaciones a través de la imitación, por lo que rápidamente se planteó si la especie humana y estas aves comparten mecanismos moleculares de aprendizaje vocal. En 2004, Teramitsu y colaboradores analizaron la expresión del gen en cerebro de pinzón cebra y observaron, en primer lugar, que la expresión de FoxP2 en el cerebro del pinzón cebra es muy similar a la observada en mamíferos, y en segundo lugar que FoxP2 y FoxP1 tienen patrones distintos y compartidos en el córtex, estriado y tálamo. Además, puesto que en el pinzón cebra, únicamente el macho aprende a cantar, se analizó la expresión en ambos sexos y se observó que aunque el patrón de expresión era similar para FoxP2, la expresión de FoxP1 mostraba dimorfismos sexuales. Así, los datos de expresión de ambos genes llevaron a estos autores a proponer que en las regiones en las que solapan, ambos genes podrían actuar como correguladores debido a su capacidad para formar homo y heterodímeros (Li *et al.*, 2004).

En un estudio paralelo Haesler y colaboradores (2004) compararon la distribución del RNA mensajero y de la proteína de FoxP2 en cerebros adultos y en desarrollo de una serie de aves capaces de aprendizaje vocal, así como otras no capaces de aprender. Al tratar de determinar si *FoxP2* se expresa de forma diferente en los cerebros de aves aprendices vocales y aves no aprendices, no se obtuvieron datos concluyentes, aunque sí muy interesantes. Por un lado, se observó que *Foxp2* se expresa de forma diferente en lo que se conoce como área X del núcleo vocal, zona necesaria para el aprendizaje vocal, y no presente en aves no aprendices vocales. Además, la expresión del gen aumenta en el área X cuando el pinzón cebra joven aprende a imitar el canto y cuando los canarios adultos remodelan sus trinos. Lesiones en el área X en pinzones cebra durante el aprendizaje vocal dan lugar a que los cantos de adultos sean más plásticos que cuando esta área se encuentra intacta. Los elevados niveles de FoxP2 encontrados en los períodos de plasticidad sugerían que FoxP2 podría reprimir genes involucrados en la estabilidad neural del área X, lo que es compatible con un papel en la vocalización aprendida. *FoxP2* también se expresa en regiones del estriado no vocal en todas las especies analizadas, y su expresión es similar, independientemente de su capacidad para el aprendizaje de la vocalización.

Los datos obtenidos por los dos grupos de Haesler y Teramitsu, primeros en estudiar *FoxP2* en aves, apuntaron hacia un papel del gen más amplio del que en principio se sugería, al proponer que *FoxP2* se expresa principalmente en rutas sensoriales y en neuronas de proyección del estriado, que son el lugar de convergencia de proyecciones dopaminérgicas de la sustancia nigra y glutamatérgicas de la corteza. Además, en los embriones de aves, las regiones de expresión temprana del gen son fuentes de señales inductivas que organizan el neuroepitelio adyacente y la migración neuronal durante el desarrollo temprano, por lo que razonablemente se planteó que *FoxP2* estuviera relacionado con el desarrollo, mantenimiento y función de los circuitos sensori-motores y sensoriales del estriado y mesencéfalo. Esto podría crear un ambiente permisivo en el que el aprendizaje vocal pudiera evolucionar si otros factores entraran en juego.

Introducción

Expresión en pez cebra

En el sistema nervioso del pez cebra, *FoxP2* se expresa en un patrón temporal y espacial que incluye el telencéfalo, diencéfalo, cerebelo, mesencéfalo, médula espinal y células de los ganglios retinales, expresión similar a la que ocupa en ratón y en el hombre. También se expresa en corazón. Sin embargo, a diferencia de ratón o pinzón cebra no se detectó en otros órganos como intestino o riñón (Bonkowsky y Chien, 2005).

IV.5.3 Animales *knockout*

Ratones *knockout* de *Foxp2*

Ratones *knockout*, con dos copias alteradas del gen *Foxp2*, viven durante tres semanas desde el nacimiento (Shu *et al.*, 2005). Además de la menor longevidad, muestran retraso del desarrollo y alteraciones en test de función motora. La ejecución de estas tareas por ratones heterocigotos es sólo moderadamente peor que la de los ratones normales y alcanzan la normalidad durante la segunda semana de vida. Los estudios de vocalización mostraron que la frecuencia de las vocalizaciones ultrasónicas se encuentra selectivamente alterada en las crías *knockout* y en los heterocigotos, aunque las estructuras anatómicas necesarias para ello eran aparentemente normales. A nivel estructural, los ratones *knockout* de *Foxp2* muestran anomalías cerebelares, que incluyen glía de Bergmann anormal y resolución postnatal incompleta de la capa granular externa, lo que sugiere alteraciones en la migración celular. En animales heterocigotos la capa molecular es más fina y las células de Purkinje se encuentran desalineadas y tienen ramificaciones dendríticas poco desarrolladas. Se ha planteado que el cerebelo puede ser especialmente vulnerable a la ausencia del gen *Foxp2* durante el desarrollo debido a que en este tejido *Foxp1* no coexpresa con el gen y no puede compensar por su falta de actividad (Scharff y Haesler, 2005).

Pinzón cebra *knockout* de *FoxP2*

La disminución de expresión de *FoxP2* en el área X del pinzón cebra, un área importante para el desarrollo adecuado del canto, da lugar a que los individuos jóvenes sean incapaces de imitar de forma apropiada y completa los trinos del tutor (Haesler *et al.*, 2007). La estructura acústica y duración de las sílabas del canto en los adultos son anormalmente variables, de forma similar a la producción de palabras que se observa en individuos con dispraxia verbal del desarrollo, lo que es consistente con un papel del gen en los ganglios basales de estas aves durante el aprendizaje vocal motor guiado por la audición.

IV.6 Estudios de neuroimagen de la familia KE

Los primeros análisis de neuroimagen (Vargha-Khadem *et al.*, 1998) mostraron anomalías estructurales bilaterales en diferentes regiones motoras en los miembros afectados de la familia KE respecto de los no afectados. Estas anomalías incluían el núcleo caudado, el cual también presentaba anomalías bilaterales en el estudio con tomografía de emisión de positrones (PET). Un análisis volumétrico más

detallado de esta estructura confirmó que ambos núcleos caudados se encuentran reducidos en volumen en los miembros afectados de la familia KE, respecto de miembros no afectados e individuos control (Watkins *et al.*, 2002). Además, el volumen de los núcleos caudados estaba correlacionado significativamente con las puntuaciones de los miembros afectados en los tests de práctica oral, repetición de palabras sin sentido, y un subtest de la escala de inteligencia Wechsler.

Posteriores análisis (Belton *et al.*, 2003) mostraron niveles anormalmente bajos de materia gris en la circunvolución frontal inferior (área de Broca), circunvolución precentral, polo temporal, cabeza del núcleo caudado y cerebelo ventral; así como niveles anormalmente altos de materia gris en la porción posterior de la circunvolución temporal superior (área de Wernicke), circunvolución angular y el putamen.

Los principales estudios de neuroimagen funcional llevados a cabo con la familia KE utilizaron dos aproximaciones, una de generación verbal silenciosa y otra verbal abierta que incorporaba generación verbal y repetición de palabras (Liégeois *et al.*, 2003). Los resultados de la primera de ellas mostraron: menor activación del área de Broca y su homóloga del hemisferio derecho, así como del putamen, junto con activación anormalmente baja de otras regiones relacionadas con el habla; mayor activación en áreas no relacionadas normalmente con el lenguaje, lo que podía significar el reclutamiento de circuitos compensatorios o simplemente un esfuerzo cognitivo añadido. Los resultados de la prueba verbal abierta mostraban que mientras que los miembros no afectados de la familia KE presentaban en las tareas de generación verbal un patrón de activación con dominancia izquierda típica, incluyendo el área de Broca y en las tareas de repetición una distribución más bilateral, los miembros afectados presentaban un patrón más bilateral y posterior en todas las tareas.

Los resultados de neuroimagen estructural y funcional, mostraron anomalías en áreas en las que se expresa el gen *FOXP2*, proporcionando evidencias directas de una relación entre las alteraciones cerebrales y el fenotipo de comportamiento de la familia KE, confirmando así un papel del gen en diferentes aspectos de la formación del lenguaje (Vargha-Khadem *et al.*, 1998; Watkins *et al.*, 2002a; Watkins *et al.*, 2002b; Belton *et al.*, 2003; Liégeois *et al.*, 2003)

IV.7 Función de la proteína FOXP2

En resumen, el marco de actuación de la proteína FOXP2 se presenta en la actualidad más amplio del que inicialmente se pensó como regulador transcripcional del desarrollo.

En pulmón, funciona como regulador de la diferenciación de células epiteliales. Sin embargo, en cerebro su función específica no está tan clara, aunque los datos recogidos hasta el momento han permitido proponer un circuito neural que permite a la corteza motora ser modulada y controlada por otras áreas corticales frontales, directamente, a través de rutas cortico-corticales, e indirectamente a través de dos rutas cortico-subcorticales paralelas, una de ellas frontoestriatal y la otra frontocerebelar. Dentro de estas rutas, los últimos estudios de identificación de genes diana de la proteína indican que FOXP2 podría participar en procesos biológicos relacionados con comunicación celular, transducción de señales, cascadas de señalización intracelular, respuesta a estímulos externos y morfogénesis principalmente (Spiteri *et al.*, 2007).

V. Selección positiva en el genoma humano

La búsqueda de los genes implicados en la determinación de los rasgos fenotípicos característicos de la especie humana constituye actualmente uno de los principales retos de la genética humana. Una de las aproximaciones utilizadas en esta búsqueda es examinar las regiones del genoma que han sido sujetas a selección positiva. En este sentido, la identificación de genes y regiones reguladoras que evolucionan adaptativamente puede ayudar a elucidar regiones específicas del genoma que han evolucionado durante los últimos seis millones de años para permitir a la especie humana desarrollar el lenguaje, la utilización de herramientas y en última instancia poblar todos los continentes (Kelley y Swanson, 2008).

La identificación de genes sometidos a selección positiva también puede tener implicaciones importantes en el campo de la medicina. Muchas enfermedades como el SIDA o Alzheimer afectan únicamente al hombre y no a otros animales (Olson y Varki, 2003). Del mismo modo, muchas estrategias terapéuticas desarrolladas en animales modelo fallan al pasar a la especie humana. Estas diferencias entre humanos y otras especies se deben en última instancia a que los programas moleculares han divergido entre linajes. Así, genes que han experimentado selección positiva pueden ser responsables de algunas de estas diferencias y algunas variantes pueden incluso participar directamente en la patogénesis de enfermedades (Vallender y Lahn, 2004). De este modo, sería posible identificar potenciales factores genéticos de enfermedad mediante la identificación de regiones del genoma humano que se encuentran en la actualidad bajo selección.

V.1 Detección de la selección positiva

Existen dos formas principales de detectar la huella de la selección positiva: los datos de divergencia permiten identificar selección positiva entre especies, mientras que datos de polimorfismos permiten identificarla dentro de una misma especie. Cada uno de los métodos detecta la selección en una escala de tiempo diferente. Los datos de divergencia son útiles para identificar eventos selectivos antiguos, mientras que los polimorfismos permiten identificar sucesos más recientes. Del mismo modo, se pueden analizar dos tipos de regiones, codificantes y no codificantes que puedan estar relacionadas con la regulación (Kelley y Swanson, 2008).

En regiones codificantes, una señal clara de selección positiva es un exceso en el número de sustituciones no sinónimas (que suponen un cambio de aminoácido) por sitio no sinónimo (d_N) comparado con el número de sustituciones sinónimas, por sitio sinónimo (d_S). Así un elevado valor de d_N/d_S es indicativo de selección positiva o falta de constricciones funcionales. Una limitación de estos test es que no permiten detectar selección sobre secuencias reguladoras, las cuales también juegan un papel importante en la evolución (King y Wilson, 1975; Prabhakar *et al.*, 2006).

La detección de selección positiva reciente necesita de la identificación de regiones que muestran evidencias de un arrastre selectivo. El arrastre selectivo da lugar aun aumento en la frecuencia de un alelo beneficioso que confiere ventajas selectivas a su portador (figura I9). Los alelos neutros ligados al alelo beneficioso aumentan también de frecuencia, por el efecto conocido como *hitchhiking*. Este efecto da lugar a un

aumento de la frecuencia del haplotipo en el que se encuentra el alelo seleccionado. Si en una región se ha dado un barrido selectivo, a medida que se acumulen nuevas mutaciones habrá un exceso de alelos raros en comparación con regiones neutras no ligadas a sitios seleccionados.

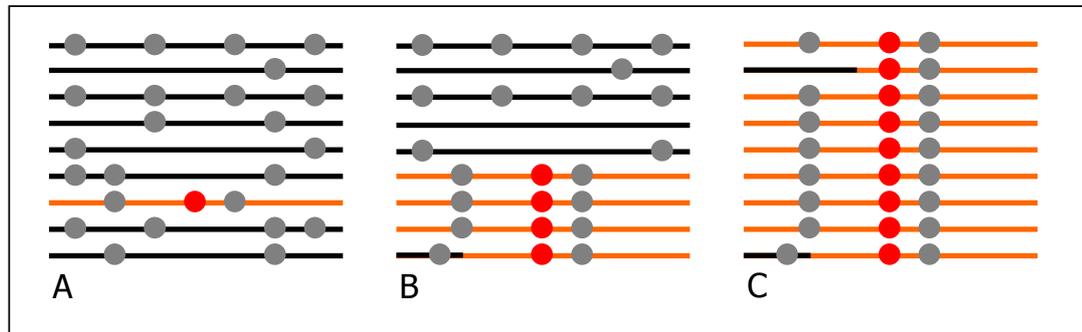


Figura I9. Arrastre selectivo con recombinación. A. La figura muestra 9 regiones cromosómicas con alelos neutros (círculos grises). El cromosoma de color naranja es el cromosoma con el alelo adaptativo (círculo rojo). B. Región cromosómica durante el arrastre selectivo. C. Resultado después del arrastre completo (Tomado de Kelley y Swanson, 2008 con modificaciones).

V.2 Selección positiva en regiones codificantes

Diferentes datos sugieren que aproximadamente un 10% del genoma humano se encuentra bajo selección positiva (Bustamante *et al.*, 2005; Williamson *et al.*, 2007), constituyendo un 4% de los genes estudiados. Estos genes incluyen los relacionados con el sistema inmune, señalización celular, metabolismo de aminoácidos y olfacción (Sabeti *et al.*, 2006; Nielsen *et al.*, 2005). Un dato interesante es el hecho de que diferentes genes involucrados en audición muestran signos de cambios acelerados en el linaje humano (Clark *et al.*, 2003) lo que permite especular sobre su posible participación en el lenguaje hablado. Sin embargo, a pesar de tener generalmente una mayor tasa de sustitución no sinónima, el número de genes seleccionados positivamente parece sustancialmente menor en humanos que en chimpancé (Bakewell *et al.*, 2007).

El estudio evolutivo de genes que se expresan en el sistema nervioso tiene un interés especial al tener en cuenta que la aparición de gran parte de las características definitorias de la especie como el lenguaje o el aprendizaje social está directamente relacionada con este sistema. Dorus y colaboradores (2004) plantearon la sugestiva hipótesis de que la intensa evolución cerebral fenotípica observada en primates debía estar correlacionada a nivel molecular y por tanto genes relacionados en biología del sistema nervioso presentarían más cambios moleculares evolutivos que mamíferos no primates.

En el contexto de esta hipótesis, compararon las tasas evolutivas de un conjunto de genes que se expresan en el sistema nervioso entre los linajes macaco-humano con las del linaje de los roedores rata-ratón (Dorus *et al.*, 2004). Encontraron que genes relacionados con el sistema nervioso, y especialmente aquellos relacionados con su desarrollo, mostraban mayores tasas de evolución protéica en primates respecto a roedores. Dentro de primates, la aceleración era más prominente en el linaje humano (Dorus *et al.*, 2004). Estos resultados sin embargo, han sido cuestionados por Shi y colaboradores, quienes no encuentran evidencias de dicha aceleración en su análisis genómico y sugieren que la selección de genes de Dorus era poco representativa (Shi *et al.*, 2006).

V.3 Evolución en regiones reguladoras

El pequeño número de diferencias aminoacídicas entre el ser humano y chimpancé llevó a plantear que la diversidad fenotípica observada sea debida a cambios en la regulación (King y Wilson, 1975). Existen diferencias de expresión entre poblaciones que pueden conferir diferentes ventajas adaptativas, y ser seleccionadas positivamente. Los niveles de transcripción y expresión de un gen pueden verse afectados tanto por la variación en el número de copias como por sustituciones en regiones no codificantes. Por tanto, la selección positiva puede potencialmente actuar en ambas.

La utilización de búsquedas genómicas basadas en polimorfismos puede permitir la identificación de polimorfismos localizados en regiones no codificantes que han conferido ventajas selectivas en el linaje humano. Dentro de estos estudios, Prabhakar y colaboradores encontraron un exceso de sustituciones específicas en el hombre en secuencias no codificantes conservadas cerca de genes que se expresan en cerebro (Prabhakar *et al.*, 2006). Este tipo de secuencias no codificantes conservadas están enriquecidas en las regiones reguladoras, lo que sugiere que la selección positiva pueda actuar sobre regiones reguladoras de genes neurales.

Comparando regiones que contenían secuencias reguladoras en *cis*, Haygood y colaboradores encontraron que la regulación de genes que se expresan en sistema nervioso y relacionados con nutrición, esenciales para la diferenciación de seres humanos respecto al resto de primates ha estado sometida a selección positiva en el linaje humano (Haygood *et al.*, 2007).

Un método alternativo para estudiar la evolución de la regulación génica es identificar elementos reguladores específicos. Mediante el método de "*phylogenetic shadowing*" o ensombrecimiento filogenético, se comparan diferentes genomas y se buscan regiones con elevado grado de conservación (Boffelli *et al.*, 2003). Se asume que estas regiones han sufrido selección purificadora, indicando que son funcionalmente importantes. Comparando los genomas de ratón, rata y chimpancé, Pollard y colaboradores encontraron 35000 regiones conservadas que habían cambiado muy poco en los 80 millones de años desde que primates y roedores compartieron un ancestro común (Pollard *et al.*, 2006). Entonces examinaron estas regiones en el genoma humano, buscando aquellos casos en los que estas regiones habían cambiado significativamente, esto es, había tenido lugar selección positiva en el linaje humano. Encontraron 49 regiones aceleradas en humanos (HAR de *human accelerated regions*), de las cuales únicamente 2 codificaban para proteínas, correspondiendo el resto a regiones reguladoras.

V.4 Estudios comparativos de expresión

Al igual que en el caso de estudios evolutivos de comparación de secuencias entre especies o linajes, dentro de los estudios en los que se compara la expresión génica en tejidos humanos respecto a otras especies, la expresión génica en cerebro humano ha atraído especial interés debido al hecho de que gran parte de las diferencias con otros organismos tienen un componente cognitivo.

Contrariamente a lo que intuitivamente se esperaría, puesto que la función del cerebro ha cambiado más entre humanos y simios que otros tejidos, cuando se compara la expresión génica de humanos y chimpancés, se encuentran menos diferencias en cerebro que en otros órganos como hígado, riñón, corazón o testículo (Khaitovich *et al.*, 2004). Así mismo, la expresión génica en regiones de la corteza cerebral que están involucradas en funciones específicas de humanos, como la producción del habla, no se diferencian más entre las dos especies que otras regiones en las que no ha habido un cambio de función tan drástico, como es la corteza visual primaria. Estos resultados podrían explicarse si la principal fuerza en determinar un cambio de expresión fuera la selección negativa, como predice la teoría neutral. En este sentido, los genes que se expresan en cerebro podrían estar involucrados en más interacciones funcionales que genes que se expresan en otros órganos, por lo que estarían sometidos a una mayor restricción funcional. De hecho, en levadura, genes involucrados en mayor número de interacciones evolucionan más lentamente que los que participan en menor número (Fraser *et al.*, 2002). Se ha observado que en cuantos más tejidos se expresa un gen, menos cambia entre especies y los cambios de expresión tienden a ser similares en los distintos tejidos.

Se han encontrado evidencias indicativas de que la selección positiva podría haber afectado la expresión génica en cerebro durante la evolución humana. Cuando se utiliza un *outgroup* para asignar los cambios en expresión génica desde el ancestro común de los linajes humano o de chimpancé, la razón del linaje humano respecto del de chimpancé es mayor para el cerebro que para hígado o corazón (Enard *et al.*, 2002a; Caceres *et al.*, 2003). Así, de los pocos cambios de expresión observados en cerebro, una mayoría parece haber ocurrido en el linaje humano respecto del de chimpancé. Este efecto podría ser explicado por aproximadamente un 10% de todos los genes. Además, hay alguna evidencia de que los genes que se expresan en cerebro podrían haber cambiado más en su secuencia aminoácida en el linaje humano que en el de chimpancé (Dorus *et al.*, 2005). Si se produce una aceleración en la evolución de la expresión génica en cerebro, podría ser debida a que más genes se seleccionan positivamente en el linaje humano o debida a una relajación de las restricciones y por tanto menor selección negativa de genes que se expresan en cerebro humano respecto del cerebro de chimpancé.

La acción de la selección positiva en el linaje humano se ve reforzada por dos observaciones. En primer lugar, la mayoría de los cambios en región reguladora en el linaje humano ha dado lugar a aumentos en los niveles de expresión (Caceres *et al.*, 2003; Gu *et al.*, 2003). La segunda observación se obtiene cuando los cambios en la expresión génica en cerebro, corazón, riñón e hígado del hombre y chimpancé se asignan a sus linajes. Únicamente en el caso del tejido cerebral hay una relación entre grupos de genes relacionados funcionalmente que cambian su expresión en el linaje humano y grandes regiones con desequilibrio de ligamiento en poblaciones humanas. Puesto que niveles aumentados de desequilibrio de ligamiento son un indicador indirecto de selección positiva, parece probable que este mecanismo esté actuando (Khaitovich *et al.*, 2006).

V.5 Selección positiva y esquizofrenia

La esquizofrenia presenta una importante paradoja evolutiva, puesto que a pesar de exhibir alta heredabilidad y efectos negativos en eficacia biológica, persiste con una

Introducción

prevalencia de aproximadamente 1% en las poblaciones humanas. Se ha sugerido que la predisposición genética a la esquizofrenia haya evolucionado como una consecuencia secundaria de la selección de rasgos cognitivos humanos (Crow, 1997; Horrobin, 1998). Esta hipótesis predice que genes que aumenten el riesgo a este desorden han sido sujetos a selección positiva en la historia evolutiva humana y otros primates.

Así, bajo la perspectiva de que genes asociados con esquizofrenia han sido sometidos a selección positiva, Crespi y colaboradores analizaron en 2007 la evolución molecular de 76 genes candidatos a aumentar el riesgo a padecer esquizofrenia (Crespi *et al.*, 2007). Se encontraron señales de selección positiva para 28 de los genes analizados. Otras evidencias de que la selección positiva ha actuado en la evolución de la esquizofrenia se muestran a continuación. En primer lugar, estudios neurológicos muestran que áreas desreguladas en esquizofrenia han sido las áreas más notablemente sujetas a cambios evolutivos diferenciales en el linaje humano (Brüne, 2004; Burns, 2006). En segundo lugar, Wayland y colaboradores encontraron que los genes que muestran selección positiva en la expresión diferencial entre humanos y chimpancés están desregulados en la corteza prefrontal dorsolateral y corteza orbitofrontal de individuos con esquizofrenia (Wayland *et al.*, 2006). Por último, recientes análisis que combinan el estudio de cambios metabólicos en esquizofrenia con diferencias en expresión entre los linajes humano y de primates sugieren que procesos metabólicos que se encuentran alterados en el cerebro de esquizofrénicos evolucionaron rápidamente en el linaje humano pero no en el de chimpancé (Khaitovich *et al.*, 2008).

V.6 Evolución del gen *FOXP2* como ejemplo de selección positiva en el linaje humano

La secuencia del gen *FOXP2* se encuentra muy conservada, incluso en especies de vertebrados lejanas, como son las aves, con las que tiene un 98.9% de identidad, o con el pez cebra, con quien tiene un 85.5% (Teramitsu *et al.*, 2004; Bonkonwsky y Chien, 2005).

Evolución en el linaje humano

A pesar del elevado grado de conservación del gen, en el linaje humano se produjo un notable incremento (>60 veces) en la tasa de sustitución aminoacídica (Enard *et al.*, 2002b; Zhang *et al.*, 2002). El gen *FOXP2* humano y su ortólogo de ratón difieren sólo en tres aminoácidos, de los cuales dos cambios ocurrieron después de la divergencia del linaje humano con el del chimpancé. Así, teniendo en cuenta que los linajes que dan lugar a humanos y ratones divergieron hace unos 70 millones de años, durante los 130 millones de años que separan al común ancestro de humanos y chimpancés del ratón, tuvo lugar un único cambio de aminoácido en la proteína FOXP2. Sin embargo, desde que el linaje humano se separó del linaje del chimpancé, hace unos 5 millones de años, tuvieron lugar dos cambios en el linaje humano mientras que, con excepción de un cambio en el linaje del orangután, ninguno ocurrió en los otros linajes de primates (figura I10).

Los cambios en el linaje humano se encuentran localizados en el exón 7 y constituyen un cambio de treonina por asparragina en la posición 303 de la proteína, y un cambio de asparragina por serina en la posición 325. Este último crea una potencial diana de

fosforilación por una proteína kinasa C, lo que puede ser importante puesto que en factores de transcripción *forkhead* la fosforilación de este dominio constituye un mecanismo importante de regulación transcripcional (Brunet *et al.*, 1999; Kops *et al.*, 2002). Este cambio también da lugar a un cambio menor de la estructura secundaria (Enard *et al.*, 2002b). Ambos cambios se encuentran fijados en la población humana (Enard *et al.*, 2002b; Zhang *et al.*, 2002).

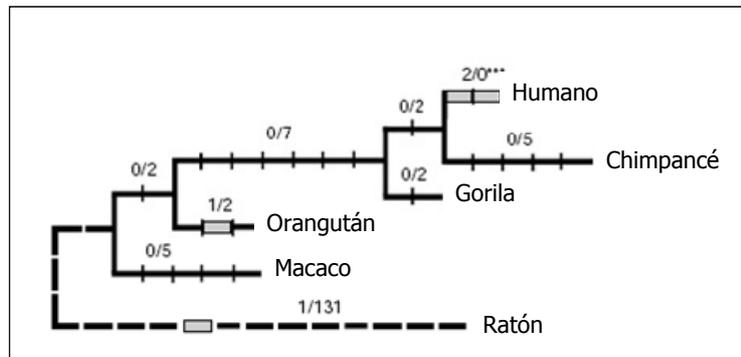


Figura I10. Filogenia de primates con sustituciones nucleotídicas silenciosas/no silenciosas del gen *FOXP2*. Las barras representan cambios nucleotídicos, los rectángulos grises indican los cambios de aminoácidos. Figura tomada de Enard *et al.*, 2002b con modificaciones.

Con el objetivo de determinar si el gen *FOXP2* había sido diana de la selección durante la evolución humana, se llevó a cabo un análisis detallado de la variación nucleotídica a lo largo de una región del gen de 14 kilobases cercana a la región de las sustituciones aminoacídicas (intrones 4, 5 y 6), en diferentes poblaciones humanas, chimpancé y orangután. Si la selección positiva había actuado recientemente en el linaje humano, tendrían que encontrarse rastros de un "arrastre selectivo", tal como se ha indicado en el párrafo anterior. Por una parte, valores negativos del estadístico D en el test de Tajima indicarían más alelos con baja frecuencia de los esperados bajo un modelo de evolución neutra y en segundo lugar valores negativos del estadístico H reflejarían una mayor frecuencia de alelos no ancestrales obtenidos por recombinación que en el modelo neutro. Se obtuvieron valores significativamente negativos para ambos estadísticos, lo que indicaba la acción de la selección positiva en el *locus FOXP2* en el linaje humano, siendo los candidatos más probables los dos cambios aminoacídicos específicos del linaje humano situados en el exón 7 (Enard *et al.*, 2002b). Mediante el test Hudson-Kreitman-Aguadé se han obtenido resultados similares que apoyan la intervención de la selección positiva (Zhang *et al.*, 2002)

Dado el interés del gen *FOXP2* en relación con el lenguaje, no tardó en plantearse la posibilidad de que los dos cambios en el linaje humano fueran decisivos en el desarrollo del lenguaje. La estima del momento de fijación de estos cambios es hace aproximadamente 200.000 años (Enard *et al.*, 2002b), tiempo correspondiente o inmediato a la emergencia del hombre moderno, lo cual es compatible con un modelo en el que la expansión del hombre moderno estuviera dirigida por la aparición de un lenguaje hablado eficiente, y en el que el gen *FOXP2* pudiera jugar un papel importante. Sorprendentemente, el cambio en la posición 325 también se encuentra en el orden de los carnívoros (Zhang *et al.*, 2002).

Introducción

En el año 2007, la secuenciación en Neandertal de las regiones donde se encuentran los cambios específicos de la especie humana permitió determinar que ambas sustituciones aminoacídicas se encontraban en Neandertal, por lo que debían estar presentes en el antecesor común (Krause *et al.*, 2007). Además, en Neandertal, los cambios estaban localizados en el haplotipo del hombre moderno, sujeto a arrastre selectivo.

No obstante, estos resultados han sido cuestionados por autores posteriores (Coop *et al.*, 2008), quienes apuntan a que el escenario propuesto por Krause y colaboradores no permite explicar dos de las características observadas de variación de los humanos modernos. En primer lugar, el exceso de alelos derivados de elevada frecuencia en intrones cercanos a las dos sustituciones, compatible con una fijación reciente de un alelo beneficioso y la recombinación entre los *loci* seleccionado y neutro durante el arrastre selectivo, junto con el hecho de que uno de los sitios intrónicos es polimórfico en los 2 neandertales analizados indica que el alelo ancestral tampoco se había perdido de la población de neandertales. En segundo lugar, Krause y colaboradores predicen que el arrastre selectivo habría terminado alrededor de hace 300.000 años aunque el haplotipo seleccionado parece haber acumulado pocas mutaciones desde entonces. Mediante datación filogenética Coop y colaboradores datan el ancestro común más cercano del haplotipo seleccionado hace aproximadamente 42.000 años, sorprendentemente reciente si la selección tuvo lugar hace 300.000 años. Con el fin de conciliar los datos de humanos modernos y neandertales, proponen diferentes explicaciones posibles. En la primera de ellas indican que si se asume que uno o dos de los cambios aminoacídicos era diana de selección positiva los datos de neandertal podrían reflejar flujo entre ambas poblaciones (asumiendo que un alelo fuera beneficioso, serían suficientes niveles bajos de flujo génico para extenderlo a otra población). En la segunda explicación sugieren que el episodio de selección positiva reciente en el hombre moderno no hubiera involucrado a los sitios aminoacídicos. Otra posibilidad no descartada y desgraciadamente común en estudios con DNA de neandertal, es la contaminación de la muestra con DNA de humano moderno.

Evolución en otros linajes de vertebrados

Los estudios sobre la estructura del gen *FoxP2* en aves con aprendizaje vocal como el pinzón cebra aportaron nuevos datos sobre la evolución del gen (Teramitsu *et al.*, 2004). Se encontró que aunque la proteína resultante no contenía los dos cambios específicos de la especie humana dentro de los primates, presentaba cinco residuos no presentes en mamíferos. Esto sirvió para postular que a pesar de la ausencia de un cambio específico en el hombre resultante de selección positiva durante la reciente evolución de los primates, esto no excluía la posibilidad de que la presión de selección actuara independientemente sobre la evolución del aprendizaje vocal en aves (Haesler *et al.*, 2004). También se planteó el que entre mamíferos y entre pájaros tuviera lugar una variación independiente de la estructura secundaria de *FoxP2* que contribuyera a la capacidad de aprendizaje vocal en ciertas especies.

Posteriores estudios con otras especies con aprendizaje vocal, como los murciélagos ecolocalizadores, algunos cetáceos e incluso un tipo de elefante africano han mostrado que estas especies no comparten de forma consistente mutaciones no sinónimas, por lo que no parece haber mutaciones asociadas con la habilidad general del aprendizaje vocal (Li *et al.*, 2007). El único caso que se aproxima se trata de dos de los tres cambios que ocurrieron en cetáceos (Prolina en posición 302 y Metionina en posición 316), que se encuentran de forma independiente en algunas familias de murciélagos. Curiosamente, el único cambio presente en humanos, una Treonina en posición 303,

se encuentra flanqueado por dos cambios en ballena y delfín: Prolina en posición 302 y Treonina en 304 (Webb y Zhang, 2005). Un dato interesante de los estudios con secuencias de murciélagos es el hecho de que el gen ha sufrido evolución acelerada en murciélagos y además la elevada variación en los exones 7 y 17 se encuentra correlacionada aproximadamente con los límites de los árboles filogenéticos para las diferentes familias de murciélagos (Li *et al.*, 2007).

V.7 RNA no codificante: HAR1A

En una reciente búsqueda de regiones genómicas conservadas ancestralmente, se describió el elemento HAR1 (de *Human Accelerated Region 1*), como aquel con tasa de sustitución más acelerada del linaje humano (Pollard *et al.*, 2006): 18 sustituciones fijadas en población humana desde la separación del ancestro común con chimpancé frente a las 0.26 esperadas en esta región en otros amniotas. HAR1 se localiza en 20q13.33 y es parte del gen de RNA *HAR1A*.



Figura I. Estructura del gen *HAR1A*. En la figura aparece representado el gen *HAR1A* en relación al gen *HAR1R*, así como la región con la predicción de estructura secundaria de RNA y el grado de conservación obtenido de la base de datos del *UCSC Genome Browser*.

Su presencia en el pollo, pero no en la rana ni en invertebrados permite estimar su tiempo de aparición entre hace 300 y 400 millones de años. Por otra parte, la ausencia de huellas de barrido selectivo indica que aunque los cambios ocurrieron en el linaje humano, debieron ocurrir hace más de un millón de años.

HAR1 se localiza en 20q13.33 y forma parte de un par de genes solapantes *HAR1A* (o *HAR1F*) y *HAR1R*. Aunque la región de solapamiento de ambos genes se encuentra conservada desde el ancestro común con pollo hasta el ancestro común con chimpancé, el resto de ambos transcritos ha evolucionado más rápidamente.

Programas de predicción de codificación de proteínas estiman que *HAR1A* y *HAR1R* no codifican para ninguna proteína, aunque ambos transcritos pueden plegarse para dar lugar a estructuras secundarias de RNA, siendo la de *HAR1A* la más estable. El análisis comparativo entre las estructuras secundarias esperadas para este transcrito entre chimpancé y humano, indica diferencias significativas entre ambas, siendo la obtenida en humano una estructura nueva sin homología con ningún RNA de las bases de datos. Esto dio lugar a su consideración como genes de RNA no codificantes nuevos.

v.7.1 Expresión de *HAR1A* y *HAR1R*

HAR1F se expresa intensa y específicamente en el desarrollo del neocórtex en el ser humano, durante un período crítico para la migración y especificación neuronal cortical. Así, durante el desarrollo temprano de la corteza cerebral se expresa en células de la placa cortical superior que están acabando la migración y coexpresa con reelina en las neuronas de Cajal-Retzius de la capa granular subpial, capa que se desarrolla únicamente en la especie humana, hasta la semana de gestación 17-19. En posteriores estadios, *HAR1A* deja de expresarse en neuronas de Cajal-Retzius y se expresa en otras regiones cerebrales diferentes al córtex, como el primordio hipocampal, el giro dentado, el córtex cerebelar en desarrollo y algunos núcleos como el complejo olivar. La expresión en el córtex se encuentra altamente conservada desde la divergencia entre homínidos y monos del viejo mundo, hace unos 25 millones de años.

En adultos, *HAR1R* y *HAR1A* se detectan en cerebro, ovario y testículo. Además *HAR1A* también se detecta en cerebelo y estructuras del cerebro anterior como córtex, hipocampo, tálamo e hipotálamo. En general, los niveles de *HAR1A* son mayores, con excepción de testículo, donde ambos transcritos están presentes en pequeñas pero similares cantidades. En ratón, sin embargo, los niveles de *HAR1A* y *HAR1R* son los mismos en cerebro fetal y adulto.

v.7.2 Función del gen *HAR1A*

La coexpresión del gen *HAR1A* con reelina en las células de Cajal-Retzius, así como posteriormente en áreas cerebrales como el córtex cerebelar y los núcleos olivares que corresponden a estructuras en cuya formación del patrón interviene la ruta de la reelina permite plantear la hipótesis de que *HAR1A* influya en la expresión de la reelina o sus receptores, directa o indirectamente (Pollard *et al.*, 2006). Interesantemente, la región donde se localiza *HAR1A*, 20q13 es una de las 8 para las que se ha encontrado mayor ligamiento con esquizofrenia en la iniciativa genética del NIMH (Freedman *et al.*, 2001) y también con psicosis con pedigrís de desorden bipolar (Park *et al.*, 2004), además de sugerirse que defectos en la expresión de la reelina pueden estar relacionados con la esquizofrenia (Impagnatiello *et al.*, 1998; Eastwood *et al.*, 2006). Sin embargo el primer estudio de asociación de polimorfismos del gen *HAR1A* con esquizofrenia, concretamente con alucinaciones auditivas, no indica diferencias significativas entre pacientes y controles, aunque sugiere una tendencia al comparar pacientes con alucinaciones auditivas con pacientes sin este síntoma (Tolosa *et al.*, 2008).

Asimismo, el patrón de expresión de *HAR1R* sugiere que tal vez se exprese más tarde en el desarrollo que *HAR1A* con el fin de regular a este por inhibición antisentido, como ocurre con otros genes del desarrollo.

Hipótesis y objetivos

Hipótesis

La hipótesis central de esta tesis es que *FOXP2*, un gen sometido a selección positiva en el linaje humano y directamente relacionado con alteraciones en el lenguaje, está implicado en la susceptibilidad a padecer esquizofrenia. Esta hipótesis general asume como hipótesis de trabajo:

- La existencia de diferencias significativas en las frecuencias alélicas y genotípicas de los polimorfismos del gen *FOXP2*, en relación con el desarrollo de alucinaciones auditivas en pacientes esquizofrénicos y con variantes clínicas relacionadas con el lenguaje.
- La existencia de diferencias en los niveles de expresión del gen en diferentes áreas cerebrales entre pacientes y controles.

Objetivos

Este trabajo está centrado en evaluar la implicación del gen *FOXP2* en el desarrollo de la esquizofrenia, especialmente en el estudio de las bases genéticas de las alucinaciones auditivas, como fenotipo alternativo a la esquizofrenia. Para ello se plantearon una serie de objetivos concretos:

- Analizar las diferencias en las frecuencias alélicas, genotípicas y haplotípicas de individuos control respecto a las de pacientes con alucinaciones auditivas y sin ellas, así como respecto a variables de las alucinaciones auditivas y del lenguaje.
- Determinar la presencia de expresión diferencial del gen en diferentes áreas cerebrales en pacientes y controles, así como correlacionarla con los patrones de metilación de la isla CG localizada en el primer exón del gen.

El gen *FOXP2* ha sido objeto de intensos estudios evolutivos, centrados especialmente en la secuencia codificante. Sin embargo, la región reguladora apenas ha sido analizada. Por estas razones, se planteó como objetivos adicionales:

- El análisis evolutivo de la región 5' del gen en primates.
- La identificación de elementos conservados y potencialmente funcionales en la región reguladora localizada en 5', así como la evaluación de su posible funcionalidad.

Material y métodos

I. Material biológico

I.1 Material biológico humano

A lo largo de este trabajo se han utilizado diferentes muestras de DNA y RNA. A continuación, se presentan todas ellas, en función del estudio en el que se hayan empleado.

I.1.1 Muestras utilizadas en el estudio de asociación

La muestra con la que se trabajó consistió en 293 pacientes y 340 controles, todos ellos de etnia caucásica y origen español, procedentes de la provincia de Valencia, área 4 de Salud, Hospital Clínico. Los criterios de exclusión para ambos grupos incluyeron síndromes cerebrales orgánicos, retraso mental, abuso severo de drogas o tener una incoherencia en el pensamiento-lenguaje que incapacitara para llevar a cabo la evaluación clínica.

El estudio fue aprobado por el Comité de Ética Local y todos los participantes dieron el consentimiento informado por escrito.

I.1.1.1 Pacientes

I.1.1.1.1 Diagnóstico

Todos los pacientes cumplían los criterios del DSM-IV-TR (APA, 2002) de diagnóstico incluidos en el apartado de Esquizofrenia y otros Trastornos Psicóticos o Trastorno Bipolar (códigos 295-299).

El diagnóstico fue confirmado para cada paciente a través de un consenso entre el psiquiatra que trataba al paciente y dos psiquiatras *senior* (con más de 10 años de experiencia) del grupo de investigación. Un resumen de los diagnósticos se muestra en la tabla M1.

Diagnóstico	Número de individuos
Esquizofrenia	214
Trastorno Esquizoafectivo	16
Psicosis breve	2
Trastorno delirante	14
Trastorno bipolar	34
Trastorno psicótico no especificado	5
Depresión	4
Psicosis atípica	3
Psicosis tóxica	1

Tabla M1. Diagnóstico de los pacientes incluidos en el estudio.

Material y Métodos

I.1.1.1.2 Características demográficas

El rango de edad oscilaba entre 18-77 años (media=40.32, Desviación estándar=11.917) y la distribución por sexos era de 33.1% mujeres y 66.9% hombres.

En cuanto al estado civil un 67.5% eran solteros, un 17.95% casados, un 10.68% separados o divorciados y el resto viudos o conviviendo con pareja durante un mínimo de seis meses.

Respecto al nivel de estudios un 50.43% tenían estudios primarios, un 24% habían cursado FP o BUP, un 11.54% sólo sabían leer y escribir, un 10.68% tenían estudios universitarios y el resto eran analfabetos.

En cuanto al tratamiento farmacológico, un 45.92% estaban siendo tratados con antipsicóticos atípicos, un 18.36% con antipsicóticos típicos, un 29.9% con antipsicóticos mixtos y el resto, un 5.78%, no estaban tomando medicación en el momento de la evaluación.

I.1.1.1.3 Variables clínicas analizadas

Dentro de las variables clínicas evaluadas, aquellas utilizadas para llevar a cabo los tests de asociación y cuyos estados y número de individuos de cada grupo se muestran en la tabla M2, son las siguientes:

- **Historia clínica de alucinaciones auditivas (AA)**

Recogida por la entrevista y la historia clínica previa.

- **AA en el momento de realizar la evaluación clínica**

Entre los pacientes con historia de AA, un grupo de pacientes refirieron historia de antecedentes de haber sufrido AA pero no de padecerlas en el momento de la prueba, mientras que en el otro grupo refirieron haber tenido alucinaciones al menos en la última semana antes de la evaluación. A ambos grupos se les evaluó la intensidad de las alucinaciones mediante la escala PSYRATS.

- **Cronicidad de las AA**

Los alucinadores crónicos se definieron según los siguientes criterios:

1. Las alucinaciones persisten a lo largo del último año
2. Están presentes al menos una vez al día
3. Las alucinaciones persisten después de tratamiento con al menos dos tipos de antipsicóticos con dosis equivalentes a 600mg de clorpromacina o superiores.

- **Edad de inicio de la enfermedad**

Se consideró la edad de inicio como la edad en que por primera vez el paciente pidió ayuda a un profesional médico o psicólogo por sus problemas psíquicos aunque no recibiera tratamiento. Existe pues un margen desconocido y no controlado de tiempo en el que el paciente pudiera desarrollar problemas sin acudir a un facultativo. Sin embargo determinar el momento exacto quedaba fuera de las posibilidades del grupo de investigación.

Se decidió establecer un punto de corte de 20 años y otro en 28, de forma que se compararan únicamente los grupos extremos con inicio temprano o inicio tardío de la enfermedad. Con esto se pretendió tener grupos más homogéneos para comparar.

Variable	Estado de la variable	Número de individuos
Presencia de alucinaciones auditivas	Con historia de alucinaciones auditivas	215
	Sin historia de alucinaciones auditivas	77
Cronicidad de las alucinaciones auditivas	Presencia de cronicidad	135
	Ausencia de cronicidad	80
Inicio de la enfermedad	Inicio temprano (hasta los 20 años)	60
	Inicio tardío (a partir de los 28 años)	89

Tabla M2. Variables clínicas utilizadas en los estudios de asociación.

I.1.1.1.4 Escalas clínicas Manchester y PSYRATS

Las escalas Manchester (Krawiecka *et al.*, 1977) y PSYRATS (Haddock *et al.*, 1999) fueron utilizadas para evaluar los síntomas clínicos psicóticos y la intensidad de las alucinaciones auditivas, respectivamente.

La **Escala Manchester**, en la versión española adaptada por Pérez-Fuster y colaboradores (Pérez-Fuster *et al.*, 1989), fue administrada para la valoración del estado clínico general. Está diseñada para la evaluación de pacientes psicóticos crónicos y recoge síntomas positivos, negativos y afectivos. Consta de 8 ítems: depresión, ansiedad, aplanamiento afectivo, retardo psicomotor, delirios, alucinaciones, incoherencia del lenguaje y pobreza del lenguaje. En el estudio de asociación se tuvieron en cuenta los dos últimos ítems.

La **Escala de Evaluación de Síntomas Psicóticos**, o escala PSYRATS (Haddock *et al.* 1999) se administró con el objetivo de evaluar las distintas dimensiones de las AA. Esta escala, validada para la población española por nuestro grupo de investigación (González *et al.*, 2003) se compone de dos subescalas para la valoración de AA y delirios. Permite el diagnóstico y evaluación psicopatológica de las distintas dimensiones de las AA, evalúa la severidad de los síntomas a lo largo de su evolución y mide el cambio tras el tratamiento.

En esta escala, las dimensiones de las AA se miden en una escala de 5 puntos (de valores 0 a 4). Las dimensiones que considera son: frecuencia, duración, localización de las voces (dentro o fuera del espacio interno), volumen, creencias acerca del origen de las voces, cantidad de contenido negativo, grado de contenido negativo, cantidad de angustia que producen las voces, intensidad de la angustia, disrupción para la vida producida por las voces y capacidad de control de las alucinaciones.

En la caja siguiente aparecen reflejados los ítems de ambas escalas utilizados en el presente estudio. Para el estudio de asociación se utilizaron todos los ítems reflejados en la escala PSYRATS, así como la puntuación total de esta escala y dos ítems de la escala Manchester: Incoherencia del lenguaje y Pobreza del lenguaje.

Material y Métodos

Caja M1	
Escala PSYRATS	
Ítems de las alucinaciones auditivas	
<i>Frecuencia de aparición.</i>	<i>Frecuencia de contenido negativo de las voces (cont.)</i>
0. Voces no presentes o presentes menos de una vez por semana. 1. Voces que suceden al menos una vez por semana. 2. Voces que se presentan al menos una vez al día. 3. Voces que se presentan al menos una vez por hora. 4. Voces que se presentan continuamente o casi continuamente, por ejemplo, se paran sólo segundos o minutos.	3. Insultos personales relacionados con la autovaloración personal, por ejemplo, "eres perezoso", "eres malo", "pervertido". 4. Amenazas personales; por ejemplo amenazas de hacerle daño a él o su familia, órdenes de autolesionarse o lesionar a otros.
<i>Duración.</i>	<i>Frecuencia con la que producen ansiedad.</i>
0. Voces no presentes. 1. Voces que duran unos cuantos segundos, fugaces. 2. Voces que duran algunos minutos. 3. Voces que duran al menos 1 hora. 4. Voces que duran horas.	0. Las voces no producen ansiedad. 1. Sólo ocasionalmente producen ansiedad. 2. Sólo producen ansiedad en algunas ocasiones. 3. Producen ansiedad la mayoría de las veces. 4. Las voces producen ansiedad siempre.
<i>Localización.</i>	<i>Repercusión en la vida diaria causada por las voces.</i>
0. Voces no presentes. 1. Voces que se oyen dentro de la cabeza únicamente. 2. Voces fuera de la cabeza, pero determinadas a los oídos o a la cabeza. También podrían oírse dentro de la cabeza. 3. Voces que suenan como dentro o cerca de los oídos y fuera de la cabeza, apartada de los oídos. 4. Voces que suenan sólo fuera de la cabeza.	0. No existencia de repercusión, es capaz de mantener relaciones sociales y familiares. 1. Las voces causan una mínima repercusión en la vida del sujeto; por ejemplo, interferencia en la concentración, aunque es capaz de mantener actividad diaria, relacionarse social y familiarmente y ser capaz de mantener independencia sin apoyo. 2. Las voces causan una repercusión moderada, provocando alguna alteración en la actividad diaria y en la familia o en las actividades sociales. El paciente no está hospitalizado, aunque puede vivir en residencias de apoyo o recibir ayuda adicional en el desarrollo de habilidades diarias. 3. Las voces causan una repercusión severa, por lo que la hospitalización es necesaria. El paciente es capaz de mantener algunas actividades diarias, cuidarse por sí mismo y relacionarse en el hospital. Podría también estar en alojamientos de apoyo, pero experimenta importantes trastornos en términos de actividades, desarrollo de habilidades y/o relaciones. 4. Las voces causan una completa alteración de la vida diaria, requiriéndose hospitalización. El paciente es incapaz de mantener ninguna actividad diaria ni relacionarse. También están severamente alterados los cuidados propios.
<i>Intensidad (volumen)</i>	
0. Voces no presentes. 1. Más bajas o silenciosas que nuestra propia voz (susurro). 2. Semejantes a nuestra propia voz. 3. Más fuerte que nuestra propia voz. 4. Extremadamente fuerte (gritando)	
<i>Grado de convicción del origen de las voces.</i>	
0. Voces no presentes. 1. Voces generadas sólo internamente y relacionadas con uno mismo. 2. Alguna convicción de que las voces procedan de causas externas. 3. Fuerte convicción de que las voces procedan de causas externas. 4. Las voces son únicamente debidas a causas externas (100% de convicción)	
<i>Cantidad de contenido negativo.</i>	<i>Control sobre las voces.</i>
0. No existencia de contenido negativo en las voces. 1. Contenido desagradable de forma ocasional (<10%). 2. La minoría del contenido de las voces es desagradable o negativo (<50%). 3. La mayoría del contenido de las voces es desagradable o negativo (>50%). 4. Todo el contenido de las voces es desagradable o negativo.	0. Los pacientes creen que pueden controlar las voces, atrayéndolas o disipándolas (rechazándolas). 1. El sujeto cree poder tener algún control de las voces, en la mayoría de las ocasiones. 2. Control de las voces la mitad del tiempo aproximadamente. 3. El sujeto cree tener control de las voces, pero sólo ocasionalmente; la mayoría del tiempo el sujeto experimenta voces que no puede controlar. 4. El sujeto no tiene control sobre las voces y no puede rechazarlas o atraerlas.
<i>Frecuencia de contenido negativo en las voces.</i>	
0. No desagradable o negativo. 1. Algún grado de contenido desagradable o negativo, pero no relacionados con uno mismo o su familia, por ejemplo, tacos o comentarios no dirigidos a uno mismo: "El lechero es feo". 2. Insultos personales, comentarios sobre el comportamiento, por ejemplo: "no deberías hacer o decir eso".	
Escala Manchester	
Ítems seleccionados.	
<i>Pobreza del lenguaje.</i>	<i>Incoherencia del lenguaje</i>
0. Ausente. 1. Leve. 2. Moderada. 3. Marcada. 4. Grave	0. Ausente. 1. Leve. 2. Moderada. 3. Marcada. 4. Grave

I.1.1.2 Controles

Los controles fueron seleccionados a partir de una muestra de población general. Se les aplicó un pequeño cuestionario para descartar antecedentes psiquiátricos personales y/o familiares de primer grado. El criterio de exclusión utilizado fue el haber recibido cualquier tipo de tratamiento psiquiátrico.

El rango de edad oscilaba entre 18-91 años (media=37.99, Desviación estándar=15.630) y la distribución por sexos era de 38.6% mujeres y 61.4% hombres.

I.1.2 Muestra con Trastorno Específico del Lenguaje (TEL)

Dentro de la búsqueda de mutaciones en el gen *FOXP2*, se estudió también una muestra de 10 niños con Trastorno Específico del Lenguaje (según criterios diagnósticos DSM-IV-TR código 315) seleccionados en una muestra de población escolar de la ciudad de Castellón, España.

I.1.3 Muestras de tejido *postmortem* humano

Para llevar a cabo el análisis de metilación del promotor del gen *FOXP2*, así como de la expresión de este gen en diferentes áreas cerebrales se partió principalmente de muestras de tejido cerebral humano *postmortem*, que habían sido donadas al centro de investigación *SANE POWIC (SANE Prince of Wales Internacional Centre)* de Oxford por el Banco de Cerebros de Enfermedades Neurodegenerativas de Londres. Las muestras correspondían a diferentes regiones de la corteza cerebral, y procedían de diferentes pacientes esquizofrénicos e individuos control. En concreto, se trabajó con muestras de ambos hemisferios de las siguientes regiones:

- Circunvolución Temporal Superior
- Circunvolución del Parahipocampo
- Circunvolución del Cíngulo
- Paracingulado

El número y detalle de muestras utilizadas para cada análisis se especificará en el apartado correspondiente.

I.2 Muestras de Primates

Para el análisis filogenético y parte del análisis de metilación de la región promotora del gen *FOXP2*, se utilizaron muestras de DNA procedentes de diferentes especies de primates. La mayoría de estas muestras procedían del *Coriel Institute for Medical Research*. En la tabla M3, aparecen reflejadas estas muestras, junto con el número identificador del proveedor y la procedencia del DNA.

Además de las muestras mencionadas, también se utilizó DNA genómico de *Gorilla gorilla* amablemente cedido por el Dr. Jaume Bertranpetit (código de las muestras pg12 y pg13); DNA genómico de cerebro de *Pan troglodytes* y DNA genómico de una línea celular de *Gorilla gorilla*, cedidas en el centro de investigación *SANE POWIC (SANE Prince of Wales Internacional Centre)* de Oxford. Estas dos últimas muestras fueron las utilizadas para el análisis de metilación.

Material y Métodos

Especie	Identificador del <i>Coriel Institute for Medical Research</i>	Procedencia del DNA
<i>Saguinus labiatus</i>	NG05308	Cultivo de fibroblasto a partir de biopsia de piel
<i>Lagothrix lagothricha</i>	NG05356	Cultivo de fibroblasto a partir de explantes de tejido de piel
<i>Macaca arctoides</i>	NA03443	Cultivo de linfoblastos
<i>Macaca fascicularis</i>	NA03446	Cultivo de fibroblasto a partir de biopsia de piel
<i>Macaca mulatta</i>	NG07109	Cultivo de fibroblasto a partir de biopsia de piel
<i>Macaca nigra</i>	NG07101	Cultivo de fibroblasto a partir de biopsia de piel
<i>Macaca nemestrina</i>	NG07921	Cultivo de fibroblasto a partir de biopsia de piel
<i>Erythrocebus patas</i>	NG06116	Cultivo de fibroblasto a partir de explantes de tejido de piel

Tabla M3. Muestras de DNA de primates procedentes del *Coriel Institute for Medical Research*.

I.3 Cepas bacterianas

Los genotipos de las cepas bacterianas utilizadas han sido los siguientes:

- ***E. coli* JM109**

recA1, endA1, gyrA96, thi, hsdR17 (rK⁻, mK⁺), relA1, supE44, Δ(lac proAB), [F', traD36, proAB, lacIqΔM15].

- ***E. coli* TOP10**

F' mcrA Δ(mrr-hsdRMS-mcrBC) Φ80/lacZΔM15 ΔlacX74 deoR recA1 endA1 araD139, Δ(ara leu)7697 galJ galK rpsL (Str^R) nupG.

- ***E. coli* XL-1Blue**

recA1 endA1 gyrA96 thi hsdR17 (rK⁻, mK⁺) supE44, relA1, λ⁻ Δ(lac) {F' proAB lacIqZ ΔM15 Tn10(tet^R)}.

I.4 Células eucariotas

- ***CHO-K1***

Línea celular procedente de fibroblastos de ovario de hámster cedida por la Dra. Isabel Martínez Garay

- ***H23***

Línea celular procedente de tumor de pulmón, cedida por D. Salvador Mena.

II. Métodos moleculares

II.1 Técnicas básicas comunes

II.1.1 Extracción de ácidos nucleicos

Extracción de DNA genómico humano a partir de sangre periférica

La extracción de DNA a partir de sangre periférica se llevó a cabo en las muestras utilizadas para el estudio de asociación. Se utilizaron dos métodos: extracción mediante fenol-cloroformo (Fritsch *et al.*, 1989) y utilización del kit de extracción *Puregene* (Gentra-System).

Extracción de DNA genómico humano a partir de saliva

Se extrajo DNA a partir de saliva en las muestras correspondientes a los niños con TEL. Para ello se utilizó el *Kit de Extracción de DNA SSS* (Real, Durviz SL).

Extracción de DNA genómico a partir de tejido cerebral postmortem

La extracción de DNA genómico de las muestras de tejido cerebral utilizadas para el análisis de metilación se llevó a cabo mediante el *kit de Extracción de DNA genómico de Nucleon* (Tepnel Life Sciences).

Extracción de DNA plasmídico

Las extracciones de DNA plasmídico integradas en diferentes experimentos se llevaron a cabo con los siguientes kits, siguiendo en cada caso las recomendaciones del fabricante:

- *High Pure Plasmid Isolation Kit* (Roche)
- *Genopure Plasmid Midi Kit* (Roche)
- *QIAprep Miniprep* (Qiagen)
- *GenElute™ Plasmid Miniprep Kit* (Sigma Aldrich)

Se especificará en los apartados posteriores correspondientes cuando se utilizaron los diferentes kits.

Extracción de RNA

La extracción de RNA a partir de muestras de tejido cerebral *postmortem* para el análisis de expresión se realizó utilizando el kit *RNeasy Lipid Tissue Mini Kit* (Qiagen).

II.1.2 Cuantificación de ácidos nucleicos

Las cuantificaciones de ácidos nucleicos se llevaron a cabo diluyendo el ácido nucleico en agua HPLC y midiendo absorbancia en un espectrofotómetro (*BioPhotometer*, Eppendorf; *Genequant pro RNA/DNA calculator*, Amersham Pharmacia Biotech).

Material y Métodos

En aquellos casos en los que el volumen de muestra no permitía llevar a cabo diluciones se obtuvo una estima de la concentración mediante comparación con un patrón de concentraciones conocidas en un gel de agarosa al 1% (ver apartado de electroforesis, epígrafe II.1.4).

II.1.3 PCR

La reacción en cadena de la polimerasa (PCR) fue utilizada ampliamente a lo largo de todo el trabajo. A continuación se muestra un listado de las polimerasas utilizadas:

- *Netzyme* (Fermentas)
- *Taq DNA Polymerase* (Roche).
- *Pwo SuperYield DNA Polymerase* (Roche).
- *Taq* polimerasa incluida en la mezcla de reacción comercial *REDTaq[®] ReadyMix[™] PCR Reaction Mix* (Sigma Aldrich).

En cada apartado se especificará cuando se utilizó cada una de ellas.

Las mezclas de reacción estándar utilizadas para llevar a cabo las amplificaciones se relacionan en la tabla M4. Además, cuando el contenido en GC del producto amplificado era elevado se utilizaron mezclas de reacción de PCR con solución *CG-Rich* (Roche) o glicerol al 50% y en el caso de obtenerse productos inespecíficos de menor tamaño que el deseado, se utilizaron las mezclas de reacción de PCR con DMSO.

Componentes de la mezcla	Mezcla estándar (μl)	Mezcla con <i>CG-Rich</i> / Glicerol (μl)	Mezcla con DMSO (μl)
DNA (25-100 ng)	1	1	1
Tampón 10X con MgCl ₂ ^a	2.5	2.5	2.5
dNTPs (10 mM)	0.5	0.5	0.5
Cebador directo (10 pmol/μl)	0.5-1	0.5-1	0.5-1
Cebador reverso (10 pmol/μl)	0.5-1	0.5-1	0.5-1
DNA Polimerasa (1-5 u/μl)	0.5	0.5	0.5
Solución <i>GC-Rich</i> /Glicerol 50%	-	5	-
DMSO 50%	-	-	5
H ₂ O	Hasta 25	Hasta 25	Hasta 25

Tabla M4. Mezclas de reacción utilizadas para la amplificación por PCR.

^a La concentración de MgCl₂ de los tampones utilizados correspondía a la proporcionada por la casa comercial correspondiente, siendo 1.5 o 2 mM.

En las reacciones de PCR realizadas en el análisis de metilación y de expresión se utilizó la mezcla de reacción comercial *REDTaq[®] ReadyMix[™] PCR Reaction Mix* (Sigma Aldrich). En este caso, puesto que ya incluía el tampón y dNTPs, la mezcla de reacción se muestra en la tabla M5.

Los componentes de la mezcla de reacción correspondiente a la polimerasa con prueba de lectura *Pwo SuperYield DNA Polymerase* (Roche), utilizada para la obtención de los fragmentos de la región reguladora utilizados en el análisis funcional se muestran en la tabla M6.

Componentes de la mezcla	Mezcla <i>REDTaq</i> (μl)
DNA (25-100 ng)	1
Cebador directo (10 pmol/μl)	1
Cebador reverso (10 pmol/μl)	1
<i>REDTaq</i> [®] <i>ReadyMix</i> [™]	10
H ₂ Od	Hasta 20

Tabla M5. Mezcla de reacción utilizada para las amplificaciones con *REDTaq*[®] *ReadyMix*[™] PCR Reaction Mix (Sigma Aldrich).

Componentes de la mezcla	Mezcla estándar (μl)
DNA (25-100 ng)	3
Tampón 10X con MgCl ₂	5
dNTPs (10 mM)	1
Cebador directo (10 pmol/μl)	1.5
Cebador reverso (10 pmol/μl)	1.5
<i>Pwo SuperYield</i>	0.75
Solución <i>GC-Rich</i> /Glicerol 50%	10
H ₂ Od	Hasta 50

Tabla M6. Mezcla de reacción utilizada para las amplificaciones con *Pwo SuperYield DNA Polymerase* (Roche).

Las reacciones de amplificación se llevaron a cabo en termocicladores Eppendorf Mastercycler (Eppendorf), MJ Research PTC 200 y MJ Research PTC 100 (MJ Research).

Los diferentes programas utilizados se muestran en la tabla M8.

En el caso de amplificarse fragmentos de más de 1 Kb y especialmente difíciles de amplificar se utilizó el siguiente programa:

Programa PCR fragmentos largos		
95°C	5'	
95°C	30"	x 15
54.5°C	45"	
72°C	2' 30"	
95°C	1'	x 25
55°C	50"	
72°C	2' + 15"/ciclo	
72°C	5'	

Tabla M7. Programa de PCR utilizado para la amplificación de fragmentos largos de DNA.

Material y Métodos

Programa estándar		Programa cDNA 1		Programa SNP-60		Programa Promotor 1		Programa Promotor 2	
95°C 5'		95°C 5'		94°C 4'		96°C 10'		95°C 10'	
95°C 30"	x 30	95°C 45"	x 35	94°C 30"	x 30	96°C 45"	x 35	95°C 40"	x 35
Tª 45"		Tª 45"		60°C 40"		Tª 45"		Tª 45"	
72°C 45"		72°C 45"		72°C 50"		72°C 50"		72°C 45"	
72°C 5'		72°C 5'		72°C 8'		72°C 10'		72°C 10'	
Programa Promotor 3		Programa Promotor 4		Programa Promotor 5		Programa Promotor 6		Programa Promotor 7	
96°C 15'		96°C 10'		96°C 15'		96°C 15'		96°C 15'	
96°C 1'	x 35	96°C 1'	x 35	96°C 1'10"	x 35	96°C 1'	x 35	96°C 1'	x 35
Tª 50"		Tª 45"		Tª 45"		Tª 1'		Tª 1'	
72°C 1'10"		72°C 45"		72°C 1'		72°C 1'		72°C 1'30"	
72°C 10'		72°C 10'		72°C 10'		72°C 10'		72°C 10'	
Poliglutaminas10		<i>GC-Rich</i>		Programa bisulfito					
96°C 5'		96°C 5'		95°C 5'					
96°C 40"	x 35	96°C 1'	x 35	95°C 1'	x 40				
Tª 40"		Tª 40"		Tª 1'					
72°C 50"		72°C 50"		72°C 1'					
72°C 5'		72°C 5'		95°C 5'					

Tabla M8. Programas de PCR utilizados

II.1.3.1 PCR de colonia

La PCR de colonia constituye un tipo de PCR especial en el que a la mezcla de reacción no se le añade DNA purificado, sino directamente una colonia bacteriana. Es un método rápido para determinar la presencia o ausencia de fragmentos de DNA en un clon bacteriano. Además, el producto obtenido se puede purificar y secuenciar como una PCR estándar. En general este tipo de PCR se ha llevado a cabo para comprobar o determinar la secuencia de fragmentos de DNA insertados en vectores plasmídicos. Es por esta razón por lo que se suelen utilizar cebadores específicos de vectores o los mismos cebadores de la reacción de PCR original, aunque también se pueden utilizar para este fin cebadores internos del producto de PCR.

En la tabla M9 se muestran los cebadores presentes en vectores plasmídicos utilizados para las amplificaciones mediante PCR de colonia.

Nombre	Secuencia (5'-3')
M13F	GTAAAACGACGGCCAG
M13R	CAGGAAACAGCTATGAC
T7	GTAATACGACTCACTATAGGGC
SP6	ATTTAGGTGACACTATA
pSEAP2-F	CTAGCAAATAGGCTGTCCC
pSEAP2-R	CCTCGGCTGCCTCGCGTTCC

Tabla M9. Cebadores presentes en vectores plasmídicos utilizados para amplificaciones mediante PCR de colonia.

Para llevar a cabo este tipo de PCR se transfiere una muestra de la colonia de interés, con un palillo o punta estéril, a la mezcla de PCR. Con el fin de no perder todo el material bacteriano, antes de transferir la colonia al tubo de PCR se hace una estría en una placa para tener copia de la colonia.

II.1.4 Electroforesis

II.1.4.1 Electroforesis en gel de agarosa

Se llevaron a cabo electroforesis en geles de agarosa para comprobar PCRs, ver resultados de digestiones con enzimas de restricción, preparar productos para extraer DNA y cuantificación relativa de ácidos nucleicos.

La concentración de agarosa empleada se modificó en función del tamaño de los fragmentos de DNA a separar, utilizándose geles con 0.8, 1, 1.5, 2, 3 y 3.5%. Las electroforesis se realizaron utilizando como tampón TBE 1x con Bromuro de Etidio y los resultados se obtuvieron en el captador de imágenes Gelprinter plus (Tecnología para Diagnóstico e Investigación).

En la tabla M10 se muestran los tampones y disoluciones utilizados para la electroforesis en geles de agarosa.

Soluciones	Componentes
Tampón de carga 6X	Azul de bromofenol 0.05% Azul de xilencianol 0.05% EDTA pH8.0 100mM Glicerol 50%
TBE 1X	Tris Base 89 mM Ácido Bórico 89mM EDTA 2mM
Bromuro de Etidio	0.5 mg/ml en gel y TBE 1X

Tabla M10. Soluciones utilizadas para la electroforesis en geles de agarosa

II.1.4.2 Electroforesis en gel de acrilamida

Se llevaron a cabo electroforesis en gel de acrilamida en la genotipación de 3 de los SNPs del estudio de asociación, debido a no poder diferenciar en geles de agarosa las diferentes bandas de DNA esperadas. Se utilizaron geles de acrilamida al 12%, con la composición reflejada en la tabla M11.

El tampón de carga de las muestras fue el mismo que se utilizó para los geles de agarosa. Se aplicó un voltaje de 800V durante 2 horas.

El revelado se llevó a cabo siguiendo el método de tinción con plata descrito en Fritsch *et al.* (1989).

Material y Métodos

Componente	Volumen
Acrilamida-bisacrilamida (Solugel 29:1)	36 ml
TBE 5X	24 ml
Persulfato de amonio 10%	1.2 ml
TEMED	24 μ l
H ₂ O	60 ml

Tabla M11. Composición geles de acrilamida 12%.

En la tabla M12 se muestran las soluciones utilizadas para la fijación, tinción y revelado de geles de acrilamida.

Solución	Composición
Etanol	10% en H ₂ O
HNO ₃	1% en H ₂ O
AgNO ₃	12 mM en H ₂ O
Revelador	Na ₂ CO ₃ anhidro 29,6 g Formaldehido 37% 540 μ l H ₂ O hasta 1 litro
Ácido acético	5% en H ₂ O

Tabla M12. Soluciones utilizadas para la fijación, tinción y revelado de geles de acrilamida.

II.1.5 Purificación ácidos nucleicos

II.1.5.1 Productos de PCR

En ocasiones se purificaron productos de PCR, bien para su posterior secuenciación, clonación o utilización del DNA en otros experimentos. Al purificar se eliminan nucleótidos, cebadores, sales y enzimas que puedan interferir con otras reacciones.

La purificación se llevó a cabo mediante precipitación con acetato amónico 4M.

El protocolo es el siguiente:

- 15 μ l - 20 μ l de producto de PCR
- 15 μ l (aprox. 1 volumen) de acetato amónico 4M
- 90 μ l (aprox. 6 volúmenes) de etanol 100%
- Mezclar bien y centrifugar a temperatura ambiente durante 15' a 13.000 rpm.
- Eliminar el sobrenadante.
- Añadir 400 μ l de etanol 70%.
- Vórtex y centrifugar a temperatura ambiente durante 5' a 13.000 rpm.
- Eliminar el sobrenadante.
- Secar el pellet durante 5' a 37°C.
- Disolverlo en 20 μ l de H₂O.

II.1.5.2 Productos de la digestión con enzimas de restricción

La purificación del DNA digerido con enzimas de restricción se realizó precipitando con acetato amónico 4M (ver apartado anterior), manteniendo la relación entre los volúmenes o bien utilizando el kit *Qiaex II* (Qiagen).

II.1.5.3 Purificación a partir de geles de agarosa

En algunos casos fue necesario purificar determinadas bandas de DNA procedentes de PCRs o digestiones con enzimas de restricción a partir de geles de agarosa. En estos casos, se cargó la totalidad de la PCR o digestión en un gel de agarosa al 0.8 o 1%. Después de correr la electroforesis se cortaron las bandas de interés y se extrajo el DNA con el kit *GFX PCR DNA and Gel Band Purification Kit* (GE Healthcare), o bien el kit *Qiaex II* (Qiagen) eluyendo o resuspendiendo respectivamente en 30 μ l de H₂O.

II.1.6 Secuenciación

La técnica de secuenciación se utilizó ampliamente a lo largo del trabajo, tanto para comprobar la secuencia de productos de PCR o clonados, como para obtener contigs de regiones de interés. El método de secuenciación aplicado está basado en el método de los dideoxynucleótidos (Sanger *et al.*, 1977) y el kit utilizado ha sido el *BigDye[®] terminator v3.1 Cycle sequencing kit* (Applied Biosystems), que incorpora en una única mezcla todos los componentes necesarios para la secuenciación, excepto el DNA y el cebador. Así, en la mezcla del kit están incluidos los cuatro dNTPs, una DNA polimerasa, el tampón de reacción y una mezcla de dideoxynucleótidos ddATPs, ddCTPs, ddGTPs y ddTTPs marcados con fluorocromos diferentes. El DNA que se quiere secuenciar actúa de molde y cuando la polimerasa incorpora un ddNTP la síntesis se interrumpe, generándose una mezcla de fragmentos de diferentes tamaños con la última base marcada.

Las reacciones se llevaron a cabo en termocicladores Eppendorf Mastercycler y MJ Research PTC 200. Los productos obtenidos fueron enviados al Servicio de Secuenciación de la Universitat de València, donde se purificaron y separaron los fragmentos por electroforesis capilar en un secuenciador (Applied Biosystems 3730 DNA Analyzer). Parte de las reacciones llevadas a cabo para el análisis de metilación fueron realizadas por el Servicio de Secuenciación del Departamento de Bioquímica de la Universidad de Oxford.

Las mezclas de reacción utilizadas para la secuenciación fueron las siguientes:

Componentes de la mezcla	Mezcla estándar (μ l)	Mezcla con <i>CG-Rich</i> /Glicerol (μ l)
DNA ^a	1-4	1-4
Cebador directo (10 pmol/ μ l)	1	1
<i>BigDye[®]</i>	2	2
Solución <i>GC-Rich</i> /Glicerol 50%	-	2
H ₂ O	Hasta 10	Hasta 10

Tabla M13. Mezclas de reacción de secuenciación.

^a La concentración de los productos de PCR purificados no se cuantificó, y la cantidad a secuenciar se estimó en función del rendimiento de la PCR en cuestión. Para los productos clonados, de forma estándar se secuenció 1 μ l.

El *BigDye[®] terminator v3.1 Cycle sequencing kit* se usó sin diluir en aquellos casos en que el fragmento a secuenciar era de gran longitud, por ejemplo al secuenciar clones

Material y Métodos

o, en el caso de PCRs, cuando estas eran difíciles de amplificar. En general, para fragmentos cortos se aplicó una dilución 1:2 en 400 mM Tris-HCl pH9 con 10 mM MgCl₂, preparada siempre en el momento.

Los programas de secuenciación fueron los siguientes:

Programa estándar Sec-L		Programa <i>GC-Rich</i>	
96°C 5'		96°C 10'	
96°C 30"	x 40 ^a	98°C 30"	x 40 ^a
50°C 15"		50°C 15"	
72°C 4'		72°C 4'	
4°C		4°C	

Tabla M14. Programas de secuenciación.

^a. El número de ciclos se aumentó a 50 cuando los productos a secuenciar eran de gran longitud.

El programa *GC-Rich* se empleó en aquellos casos en los que había un elevado contenido en GC en el producto.

II.1.7 Clonación

Las clonaciones realizadas en esta tesis se llevaron a cabo en tres vectores diferentes, pCR[®]2.1-TOPO[®] (Invitrogen), pGEM-T[®] Easy y pSEAP2-Basic (Promega). Los dos primeros contienen dos timidinas desapareadas en los extremos linearizados, por lo que resultan de gran utilidad cuando se quiere clonar productos de PCR, ya que gran parte de las DNA polimerasas utilizadas incorporan de forma inespecífica una adenina en los extremos del amplificado.

El vector pCR[®]2.1-TOPO[®] forma parte del kit de clonación *TOPO[®] TA Cloning Kit* (Invitrogen). Este kit está basado en la utilización de un enzima topoisomerasa localizada sobre el sitio de clonación, que es la que cataliza la ligación, de forma que ésta puede realizarse en menos tiempo.

El vector pGEM-T[®] Easy forma parte del kit de clonación *pGEM-T[®] Easy Vector System* (Promega). Este kit contiene una ligasa específica que también permite acortar los tiempos de ligación.

El vector pSEAP2-Basic forma parte del kit *Great EscAPe SEAP Reporter System 3* (BD Bioscience), y los productos de interés se introdujeron en el sitio de policlonación, tras digerir con enzimas de restricción específicos.

En todos los experimentos de clonación se siguieron las recomendaciones del fabricante, con alguna modificación. A continuación se detallan los pasos realizados.

II.1.7.1 Ligación

- **Utilizando el *TOPO[®] TA Cloning Kit***

- Mezclar los siguientes componentes:

Producto de PCR	4.5 μ l
<i>Salt Solution</i> (proporcionada en el kit)	1 μ l
Vector pCR [®] 2.1-TOPO [®] (10 ng/ μ l)	0.5 μ l

- Incubar a temperatura ambiente durante 15'.
- Poner en hielo.

- **Utilizando el *pGEM-T[®] Easy Vector System***

- Mezclar los siguientes componentes:

Producto de PCR	4 μ l
Tampón 2X	5 μ l
Vector pGEM-T [®] Easy	0.5 μ l
T4-Ligasa (3 unidades Weiss/ μ l)	0.5 μ l

- Incubar 16 horas a 4°C.

- **Ligaciones con vector *pSEAP2-Basic***

Previa la reacción de ligación se realizaron dos pasos, linearización del vector y defosforilación del mismo:

- Linearización del vector:

La linearización del vector se llevó a cabo mediante la digestión doble con los enzimas *Asp718* y *EcoRI* (Roche). La reacción de digestión se muestra en la tabla M15. La reacción se incubó durante 16 horas a 37°C.

Componentes de la mezcla	(μ l)
DNA de midiprep (2356 ug/ml)	3
Tampón B 10X	5
<i>EcoRI</i>	5
<i>Asp718</i>	5
H ₂ O _d	32

Tabla M15. Reacción de digestión doble para linearizar el vector pSEAP-Basic

- Defosforilación del vector:

Los componentes de la mezcla de reacción para la defosforilación del vector se muestran en la tabla M16. La reacción se incubó durante 10' a 37°C y posteriormente se inactivó el enzima a 65°C durante 15'.

Material y Métodos

Componentes de la mezcla	(μ l)
pSEAP2 Basic digerido	μ l hasta llegar a concentración de 25 ng/ μ l
Tampón 10X SAP	1
SAP (fosfatasa alcalina con 1u/ μ l)	1
H ₂ O	6.9

Tabla M16. Reacción de defosforilación.

- Ligación:
 - Mezclar los siguientes componentes:

Inserto (*)	
Vector (25 ng/ μ l)	2 μ l
Tampón Ligasa T4 (10X)	2 μ l
T4-Ligasa (5 Weiss u/ μ l)	1 μ l
H ₂ O	Hasta 20 μ l

(*) Concentración calculada para que la relación molar vector:inserto sea de 1:3

- Incubar durante 16 horas a 19°C y durante 2 horas a 4°C

II.1.7.2 Transformación

El procedimiento de transformación fue ligeramente diferente en función del tipo de células utilizadas.

- ***E. coli* JM109**

(Estas células provenían del *pGEM-T[®] Easy Vector System* y sólo fueron utilizadas con el mismo)

- Añadir la mezcla de ligación a las células (50 μ l) , en hielo.
- Incubar durante 20', en hielo.
- Choque térmico 45'' a 42°C.
- Poner células en hielo durante 2'.
- Añadir 950 μ l de medio SOC estéril.
- Incubar durante 90' a 37°C en agitación (150 rpm.)
- Sembrar en placas con ampicilina, X-Gal y IPTG.
- Incubar 16-20 horas a 37°C.

- ***E. coli* TOP10**

(Estas células eran proporcionadas en el kit *TOPO[®] TA Cloning Kit* y sólo fueron utilizadas con el mismo).

- Añadir la mezcla de ligación a las células (50 μ l), en hielo.
- Incubar durante 30', en hielo.
- Choque térmico 45'' a 42°C.
- Poner células en hielo durante 2'.
- Añadir 250 μ l de medio SOC estéril (proporcionado con el *TOPO[®] TA Cloning Kit*)
- Incubar durante 1 hora a 37°C en agitación (200 rpm aprox.)
- Sembrar en placas con ampicilina o kanamicina, X-Gal y IPTG.
- Incubar 16-20 horas a 37°C.

▪ ***E. coli* XL-1Blue**

(Estas células procedían del stock del Departamento de Genética, y fueron preparadas por la técnico de laboratorio Carmen Martínez. Fueron utilizadas con clonaciones de los tres tipos de vector mencionados).

- Añadir la mezcla de ligación a las células (100µl), en hielo.
- Incubar durante 30', en hielo.
- Choque térmico 90" a 42°C.
- Poner células en hielo durante 2'.
- Añadir 400 µl de LB estéril.
- Incubar durante 1 hora a 37°C en agitación (200 rpm aprox.)
- Sembrar en placas con ampicilina o kanamicina, en función del marcador de resistencia del vector utilizado y con IPTG y X-Gal en los casos necesarios.
- Incubar 16-20 horas a 37°C.

Medios y aditivos utilizados en el cultivo de bacterias:

Medios	Componentes
LB líquido (1 L)	Triptona 10 g Extracto de levadura 5 g NaCl 10 g H ₂ O hasta 1 L Esterilizar en autoclave
LB agar (1 L)	Triptona 10 g Extracto de levadura 5 g NaCl 10 g Agar 15 g H ₂ O hasta 1 L Esterilizar en autoclave
LB Broth (Sigma Aldrich) (1 L)	20 g H ₂ O hasta 1 L Esterilizar en autoclave
LB Agar (Sigma Aldrich) (1 L)	35 g H ₂ O hasta 1 L Esterilizar en autoclave
SOC (100 ml)	Triptona 2 g Extracto de levadura 0.5 g NaCl 1M 1ml KCl 1M 0.25 ml H ₂ O Hasta 1 L Esterilizar en autoclave y añadir: Mg ²⁺ 2M 1 ml glucosa 2M 1 ml (esterilizados por filtración)

Tabla M17. Medios utilizados para el cultivo de bacterias.

Antibióticos y otros aditivos	Concentración
Ampicilina	0.8 ml del stock a 100 mg/ml por 1 L de medio
Kanamicina	0.5 ml del stock a 30 mg/ml por 1 L de medio
X-Gal+IPTG	X-Gal (20 mg/ml en dimetil-formamida) 40 µl IPTG (50 mM) 2 µl H ₂ O estéril 58 µl Añadir sobre placa 30' antes de la siembra y dejar secar.

Tabla M18. Antibióticos y otros aditivos para el cultivo de bacterias

II.2 Genotipación

II.2.1 Selección de los SNPs

Se seleccionaron SNPs para los genes *FOXP2* y *HAR1A*, todos ellos presentes en la base de datos dbSNP, siguiendo criterios ligeramente diferentes para ambos.

▪ Gen *FOXP2*

Puesto que no había ningún polimorfismo descrito en los exones con excepción de la mutación en el exón 14 previamente mencionada en la introducción, se siguieron inicialmente las siguientes pautas de selección:

- SNPs utilizados en otros estudios de asociación, localizados en los intrones.
- SNPs de la región promotora.

En este caso, debido al elevado número de polimorfismos disponibles en las bases de datos, se tuvieron en cuenta las siguientes características, sin que se considerara necesario que todos los SNPs cumplieran las tres al mismo tiempo:

- Posibilidad de analizar el polimorfismo mediante RFLPs.
- Existencia de una validación del polimorfismo.
- Datos de frecuencia.

Excepcionalmente, uno de los SNPs estudiados se seleccionó sin que cumpliera el criterio de poder ser analizado mediante RFLPs. En concreto, el SNP rs6961558. Debido a su especial localización cercana al primer exón del gen y en el interior de una isla GC, se decidió abordar la genotipación de este SNP modificando una diana de restricción, como se explicará en el apartado II.2.3.1.

Posteriormente, se amplió la selección de SNPs, para cubrir completamente la longitud del gen, mediante la utilización de la aplicación *Tagger* del programa Haploview 4.0 (Barrett *et al.*, 2005). En la tabla M19 aparece reflejado el origen de cada SNP, junto con el método de genotipación utilizado.

▪ Gen *HAR1A*

Se seleccionaron SNPs que cubrieran la región donde se encuentra el gen, dando prioridad a aquellos que aparecieran como validados en las bases de datos y hubiera datos de frecuencia. En la tabla M20 aparecen los SNPs estudiados, referencias y método utilizado para cada uno.

II.2.2 Métodos de genotipación utilizados

En el presente estudio se han utilizado dos métodos principales de genotipación. El primero de ellos, genotipación por RFLPs (Polimorfismo en la Longitud de los Fragmentos de Restricción), fue llevado a cabo enteramente en el laboratorio de Genética Molecular de la Universidad de Valencia. El segundo, consistió en un ensayo iPLEX™ utilizando la tecnología MassArray de Sequenom® (Oeth *et al.*, 2005), y fue llevado a cabo en la plataforma de genotipación de Santiago de Compostela de los servicios del CEGEN (Centro Español de Genotipado).

SNP	Localización	Referencias	Método genotipación
rs6466478	Región 5'	Este estudio	RFLPs
rs7803667	Región 5'	Este estudio	RFLPs
rs10447760	Región 5'	Este estudio	RFLPs
rs13308496	Región 5'	Este estudio	iPLEX™-MassArray
rs7784307	Región 5'	Este estudio	RFLPs
rs10276237	Intrón s1	Este estudio	RFLPs
rs6961558	Intrón s1	Este estudio	RFLPs
rs10254225	Intrón s1	Este estudio	iPLEX™-MassArray
rs7798050	Intrón s1	Este estudio	RFLPs
rs923875	Intrón s1	Newbury <i>et al.</i> , 2002;	RFLPs
rs1597548	Intrón s1	Este estudio	RFLPs
rs10500038	Intrón s1	Este estudio	iPLEX™-MassArray
rs4730626	Intrón s1	Este estudio	iPLEX™-MassArray
rs717233	Intrón s2	Este estudio	iPLEX™-MassArray
rs12535428	Intrón s2	Este estudio	iPLEX™-MassArray
rs1668335	Intrón s2	Este estudio	iPLEX™-MassArray
rs11771168	Intrón s2	Este estudio	iPLEX™-MassArray
rs1916977	Intrón s3	Este estudio	iPLEX™-MassArray
rs2396722	Intrón s3	Este estudio	RFLPs
rs1358278	Intrón s3	Este estudio	RFLPs
rs2253478	Intrón s3	Este estudio	iPLEX™-MassArray
rs2049603	Intrón s3	Este estudio	iPLEX™-MassArray
rs2694941	Intrón s3	Este estudio	iPLEX™-MassArray
rs1852469	Intrón 1	Gong <i>et al.</i> , 2004;	RFLPs
rs10255943	Intrón 2	Este estudio	iPLEX™-MassArray
rs10486026	Intrón 2	Este estudio	iPLEX™-MassArray
rs2396753	Intrón 2b	Gong <i>et al.</i> , 2004;	RFLPs
rs17137124	Intrón 3a	Newbury <i>et al.</i> , 2002;	RFLPs
rs7799652	Intrón 8	Este estudio	iPLEX™-MassArray
rs1456029	Intrón 8	Marui <i>et al.</i> , 2005;	iPLEX™-MassArray
rs12670585	Intrón 8	Este estudio	iPLEX™-MassArray
rs1456031	Intrón 10	Gong <i>et al.</i> , 2004;	RFLPs
rs2396765	Intrón 16	Este estudio	iPLEX™-MassArray
rs1456021	Intrón 16	Este estudio	iPLEX™-MassArray

Tabla M19. Características de los SNPs del gen *FOXP2* estudiados. La denominación de los SNPs corresponde al identificador de la base del dbSNP.

SNP	Localización	Referencias	Método genotipación
rs6011613	Región 5'	Este estudio	iPLEX™-MassArray
rs2427496	Región 5'	Este estudio	iPLEX™-MassArray
rs6122371	Exón 1	Este estudio	iPLEX™-MassArray
rs750696	Exón 2	Este estudio	iPLEX™-MassArray
rs750697	Exón 2	Este estudio	iPLEX™-MassArray
rs2427498	Región 3'	Este estudio	iPLEX™-MassArray

Tabla M20. Características de los SNPs del gen *HAR1A* estudiados. La denominación de los SNPs corresponde al identificador de la base del dbSNP.

II.2.3 Genotipación por RFLPs

II.2.3.1 Diseño de cebadores

El método de genotipación por RFLPs está basado en la ausencia o presencia de un sitio de restricción, ocasionada por el cambio de nucleótido. Así, la determinación del genotipo se lleva a cabo en dos pasos, amplificación por PCR y digestión con un enzima de restricción.

De este modo, para cada polimorfismo se diseñó un par de oligonucleótidos que amplificara un fragmento que incluyera el SNP, y que al ser digerido permitiera la discriminación de los diferentes alelos en función del tamaño de banda obtenido.

Polimorfismo	Cebadores (5'-3')
rs6466478	UPS-P6F: ACCTCTTAGCAGTGATATGCTG UPS-P6R: TGTGTTTTCACATCCACTACAG
rs7803667	UPS-P5F: ACCCTTGATTTGACTTCGGTG UPS-P5R: TCGAGACACTGCTCTAGTACG
rs10447760	UPS-P13F: CTATTCGATCGCTGTTTGCC UPS-P13R: GTGGCACGTAGTTTGTTGATG
rs7784307	UPS-P3F: GGCTTTCTGCACTTCAAGTGG UPS-P3R: TGTTGGCAATGCGAACTGTAC
rs10276237	FP2P12F: GGTCGCGGTAAGTTTCTTGCC FP2P12R: CGGACGAAGCAACATTTTGC
rs6961558	FP2P8-F2: GAGCGCAGACACCTTTCGGTG FP2P8Rm2: CACTGAGGTCGGGTGTCAAGG
rs7798050	IS1-P7F: CTCAATACTTAGAACAGTGCC IS1-P7R: TACTTGGCTTGTCTTGCTG
rs923875	FOXP2-P5F: CTTGGGAAACTGAAGCCAG FOXP2-P5R: ACTCACCCAATATCATGCAATAG
rs1597548	IS1-P4F: CAATTCACCAACAGTTGTACC IS1-P4R: GATGGCTACATCTGCTTTGAC
rs2396722	5UPS-P1F: GTCATCAATGTCACAGAGAACTTG 5UPS-P1R: AGACTGTACTTGTTCCTGGGAG
rs1358278	5UPS-P1F: GTCATCAATGTCACAGAGAACTTG 5UPS-P1R: AGACTGTACTTGTTCCTGGGAG
rs1852469	FP2I1-P9F: GGCTACAGTTTACAAGACACCAGG FP2I1-P9R: GTCCAGCCTTTGGGAATTTGAC
rs2396753	I2B-P10F: TGGCTATGAAATAACAAGCACAAAC I2B-P10R: CCCAGTTATTGGCTCACTCTACC
rs17137124	FOXP2I3A-F: GGTTCTACAGCAGTATCATGG FOXP2I3A-R: TTATCTGCACCAATGGAAGG
rs1456031	I10-P11F: CAAAGTTATCAAGGCTGCGAGTC I10-P11R: CATCTTTTTCAATGCAAACCACTCA

Tabla M21. Cebadores para la genotipación por RFLPs de los polimorfismos del gen *FOXP2*.

Los cebadores se diseñaron para que cumplieran las siguientes características:

- Longitud de 19-30 nucleótidos.
- Contenido en GC 40-60%.

- Distribución uniforme de los cuatro nucleótidos a lo largo del cebador, evitando repeticiones de nucleótidos o complementariedad que pudiera dar lugar a estructuras secundarias.
- Tm similar, y no más diferente de 5°C para los pares de cebadores utilizados en una misma reacción.
- Presencia de G o C en el extremo 3' del primer, para favorecer la unión al DNA.
- Generación de fragmentos que permitieran llevar a cabo el análisis por RFLPs.

En el caso del polimorfismo rs6961558, mencionado anteriormente, se diseñó un par de cebadores diferente. Mientras que uno de los cebadores del par era completamente complementario a la secuencia de DNA genómico, en el otro, concretamente el reverso, se introdujo un cambio que permitiera introducir una secuencia diana para un enzima de restricción en el caso de uno de los dos alelos del SNP. De esta forma, después de amplificar sería posible discriminar los dos alelos mediante una digestión.

En la tabla M21 se muestran los oligonucleótidos utilizados para el análisis de los diferentes polimorfismos. En el caso del cebador reverso del polimorfismo rs6961558, se muestra resaltada en rojo la base modificada de la secuencia genómica.

II.2.3.2 Establecimiento de las condiciones de amplificación por PCR para cada polimorfismo

Para cada uno de los polimorfismos se pusieron a punto las condiciones de amplificación por PCR. En aquellos procedentes de bibliografía, se tomaron las condiciones descritas en ésta.

Como pauta general para cada uno se realizó una primera amplificación con las características siguientes, que sería modificada en caso de no obtenerse producto:

- Temperatura de asociación un par de grados por debajo de la Tm.
- Programa estándar de amplificación (ver tabla M8)
- Mezcla estándar de amplificación (ver tabla M4)

La puesta a punto de las condiciones de PCR se llevó a cabo con DNAs control y hasta que las condiciones no fueron óptimas no se comenzó a utilizar material de la muestra de estudio. Además, los productos obtenidos en la puesta a punto se secuenciaron para confirmar la especificidad de las reacciones (ver epígrafe II.1.6).

A continuación se muestran las condiciones de PCR ajustadas para cada polimorfismo.

SNP	Mezcla de reacción	Programa de reacción	Temperatura de asociación	Modificaciones	Tamaño (pb)
rs6466478	Mezcla estándar	Programa estándar	59°C	-	658
rs7803667	Mezcla estándar	Programa estándar	61°C	-	641
rs10447760	Mezcla estándar	Programa estándar	56°C	-	248
rs7784307	Mezcla estándar	Programa estándar	50°C	35 ciclos	581
rs10276237	Mezcla con GC-Rich	Programa promotor 4	61°C	Desnaturalización 1'30" en cada ciclo	347

Tabla M22. Condiciones de las reacciones de PCR para cada polimorfismo.

- se refiere a ausencia de modificaciones respecto al programa correspondiente.

Material y Métodos

SNP	Mezcla de reacción	Programa de reacción	Temperatura de asociación	Modificaciones	Tamaño (pb)
rs6961558	Mezcla con GC-Rich	Programa promotor 4	56°C	-	263
rs7798050	Mezcla estándar	Programa estándar	61°C	-	532
rs923875	Mezcla estándar	Programa estándar	61°C	-	625
rs1597548	Mezcla estándar	Programa estándar	59°C	-	410
rs2396722	Mezcla estándar	Programa estándar	62°C	35 ciclos	830
rs1358278	Mezcla estándar	Programa estándar	62°C	35 ciclos	830
rs1852469	Mezcla estándar	Programa SNP-60	60°C	-	299
rs2396753	Mezcla estándar	Programa SNP-60	60°C	-	251
rs17137124	Mezcla estándar	Programa estándar	57°C	35 ciclos	470
rs1456031	Mezcla estándar	Programa SNP-60	60°C	-	241

Tabla M22. Continuación.

II.2.3.3 Digestión de los productos de PCR

Una vez obtenido el producto de PCR, se digirió parte del mismo con enzimas de restricción para determinar el genotipo de cada muestra. En la tabla M23 se muestran las condiciones generales de digestión. Normalmente, el volumen de PCR utilizado era de 10 μ l, aunque si la concentración de DNA obtenido en la amplificación era baja el volumen se aumentaba hasta 15 μ l y si, por el contrario era alta, se disminuía a 6 μ l. Todas las incubaciones fueron realizadas a la temperatura recomendada por el fabricante para cada uno de los enzimas utilizados, durante un periodo superior a 6 horas.

Componentes de la mezcla	Mezcla estándar (μ l)
DNA	6-15
Tampón 10X	2
Enzima	0.2-0.5
H ₂ O	Hasta 20

Tabla M23. Mezclas de digestión en la genotipación por RFLPs

En la siguiente tabla aparecen mostrados para cada polimorfismo, el enzima de restricción utilizado, el tamaño de los fragmentos generados al digerir el producto de PCR correspondiente para cada uno de los alelos y la concentración de agarosa (o en su caso, acrilamida) de los geles utilizados para discriminar entre los diferentes alelos.

SNP	Enzima de restricción	Temperatura de digestión	Tamaño fragmentos de PCR (pb)	Tamaño de los fragmentos de restricción (pb) ^a	%Gel de agarosa
rs6466478	<i>TruI</i>	65°C	658	Alelo G: 207+64+132+165+85+5 Alelo A: 207+41+23+132+165+85+5	Gel de acrilamida 12%
rs7803667	<i>TasI</i>	65°C	641	Alelo A: 42+156+18+53+37+201+105+6+23 Alelo T: 42+71+85+18+53+37+201+105+6+23	Gel de acrilamida 12%
rs10447760	<i>HpaII</i>	37°C	248	Alelo T: 50+198 Alelo C: 50+62+136	2
rs7784307	<i>AcI</i>	37°C	581	Alelo A: 33+24+20+20+148+28+114+194 Alelo C: 33+24+20+20+148+28+46+68+194	Gel de acrilamida 12%
rs10276237	<i>Hir6I</i>	37°C	347	Alelo T: 2+30+117+198 Alelo G: 2+30+75+42+198	3
rs6961558	<i>TaaI</i>	65°C	263	Alelo G: 62+201 Alelo A: 62+180+21	3
rs7798050	<i>Alw26I</i>	37°C	532	Alelo A: 532 Alelo G: 305+227	2
rs923875	<i>Alw44I</i>	37°C	625	Alelo A: 625 Alelo C: 237+388	2
rs1597548	<i>AccI</i>	37°C	410	Alelo C: 410 Alelo A: 179+231	2
rs2396722	<i>VspI</i>	37°C	830	Alelo C: 315+515 Alelo T: 315+443+72	2.5
rs1358278	<i>XapI</i>	37°C	830	Alelo G: 361+102+116+251 Alelo A: 361+102+116+133+118	3
rs1852469	<i>TruI</i>	65°C	299	Alelo T: 84+215 Alelo A: 84+124+91	2
rs2396753	<i>Eam1104I</i>	37°C	251	Alelo C: 251 Alelo A: 125+126	2
rs17137124	<i>BspTI</i>	37°C	470	Alelo C: 470 Alelo T: 240+230	2
rs1456031	<i>Cac8I</i>	37°C	241	Alelo C: 241 Alelo T: 149+92	2

Tabla M24. Enzimas de restricción utilizadas para la genotipación por RFLPs, con temperatura de digestión, tamaño de los fragmentos de PCR, fragmentos de restricción generados para cada alelo, y concentración de agarosa utilizada para discriminar entre ambos.

^a En los casos en los que al digerir con enzimas de restricción se genera más de dos fragmentos, se muestran sombreados los fragmentos que permiten diferenciar los distintos alelos.

Todos los enzimas fueron de la casa comercial Fermentas, con excepción de *AcI* y *Cac8I* que eran de NEB, y *AccI* que fue de Roche.

Material y Métodos

Una vez digeridas todas las muestras, se repitieron al azar 46 controles y 46 pacientes, para confirmar la replicabilidad de la genotipación y que tanto PCR como digestión se habían llevado a cabo correctamente.

II.3 Búsqueda de mutaciones descritas en región codificante

Se buscaron las mutaciones en región codificante del gen *FOXP2* descritas hasta el momento de la realización de este trabajo: la mutación R553H en el exón 14, encontrada en la familia KE (Lai *et al*, 2001); un cambio no sinónimo en el exón 2 y la mutación R328X en el exón 7 encontradas en pacientes con dispraxia verbal (MacDermot *et al*, 2005). La primera de ellas se buscó tanto en las muestras de pacientes y controles del estudio de asociación como en los niños con trastorno específico del lenguaje. Las mutaciones en el exón 2 y exón 7 se buscaron únicamente en la muestra de niños con TEL.

Esta búsqueda se llevó a cabo de dos formas: en el caso de la mutación encontrada en la familia KE se realizó por RFLPs, mientras que en el caso de los dos cambios encontrados en los exones 2 y 7 se amplificó la región por PCR, se purificaron las reacciones y se secuenciaron.

Los cebadores utilizados para la amplificación de los 3 fragmentos se muestran en la tabla M25 y las condiciones de las reacciones de PCR en la tabla M26.

Polimorfismo	Cebadores (5'-3')
Mutación familia KE	FOXP2E14-F: TGCTTGATCAGGGACACAAC FOXP2E14-F: CCACGAGAATGTTAGCATGC
Cambio en exón 2	FP2-E2F: GATTAGCATCTGGATGAAAGC FP2-E2R: AAGTATCACGTTCTGCAGCAC
Cambio en exón 7	FP2-E7F: ACCTTGTTATGCTAGTGAAGC FP2-E2R: ATAGAGACCATCCTGGCTAAC

Tabla M25. Cebadores utilizados en la detección de mutaciones conocidas en el gen *FOXP2*.

Fragmento analizado	Mezcla de reacción	Programa de reacción	Temperatura de asociación	Tamaño (pb)
Mutación familia KE	Estándar	Estándar	59°C	584
Cambio en exón 2	Estándar	Estándar	59°C	681
Cambio en exón 7	Estándar	Estándar	61°C	664

Tabla M26. Condiciones de las reacciones de amplificación por PCR para la detección de mutaciones conocidas en el gen *FOXP2*.

La digestión del producto de PCR obtenido para el fragmento en el que se encuentra la mutación descrita en la familia KE con el enzima de restricción *TalI*, da lugar a dos fragmentos de 228 y 356 pb en el alelo normal, mientras que en presencia de la mutación (G/A) el sitio de restricción desaparece.

II.4 Búsqueda de variación en el número de repeticiones de trinucleótidos mediante electroforesis capilar

Dentro del análisis de variaciones en el gen *FOXP2*, se incluyó el análisis de las variaciones en el número de repeticiones de los dos tramos de poliglutaminas localizados en los exones 5 y 6, por medio de análisis del tamaño de fragmentos utilizando electroforesis capilar. También se aplicó este método para determinar si había alguna expansión en un microsatélite localizado en la región promotora y para el que se detectó una delección de tres trinucleótidos en uno de los pacientes durante la genotipación.

El análisis del tamaño fragmentos se llevó a cabo en los siguientes pasos:

- Diseño de cebadores. En la tabla M27 se muestran los cebadores diseñados para la amplificación por PCR de la región de interés, y el tipo de fluorocromo utilizado en el marcaje.

Fragmento analizado	Cebadores (5'-3')	Fluorocromo
Tramo 40 repeticiones de glutaminas	PG40F: GCAAGAGCAGTTACATCTTC PG40R2: ATGAGATAACCGGATCCTAC	6-FAM -
Tramo de 10 repeticiones de glutaminas	PG10F: CTCTAGACCTTGCTCCATAC PG10R: TGTTGCATCTGGAGAAGCTG	HEX -
Región CGG en promotor	FP2P8-F2: GAGCGCAGACACCTTTTCGGTG FP2P8-R2: TGCGGAGCGTCCCAAGCGGTG	6-FAM -

Tabla M27. Cebadores utilizados para la amplificación por PCR de fragmentos de DNA analizados en electroforesis capilar.

- Reacción de PCR de la región de interés.
En la tabla M28 se muestran las condiciones de PCR aplicadas en cada caso.

Fragmento analizado	Mezcla de reacción	Programa de reacción	Temperatura de asociación	Tamaño (pb)
Tramo 40 repeticiones de glutaminas	Mezcla con DMSO	Programa estándar	58°C	213
Tramo de 10 repeticiones de glutaminas	Mezcla estándar	Poliglutaminas10	60.5°C	320
Región CGG en promotor	Mezcla con glicerol 50%	<i>GC-Rich</i>	61°C	517

Tabla M28. Condiciones de las reacciones de PCR utilizadas para amplificar los fragmentos de DNA analizados en electroforesis capilar.

- Comprobación de la amplificación por PCR en gel de agarosa.
- Separación de los posibles fragmentos en electroforesis capilar. Este paso y el siguiente se llevaron a cabo por el Servicio de Secuenciación de la Universitat de València.
- Lectura de resultados mediante el software *GeneScan-v3.7*.

II.5 Análisis de la secuencia de la región promotora del gen *FOXP2* en niños con Trastorno Específico del Lenguaje.

El análisis de la secuencia de la región promotora del gen *FOXP2* en niños con trastorno específico del lenguaje (TEL), se llevó a cabo mediante amplificación por PCR y secuenciación de los fragmentos obtenidos hasta formar un *contig* de la zona de interés. La estrategia utilizada ha sido el paseo cromosómico mediante PCR, basada en el diseño de cebadores que amplifiquen fragmentos solapantes que puedan ensamblarse para dar lugar a un *contig* que incluya toda la región que se quiere estudiar.

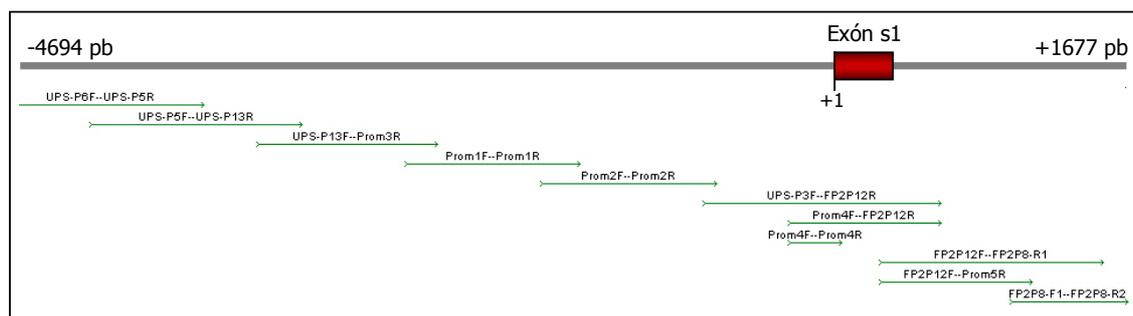


Figura M1. Estrategia de secuenciación de la región promotora del gen *FOXP2*, basándose en paseo cromosómico mediante PCR.

La región secuenciada cubre 6371 pares de bases, entre -4694 y +1677, tomando como sitio +1 la primera base del exón s1. En la figura M1 se muestra la estrategia inicial de fragmentos a secuenciar. En la tabla M29 se incluyen todos los cebadores utilizados específicamente para amplificar y secuenciar los fragmentos del promotor. En la medida de lo posible, se intentó utilizar cebadores ya disponibles procedentes del estudio de asociación (ver tabla M21). En ocasiones, además o en sustitución de los cebadores utilizados para la amplificación se secuenciaron los productos con cebadores internos.

Nombre	Secuencia (5'-3')
Prom1F	ACTTACTCGGTACACCTGGAG
Prom1R	CTTTGCTCGCAAGTCAGTCGC
Prom2F	GCACAAGTTAGTGCGGTTTCG
Prom2R	AAGTACGTCCTCTCAGCAGAC
Prom3R	GAGTAAACGACGTGGAAATCC
Prom4F	GGTACTTCAGAATCCTCTGC
Prom4R	GCTAGAGCTAGGGCGAAGCG
FP2P8-F1	GTCCACGCGTTAGCACGACCG
FP2P8-R2	TGCGGAGCGTCCCAAGCGGTG
FP2P8-R1	GCGCCAGGTGCCCTAAGAGAC
Prom5R	CCGCGAGGTGCGATTTTAC
Prom12F	GGCAACAATGAAGAGAAAGACAC
Prom13R	AACGCGATCAAGGCAGCAGTC

Tabla M29. Cebadores empleados específicamente en la amplificación de la región promotora del gen *FOXP2*.

Las condiciones de las reacciones de PCR llevadas a cabo se muestran en la tabla M30. En cuanto a las reacciones de secuenciación, éstas se llevaron a cabo según lo detallado en el epígrafe II.1.6.

Cebadores de las reacciones de PCR	Mezcla de reacción de PCR	Programa de PCR	Temperatura de asociación	Modificaciones en el número de ciclos	Tamaño amplificado (pb)	Cebadores para secuenciar los productos de PCR
UPS-P6F UPS-P5R	Mezcla estándar	Programa estándar	57°C	35 ciclos	1159	UPS-P6F, UPS-P5R
UPS-P5F UPS-P13R	Mezcla estándar	Programa estándar	56°C	35 ciclos	1193	UPS-P5F, UPS-P13R
UPS-P13F Prom3R	Mezcla estándar	Programa promotor 1	60.5°C	-	1012	UPS-P13F, Prom3R
Prom1F Prom1R	Mezcla estándar	Programa promotor 1	63°C	-	975	Prom1F, Prom1R
Prom2F Prom2R	Mezcla estándar	Programa promotor 1	62°C	-	989	Prom2F, Prom2R
UPS-P3F FP2P12R	Mezcla estándar	Programa promotor 3	62°C	-	1339	UPS-P3F, FP2P12R, Prom4F
Prom4F Prom4R	Mezcla estándar	Programa promotor 2	50.5°C	-	296	Prom4F, Prom4R
Prom4F FP2P12R	Mezcla estándar	Programa promotor 1	62°C	-	858	Prom4F, FP2P12R, FP2P12F
FP2P12F Prom5R	Mezcla con CG-Rich	Programa promotor 6	60°C	-	859	FP2P12F, Prom5R, FP2P12R
FP2P12F FP2P8-R1	Mezcla con CG-Rich	Programa promotor 5	58°C	-	1259	FP2P12F, FP2P8-R1, Prom5R
FP2P8-F1 FP2P8-R2	Mezcla con CG-Rich	Programa promotor 4	55°C	-	659	FP2P8-F1, FP2P8-R2

Tabla M30. Condiciones de las reacciones de PCR utilizadas para la obtención de la secuencia promotora del gen *FOXP2* en niños con TEL.

En el caso del fragmento amplificado con los cebadores Prom4F y Prom4R, éste comprende un polimorfismo consistente en un indel de la repetición ACAC (rs10539256). En las muestras en las que se observó que no eran homocigóticas para la presencia o ausencia de la repetición, el fragmento de PCR obtenido se clonó utilizando el *TOPO-TA Cloning Kit* (ver epígrafe II.1.7) y se determinó el genotipo tras llevar a cabo PCR de colonia de 4-6 clones con los cebadores del vector, M13R y M13F (Mezcla estándar de reacción para PCR de colonia y Programa estándar con temperatura de asociación 51°C) y secuenciar.

El análisis de las secuencias obtenidas se llevó a cabo con el programa *Sequencher* de forma que se obtuvo un *contig* para cada uno de los individuos analizados.

II.6 Análisis de la región promotora del gen *FOXP2* en primates

La estrategia de secuenciación para analizar la región promotora del gen *FOXP2* en primates fue la misma que fue utilizada para el análisis de esta región en niños con TEL, el paseo cromosómico mediante PCR.

Las muestras de DNA utilizadas procedían del *Coriel Institute for Medical Research* (tabla M3) o habían sido donadas por Jaume Bertranpetit.

Además, para el análisis posterior de las secuencias, también se utilizaron secuencias procedentes de base de datos, recogidas en la tabla siguiente:

Especie	Número de acceso en base de datos	Fragmento extraído	Base de datos
<i>Homo sapiens</i>	NT_007933.14	38905318-38911575	NCBI
<i>Pan troglodytes</i>	AC145868.2	190293-184036	NCBI
<i>Macaca mulatta</i>	Jan 2006 Assembly	Chr3:151,448,158-151,448,457	UCSC Genome browser
<i>Mus musculus</i>	AC150746.2	186000-195000	NCBI
<i>Rattus norvegicus</i>	AC106374.6	32000-40000	NCBI

Tabla M31. Secuencias de primates obtenidas de bases de datos.

Para llevar a cabo el estudio se utilizaron dos tipos de cebadores:

- Los mismos cebadores que los utilizados para el análisis del promotor en niños con TEL
- Cebadores específicos de las distintas especies de primates diseñados a partir de las secuencias obtenidas de las mismas (tabla M32).

Nombre	Secuencia (5'-3')	Especie
P13F-P5	CTATTCGATTGCTGTTTGTC	<i>Saguinus labiatus</i>
Prom7F	TGTGTACAGAAAAGTGAAAGGC	<i>Macaca sp., S. labiatus, L. lagotricha, E. patas, G. gorila</i>
Prom8F	GGGTGCCAGGGCTGCTTCCTC	<i>Macaca sp., S. labiatus, L. lagotricha, E. patas, G. gorila</i>
Prom10F	GGCGCCTCGGCTTTGAAGTGC	<i>Macaca sp., S. labiatus, L. lagotricha, E. patas, G. gorila</i>
Prom11R	CGGGCTTCCTCGTGGTCGCG	<i>Macaca sp., S. labiatus, L. lagotricha, E. patas, G. gorila</i>

Tabla M32. Cebadores específicos de primates utilizados para la amplificación de la región promotora del gen *FOXP2*.

Las condiciones de las reacciones de PCR en las que se utilizaron son las siguientes:

Cebadores	Mezcla de PCR	Programa de PCR	Temperatura de asociación	Modificaciones al programa	Tamaño del producto (pb)
Prom8F Prom11R	Mezcla estándar	Programa promotor 7	60°C	Desnaturalización inicial 10'	857
Prom7F Prom3R	Mezcla estándar	Programa promotor 6	54°C	Desnaturalización inicial 5'	308
Prom10F UPS-P3R	Mezcla <i>GC-Rich</i>	Programa promotor 7	53°C	-	976
P13F-P5 Prom3R	Mezcla estándar*	Programa promotor 1	54°C	-	1012

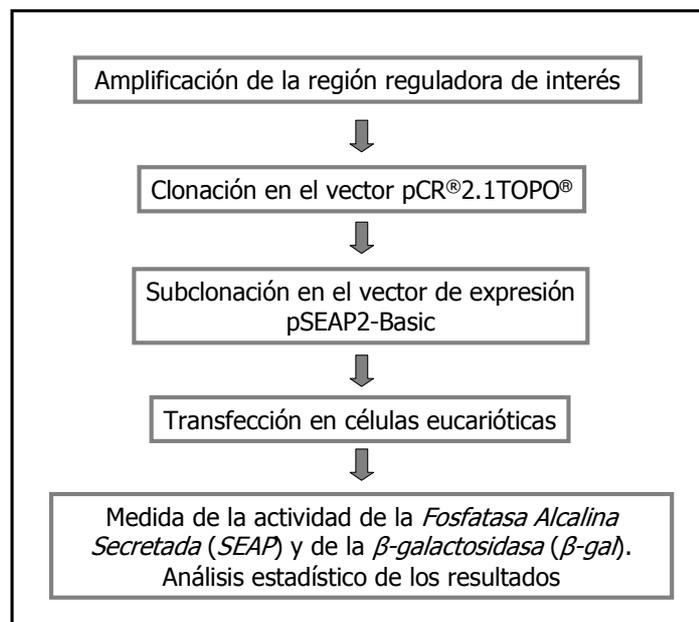
Tabla M33. Condiciones de las reacciones de PCR específicas de primates.

*Utilización de DNA polimerasa con 5 u/μl (Roche).

Al igual que en la amplificación de la región promotora de niños con TEL, los fragmentos correspondientes a los productos obtenidos con los cebadores Prom4F y Prom4R comprenden un polimorfismo consistente en un indel de la repetición ACAC (rs10539256). En los primates en los que se observó que no eran homocigotos para la presencia o ausencia de esta repetición se procedió igual que en el caso de los niños con TEL: el fragmento de PCR obtenido se clonó en el vector del *TOPO-TA Cloning Kit* (ver epígrafe del Material y Métodos II.1.7) y se determinó el genotipo tras llevar a cabo PCR de colonia de 4-6 clones con los cebadores del vector, M13R y M13F utilizando la mezcla estándar de reacción para PCR de colonia, y el programa estándar con temperatura de asociación de 51°C y secuenciar.

II.7 Análisis funcional de la región promotora

El plan de trabajo para llevar a cabo el análisis funcional se muestra en el siguiente diagrama:



II.7.1 Regiones reguladoras analizadas

Se seleccionaron tres fragmentos a analizar, en base a los resultados obtenidos a partir del análisis de la región promotora en primates y la información sobre el grado de conservación de la base de datos del *UCSC Genome Browser*.

En la figura M2 se muestra las posiciones de los fragmentos analizados, tomando como sitio +1 la primera base del exón s1.

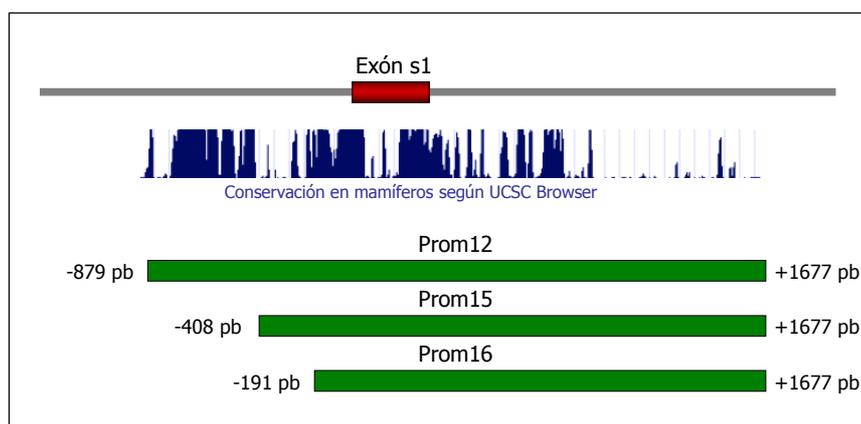


Figura M2. Esquema de las potenciales regiones reguladoras analizadas. Se muestran las tres regiones analizadas en el contexto del exón s1, así como el diagrama de conservación de la región en mamíferos obtenido del *UCSC Genome Browser*.

II.7.2 Amplificación de los fragmentos de interés de la región reguladora

La amplificación de los tres fragmentos a analizar se llevó a cabo en dos pasos. En primer lugar se amplificó fragmento de mayor tamaño, al que se denominó Prom12, y una vez clonado y subclonado se amplificaron los otros tomando a éste como molde. Con este paso se aseguró que la secuencia compartida de los tres fuera la misma, y que no hubiera polimorfismos en las secuencias.

Los cebadores utilizados para la amplificación por PCR de los fragmentos se muestran en la tabla M34. Estos cebadores llevaban en los extremos la secuencia correspondiente a las dianas de restricción del sitio de clonación múltiple donde se pretendía clonar el fragmento.

Fragmento	Cebadores (5'-3')	Tamaño
Prom12	KpnI-Prom12F: ggtaccGGCAACAATGAAGAGAAAGACAC FP2P8R2-EcoRI: gaattcTGCGGAGCGTCCCAAGCGGTG	2556
Prom15	KpnI-Prom15F: ggtaccCTGCAAATCCGAACAATATTCTTCCC FP2P8R2-EcoRI: gaattcTGCGGAGCGTCCCAAGCGGTG	2085
Prom16	KpnI-Proim16F: ggtaccCCCTCCGAGGAGAGGCAGCCC FP2P8R2-EcoRI: gaattcTGCGGAGCGTCCCAAGCGGTG	1868

Tabla M34. Cebadores utilizados para la amplificación de los fragmentos de interés del ensayo funcional.

La amplificación de los fragmentos se llevó a cabo utilizando la polimerasa con prueba de lectura *Pwo SuperYield DNA Polymerase* (Roche), con las condiciones de reacción estándar para esta polimerasa (Tabla M6 del apartado II.1.3) y el programa de PCR diseñado para fragmentos largos (Tabla M7 del mismo apartado). Una vez amplificado se llevó a cabo extracción de banda, debido a que el enzima utilizado en la amplificación, *Pwo*, tiene actividad exonucleasa 5-3' que podría competir con la posterior extensión de adeninas y además, en el caso del fragmento Prom12, porque de la amplificación se obtenían bandas no específicas.

II.7.3 Clonación en el vector pCR®2.1TOPO®

Previa a la clonación, y puesto que la DNA polimerasa utilizada en la amplificación por PCR no añade 1 adenina al extremo del fragmento amplificado, necesaria para la clonación en vectores TOPO-TA, se realizó una extensión con dATPs de los fragmentos de PCR purificados. Los componentes de la reacción se muestran en la tabla M35. El tiempo de incubación de la reacción fue de 15 minutos a 72°C.

Componentes de la mezcla	Mezcla estándar (µl)
DNA de la banda	2
Tampón 10X con MgCl ₂	1.25
dATP (10 mM)	0.5
<i>Netzyme</i> (1 u/µl)	1
H ₂ O _d	8

Tabla M35. Componentes de la mezcla de reacción para la extensión con adeninas de fragmentos de DNA.

La clonación de los fragmentos de interés en el vector pCR®2.1TOPO® se llevó a cabo siguiendo las recomendaciones de la casa comercial, con algunas modificaciones, como se indica en el apartado II.1.7.

Una vez incubadas las placas sembradas, se procedió a seleccionar las colonias con inserto. A continuación, se extrajo el DNA plasmídico y se testó la presencia del inserto mediante digestión con *EcoRI*. La reacción de digestión para ello es la siguiente:

Componente	Cantidad
DNA plasmídico	1 µl
Tampón <i>EcoRI</i>	2 µl
<i>EcoRI</i>	1 µl
H ₂ O _d	16 µl

Tabla M36. Condiciones digestión para comprobar presencia de inserto en vector pCR®2.1TOPO®.

Para confirmar la ausencia de mutaciones en el inserto, éste se secuenció al completo con cebadores del vector y cebadores internos (T7, M13R, Prom4F, FP2P12F, FP2P12R, Prom3R, FP2P8-R2).

II.7.4 Subclonación en pSEAP2-Basic

Una vez confirmada la secuencia de los insertos, se llevó a cabo una digestión doble del DNA plasmídico de los clones seleccionados, con el fin de liberar fragmentos flanqueados por extremos *Asp*718 (isosquizómero de *Kpn*I) y *Eco*RI, compatibles con el sitio donde se pretendía subclonar en el vector pSEAP-Basic.

Las digestiones se corrieron en geles de agarosa y se extrajo el DNA de las bandas correspondientes a los insertos, según lo especificado en el apartado II.1.5.3.

Componentes de la mezcla	Mezcla estándar (μl)
DNA de miniprep	15
Tampón B 10X	5
<i>Eco</i> RI	5
<i>Asp</i> 718	5
H ₂ O	20

Tabla M37. Componentes de la reacción de digestión para liberar los fragmentos de interés del vector PCR[®]2.1TOPO[®].

La subclonación en el vector pSEAP-Basic se llevó a cabo según lo especificado en el epígrafe II.1.7.1.

La amplificación del DNA de todos los vectores utilizados en la transfección se llevó a cabo mediante el kit *Genopure Plasmid Midi Kit* (Roche).

II.7.5 Vectores de expresión utilizados

En la figura M3 se muestran los vectores de expresión utilizados en los experimentos de transfección.

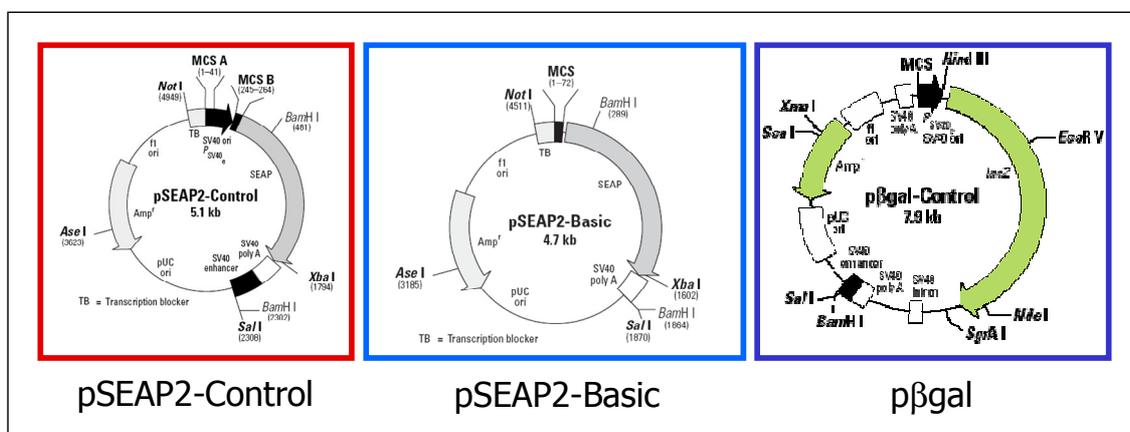


Figura M3. Vectores de expresión utilizados en los experimentos de transfección.

El vector **pSEAP2-Basic** se utilizó como control negativo puesto que carece de promotor de células eucarióticas, así como de secuencias activadoras. En el sitio de clonación múltiple (MCS), localizado en 5' respecto al gen de la Fosfatasa Alcalina Secretada, fue donde se clonaron los fragmentos cuya actividad promotora se pretendía evaluar.

El vector **pSEAP2-Control** se utilizó como control positivo, ya que consiste en el vector pSEAP2-Basic en el que se han incluido el promotor SV40, aguas arriba del gen que codifica para la Fosfatasa Alcalina Secretada y secuencias activadoras aguas abajo. El vector **pβgal** que contiene también el promotor de SV40, en este caso en 5' respecto del gen que codifica para la β-galactosidasa, se utilizó como control para normalizar todos los experimentos de transfección.

II.7.6 Transfección en células eucarióticas

La transfección de las construcciones obtenidas se llevó a cabo en dos tipos celulares: células CHO-K1, procedentes de ovario de hámster chino, ampliamente utilizadas para este tipo de estudios y células H23, procedentes de una cepa de células tumorales de pulmón humano.

Las células CHO-K1 se cultivaron en medio F12 Ham (Gibco) y las células H23 con medio D-MEM (*Dulbecco's modified Eagle's media*) con GlutaMAX™ I, 4500 mg/L D-Glucosa y Piruvato de Sodio (Gibco), en ambos casos con 10% de suero bovino fetal y 1% de penicilina/estreptomina como antibióticos.

En cada tipo celular se llevaron a cabo dos experimentos independientes de transfección, en los que se transfectaron 3 réplicas para cada una de siguientes construcciones:

- pSEAP2-Prom12 + pβgal
- pSEAP2-Prom15 + pβgal
- pSEAP2-Prom16 + pβgal
- pSEAP2-Basic + pβgal (control negativo de la actividad SEAP)
- pSEAP2-Control + pβgal (control positivo de la actividad SEAP)

El método de transfección utilizado fue mediante la utilización del agente *Fugene® HD* (Roche), siguiendo las recomendaciones del fabricante. Brevemente, el protocolo fue el siguiente:

- Realizar varios pases de las células desde su descongelación.
- Sembrar en placas de 6 pocillos 300.000-600.000 células (misma cantidad en todos los pocillos del ensayo) y esperar hasta que el cultivo alcance una confluencia de >80% (12-48 horas en función de la cepa utilizada).
- Diluir el DNA para transfectar en medio *Optimem* atemperado a temperatura ambiente. En nuestro caso se añadieron 1 μg de cada uno de los vectores a cotransfectar: pβgal y la construcción correspondiente.
- Añadir 4 μl de reactivo *Fugene® HD* a la mezcla anterior.
- Incubar a temperatura ambiente durante 15 minutos.
- Añadir la mezcla a las células.
- Incubar a 37°C con atmósfera de 5% de CO₂ durante 48 horas.

II.7.7 Medida de la actividad de la Fosfatasa Alcalina Secretada y de β -galactosidasa

II.7.7.1 Preparación de extractos para la medida de actividades de la Fosfatasa Alcalina Secretada y la β -galactosidasa

La preparación de extractos para medir la actividad de la Fosfatasa Alcalina Secretada (SEAP) se llevó a cabo siguiendo las recomendaciones del kit de detección *Great EscAPe SEAP Reporter System 3* (BD Bioscience).

En el caso de la obtención de extractos celulares para medir la actividad de la β -galactosidasa se siguieron dos protocolos diferentes en los dos tipos de células empleados. En el caso de las células CHO-K1 las células se tripsinizaron, se resuspendieron en PBS 1X y se siguió el protocolo recomendado para células en suspensión por la casa comercial del kit utilizado para medir la actividad de la β -galactosidasa (*β -Gal ELISA*, Roche). En el caso de las células H23, se siguió el protocolo recomendado para células adherentes.

II.7.7.2 Medida de la actividad de la Fosfatasa Alcalina Secretada y la β -galactosidasa

La medida de ambas actividades se llevó a cabo siguiendo las recomendaciones de la casa comercial correspondiente. La actividad de la β -galactosidasa se midió en el lector de placas Fluostar OPTIMA de BMG Labtech. La actividad de la SEAP se determinó por el método de luminiscencia y se midió en un luminómetro de placa (WALLAC Victor 1420 Multilabel HTS Counter).

Los valores de las concentraciones se obtuvieron a partir de curvas de calibración realizadas con Fosfatasa Alcalina Secretada y β -galactosidasa. Para cada una de las muestras obtenidas (tres repeticiones para cada construcción cotransfectada) se midieron dos réplicas y se tomó como valor final el valor medio de ambas.

Los valores finales de concentración de SEAP se normalizaron para cada muestra con los valores obtenidos para la β -galactosidasa.

II.8 Análisis de metilación

A través de las bases de datos *UCSC Genome Browser (Human Mar. 2006 Assembly)* y *Human Genome Segmental Duplication Database* se localizó una isla GC en la región 5' del gen *FOXP2*, solapante con el primer exón transcrito, el exón s1 (figura R1 de resultados).

Para determinar el grado de metilación de la isla GC encontrada, se llevaron a cabo dos aproximaciones, una más general mediante enzimas de restricción sensibles a metilación y otra más específica como es la secuenciación por bisulfito.

II.8.1 Análisis de metilación mediante la utilización de enzimas de restricción sensibles a metilación

La determinación del grado de metilación mediante enzimas de restricción está basada en la amplificación de fragmentos de DNA en los que se encuentran dianas para enzimas sensibles a metilación, antes y después de digerir el DNA con dichos enzimas. La mayoría de enzimas de restricción tipo II son sensibles a metilación, es decir, no reconocen el sustrato cuando está metilado. Por tanto, si la amplificación por PCR se produce de forma normal después de la digestión, será un indicativo de metilación y viceversa. Por el contrario, en el caso de un enzima que únicamente corta en presencia de metilación, la amplificación por PCR después de la digestión será indicativa de dianas metiladas.

Para valorar el estado general de metilación en diferentes áreas cerebrales y analizar posibles diferencias globales entre pacientes y controles, se prepararon mezclas de DNA de varios individuos. En la tabla M38, se muestra el número de DNAs de cada una de las mezclas, así como la región cerebral utilizada.

Región	Pacientes o controles	Nº de DNAs en la mezcla
Circunvolución temporal superior izquierda	Controles	4
	Pacientes	5
Circunvolución temporal superior derecha	Controles	4
	Pacientes	3
Circunvolución del cíngulo izquierda	Controles	4
	Pacientes	4
Circunvolución del cíngulo derecha	Controles	5
	Pacientes	3

Tabla M38. Muestras utilizadas para el análisis de metilación con enzimas de restricción.

El protocolo que se siguió fue el siguiente:

- Digestión con las enzimas de restricción *Hpa*II (sensible a metilación) o *Mcr*BC (que únicamente corta si la diana está metilada) de 100 ng de DNA en un volumen total de 50 μ L. Incubación de 16 horas a 37°C.
- Inactivación del enzima, 10 minutos a 70°C.
- PCR con cebadores de la región de interés del DNA digerido y del DNA sin digerir, que funciona como control positivo. En la tabla M39 se presentan los cebadores utilizados, y en la tabla M40 las condiciones de las reacciones de PCR así como, el tamaño del producto amplificado y el número de dianas de las enzimas empleadas.
- Comprobación y análisis de los fragmentos obtenidos en geles de agarosa.

Material y Métodos

Nombre	Secuencia 5'-3'
UPS-P3F UPS-P3R	GGCTTTCTGCACTTCAAGTGG GTTGTTGGCAATGCGAACTGTAC
Prom4F Prom4R	GGTACTTCAGAATCCTCTGC GCTAGAGCTAGGGCGAAGCG
FP2-P8F2 FP2-P8R2	GAGCGCAGACACCTTTCGGTG TGCGGAGCGTCCAAGCGGTG

Tabla M39. Cebadores utilizados en el análisis del grado de metilación mediante la utilización de enzimas de restricción sensibles a metilación.

Nombre	Mezcla de PCR	Programa de PCR	Temperatura de asociación	Tamaño del producto (pb)	Dianas <i>HpaII</i> (CCGG)	Dianas <i>McrBC</i> (PuCmG)
UPS-P3F UPS-P3R	Mezcla estándar	Programa cDNA1	57°C	581	1	12
Prom4F Prom4R	Mezcla estándar	Programa Promotor 2	50.5°C	296	3	16
FP2P8-F2 FP2P8-R2	Mezcla con glicerol 50%	Programa Promotor 3	56°C	517	2	34

Tabla M40. Condiciones de las reacciones de PCR realizadas en el análisis del grado de metilación mediante la utilización de enzimas de restricción sensibles a metilación.

En la figura M4 se muestra un esquema de la localización de los fragmentos analizados.

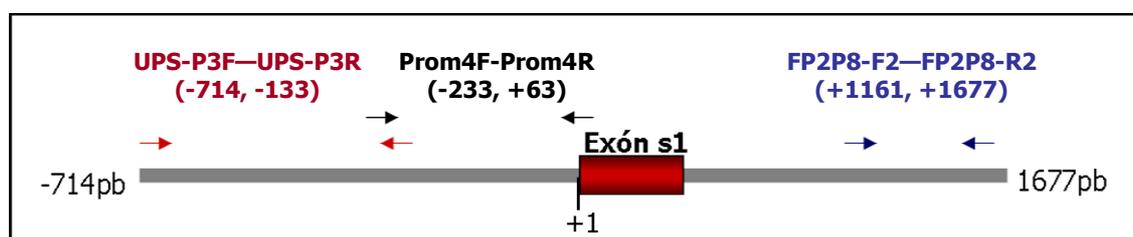


Figura M4. Localización de los fragmentos de PCR utilizados para analizar el estado de metilación mediante enzimas de restricción, respecto de la posición de la primera base del exón s1.

II.8.2 Análisis del estado de metilación mediante la técnica del bisulfito

La técnica del bisulfito permite mostrar las 5-metilcitosinas en cadenas individuales, siendo la metodología más exacta para monitorizar el estado de metilación (Frommer, 1992). El protocolo está basado en el tratamiento de DNA genómico con bisulfito, que convierte los residuos de citosina en uracilo, mientras que las 5-metilcitosinas permanecen sin modificar. En las subsecuentes amplificaciones por PCR, utilizando cebadores específicos para las cadenas individuales modificadas, todos los residuos de uracilo y timinas son amplificados como timinas, y únicamente la 5-metilcitosina es amplificada como citosina. El producto resultante puede ser secuenciado directamente o clonado en bacterias y secuenciado posteriormente. En la figura M5 se muestra una representación de las modificaciones en la composición de bases que tiene lugar durante el proceso.

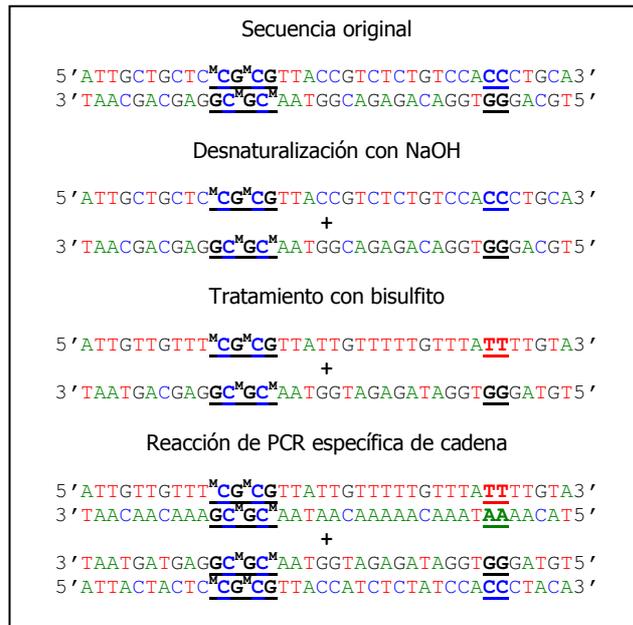


Figura M5. Descripción de los cambios en la composición de bases nucleóticas de una secuencia ejemplo de DNA a lo largo del tratamiento de modificación por bisulfito. Al ser modificadas, las dos cadenas de DNA dejan de ser complementarias y la amplificación por PCR se lleva a cabo de forma independiente en cada una de ellas. (Figura cedida por Dr. Dagnall, con modificaciones).

II.8.2.1 Muestras analizadas por medio de la modificación con bisulfito

Para llevar a cabo el análisis se partió de DNA genómico obtenido a partir de muestras de cerebro postmortem humano de diferentes regiones de la corteza cerebral, procedentes de pacientes esquizofrénicos e individuos control. Estas muestras habían sido donadas al centro SANE POWIC (*SANE Prince of Wales Internacional Centre*) de Oxford. En la tabla M41 aparecen reflejadas las muestras individuales utilizadas en este análisis.

Muestra	Localización	Hemisferio	Paciente/Control
E11	Circunvolución temporal superior	Izquierdo	Paciente
E21			Control
G17	Circunvolución del parahipocampo		Paciente
G29			Control
A12	Circunvolución del cíngulo		Paciente
A21			Control
F1	Circunvolución temporal superior	Derecho	Control
F18			Paciente
H17	Circunvolución del parahipocampo		Paciente
H28			Control
B5	Circunvolución del cíngulo		Control
B15			Paciente
<i>Homo sapiens</i>	Leucocitos		Control
<i>Pan troglodytes</i>	Cerebro		
<i>Gorilla gorilla</i>	Leucocitos		

Tabla M41. Muestras utilizadas en el análisis del estado de metilación por medio de la modificación con bisulfito.

II.8.2.2 Diseño de cebadores

Debido a la especial composición del DNA convertido (tratado con bisulfito) respecto al original, el diseño de cebadores específicos para la amplificación de DNA convertido con bisulfito se llevó a cabo siguiendo las siguientes consideraciones especiales:

- Tamaño de 25-30 bases.
- Temperaturas de asociación similares en cada pareja de cebadores.
- Temperatura de asociación del cebador directo mayor que la temperatura de asociación teórica del cebador reverso para DNA no convertido.
- Minimización de tramos repetitivos, especialmente Cs no CpGs y Ts.
- Extremos 3' finalizando en repetición de citosinas no CpG, para facilitar la discriminación entre DNA convertido por bisulfito y DNA no convertido.
- Evitar dinucleótidos CG dentro de la secuencia del cebador, puesto que se pierde información del estado de metilación en el sitio.
- El tamaño del producto amplificado debe reflejar un equilibrio entre maximizar el número de dinucleótidos CG amplificados y las limitaciones de tamaño debido a la naturaleza fragmentada del DNA tratado. Tamaño máximo de unas 300 pb.

Una vez convertido el DNA por medio del bisulfito, la composición de bases de las cadenas originales deja de ser complementaria. En nuestro trabajo únicamente se diseñaron cebadores para una de las cadenas originales, en concreto la cadena con sentido 5'-3'.

La localización de los fragmentos analizados respecto al primer exón del gen y la isla CG se muestran en la figura M6. Los oligonucleótidos utilizados, así como el tamaño del producto obtenido se muestran en la tabla M42.

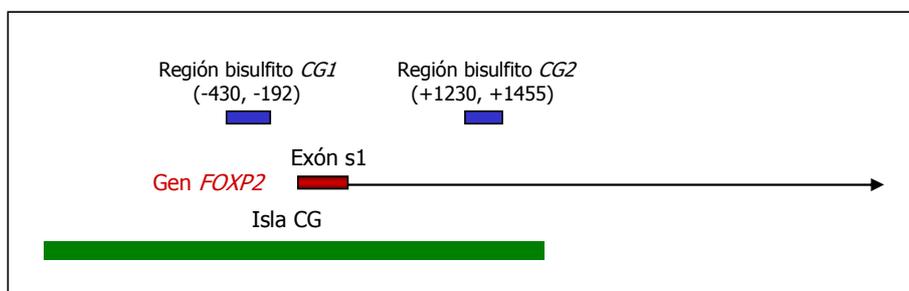


Figura M6. Localización de los fragmentos utilizados para analizar el estado de metilación mediante bisulfito, respecto de la posición de la primera base del exón s1.

Nombre	Secuencia 5'-3'	Tamaño del producto (pb)
CGF3 CGR2	GCGGGTTGTTTATATAGTAGGTGGATT CAAACTACCTCTCCTCGAAAAAA	289
CG2F3 CG2R3	GTTATTTGGAAGTTTATAGTGGTT TAACTTTTCTCCCTACTCTAAAA	274

Tabla M42. Cebadores utilizados para la amplificación de DNA convertido con bisulfito.

II.8.2.3 Tratamiento con bisulfito

Previa a la modificación con bisulfito se llevó a cabo un pretratamiento del DNA consistente en digerir 1 µg del DNA genómico con 10 unidades de *EcoRI* durante 2 horas, con posterior adición de 1µg de Proteasa K e incubación durante 12 horas a 37°C. El producto resultante se purificó mediante precipitación con NH₄Ac 4M (ver apartado II.1.5.2) o utilización del kit *Qiaex II* (Qiagen).

La modificación con bisulfito se llevó a cabo mediante dos kits:

- *GpGenome DNA Modification kit* (Chemicon Internacional, Inc.)
- *EpiTect Bisulfite Kit* (Qiagen)

Ambos kits están basados en el mismo proceso. Inicialmente, las bases nitrogenadas quedan expuestas mediante la desnaturalización del DNA por medio de un medio alcalino y de temperatura suave. A continuación, una sal de bisulfito da lugar a que las citosinas no metiladas sean sulfonadas y desaminadas hidrolíticamente, dando lugar a uracil sulfonato como intermediario. El DNA se une entonces a partículas o a un soporte de tipo columna y se eliminan las sales. La conversión a uracilo se completa mediante desulfonación alcalina, tras la que se purifica de nuevo el producto para eliminar las sales.

A continuación, se presenta brevemente el protocolo utilizado para cada uno de los kits utilizados.

En caso de utilizar el *GpGenome DNA Modification kit*, el producto de la digestión se resuspendía en 18 µl de H₂O, mientras que cuando se utilizó el *EpiTect Bisulfite Kit* se resuspendió en 20 µl. En ambos casos, la cantidad total de DNA para convertir era la misma.

- ***GpGenome DNA Modification kit*** (Chemicon Internacional)

Modificación con bisulfito

- Incubar los siguientes componentes a 50°C durante 10':

Digestión DNA genómico purificada (1 µg)	18 µl
NaOH 3M (preparado en el momento)	7 µl
H ₂ O	75 µl

- Añadir 550 µl de *DNA Modification Reagent I* (solución que contiene la sal de bisulfito), preparada en el momento y mezclar con vórtex.
- Incubar a 50°C durante 20 horas.

Eliminación de sales inicial

- Añadir 5 µl de *DNA Modification Reagent III*, agitado vigorosamente previamente para su correcta resuspensión.
- Añadir 750 µl de solución *DNA Modification Reagent II*, preparada en el momento y mezclar.
- Incubar a temperatura ambiente durante 5-10'.
- Centrifugar a 10000 rpm durante 30". Eliminar sobrenadante.

Material y Métodos

- Añadir 1 ml de Etanol al 70%, mezclar con vortex y centrifugar a 10000 rpm durante 30". Descartar sobrenadante. Repetir este paso 3 veces.
- Centrifugar a 10000 rpm durante 2'. Eliminar sobrenadante residual con pipeta.

Desulfonación, segunda eliminación de sales y elución

- Añadir 50 µl de NaOH 20mM/Etanol 90%. Mezclar por vortex.
- Incubar 5' a temperatura ambiente.
- Centrifugar brevemente para que todo el contenido vaya al fondo.
- Añadir 1 ml de Etanol al 90%. Mezclar por vortex y centrifugar a 10000 rpm durante 30". Descartar sobrenadante. Repetir este paso 2 veces.
- Centrifugar a 10000 durante 3'. Eliminar sobrenadante residual con pipeta.
- Dejar secar durante 15' al aire.
- Añadir 25 µl de tampón TE
- Incubar en TE a 50°C durante 15'.
- Centrifugar a máxima velocidad durante 3'.
- Transferir sobrenadante con DNA modificado a un nuevo tubo.

Solución	Composición
<i>DNA Modification Reagent I</i>	Por muestra: 0.227 g de <i>DNA Modification Reagent I</i> (atemperado) 20 µl de NaOH 3M (recién preparado) 571 µl H ₂ O
<i>DNA Modification Reagent II</i>	Por muestra: 1.35 g de <i>DNA Modification Reagent II</i> 750 µl de la mezcla: 1 µl β-mercaptoetanol/20 ml H ₂ O
NaOH 3M	1 g de NaOH 8.3 ml de H ₂ O
NaOH 20mM/Etanol 90%	En 1 ml: 900µl Etanol 100% 6.6 µl NaOH 3M 93.4 µl H ₂ O
TE	10 ml de Tris-HCl (pH 8.0) 400µl EDTA 0.25 M 990 ml H ₂ O

Tabla M43. Composición de las soluciones utilizadas en la modificación por bisulfito mediante el kit *GpGenome DNA Modification kit*.

▪ ***EpiTect Bisulfite Kit (Qiagen)***

Modificación con bisulfito

- Mezclar los siguientes componentes:

Digestión DNA genómico purificada	20 µl
Mezcla de bisulfito	85 µl
Tampón de protección de DNA	35 µl

- En un termociclador, llevar a cabo el siguiente programa de ciclos:

Paso	Tiempo	Temperatura
Desnaturalización	5 min	99°C
Incubación	25 min	60°C
Desnaturalización	5 min	99°C
Incubación	85 min	60°C
Desnaturalización	5 min	99°C
Incubación	175 min	60°C
	-	20°C

Purificación del DNA convertido

- Centrifugar brevemente para que todo el contenido del tubo quede en el fondo.
- Transferir la reacción a un tubo de 1.5 ml.
- Añadir 560 μ l de *Tampón BL* recién preparado con 10 μ g/ml de *RNA carrier*. Mezclar con vórtex.
- Trasferir mezcla a una columna *Epitect spin* con tubo colector.
- Centrifugar a máxima velocidad durante 1'. Descartar eluido.
- Añadir 500 μ l de *Tampón BW*.
- Centrifugar a máxima velocidad durante 1'. Descartar eluido.

Desulfonación

- Añadir 500 μ l de *Tampón BD*. Incubar 15' a temperatura ambiente.
- Centrifugar a máxima velocidad durante 1'.

Purificación y elución

- Añadir 500 μ l de *Tampón BW*. Centrifugar a máxima velocidad durante 1'. Descartar eluido. Repetir este paso.
- Colocar la columna en un tubo colector de 2 ml nuevo y centrifugar durante 1' para eliminar cualquier residuo líquido.
- Colocar la columna en un tubo de 1.5 ml nuevo.
- Añadir 20 μ l de *Tampón de Elución*.
- Centrifugar a máxima velocidad durante 1'.
- DNA convertido queda en el tubo de 1.5 ml.

II.8.2.4 Amplificación por PCR del DNA convertido

La amplificación de los fragmentos de interés se llevó a cabo con *REDTaq[®] ReadyMix[™] PCR Reaction Mix* (Sigma Aldrich) bajo las siguientes condiciones:

Nombre	Mezcla de PCR	Programa de PCR	Temperatura de asociación
CGF3 CGR2	Mezcla <i>REDTaq</i>	Programa bisulfito	61°C
CG2F3 CG2R3	Mezcla <i>REDTaq</i>	Programa bisulfito	59°C

Tabla M44. Condiciones de las reacciones de PCR utilizadas para amplificar fragmentos de DNA convertidos con bisulfito.

Material y Métodos

El volumen total de reacción se cargó en geles de agarosa al 1.5% y una vez confirmada la especificidad y tamaño de los productos, se purificaron a partir del gel mediante el kit *Qiaex II* (Qiagen).

II.8.2.5 Clonación de los productos de PCR

El producto resultante de la amplificación de DNA convertido con bisulfito puede ser secuenciado directamente o clonado en bacterias y secuenciado posteriormente. El primer caso es técnicamente más complicado, mientras que el segundo además de ser más sencillo, proporciona más información, permitiendo determinar la proporción de células en las que un sitio está metilado. Por esta razón, en nuestro caso clonamos los productos de PCR obtenidos.

Así, los productos de PCR purificados se clonaron utilizando el sistema *pGEM-T[®] Easy Vector*, (descrito en el epígrafe II.1.7), transformando en células de *E.coli* JM109 o *E. coli* XL-Blue (epígrafe II.1.7.2).

Las células transformadas se sembraron en placas con Ampicilina, X-Gal e IPTG (ver tabla M18, para concentraciones) y se seleccionaron las colonias blancas como aquellas que integraron el vector con el producto de PCR insertado.

II.8.2.6 Obtención de la secuencia de cada clon

A continuación se llevó a cabo una PCR de colonia con los cebadores M13R y T7, o SP6 y T7 presentes en el vector de clonación (las secuencias se muestran en tabla M9, epígrafe II.1.3.1). La mezcla de reacción de la amplificación fue la mezcla estándar y el programa el programa bisulfito, con temperatura de asociación de 51°C. Se amplificó un mínimo de 4 clones positivos. Los productos de PCR se purificaron por precipitación con NH₄Ac 4M y secuenciaron.

La mayor parte de las reacciones de secuenciación, con mezcla y programa estándar, se llevaron a cabo en el laboratorio de Genética Molecular de la Universitat de València y los productos se enviaron al Servicio de Secuenciación. Parte de las reacciones fueron llevadas a cabo por el Servicio de Secuenciación del Departamento de Bioquímica de la Universidad de Oxford.

II.9 PCR cuantitativa

II.9.1 Muestras analizadas

En la tabla M45 se muestran las muestras de RNA procedentes de diferentes áreas cerebrales disponibles para el análisis por PCR cuantitativa. Como se mencionó previamente, estas muestras habían sido donadas al centro SANE POWIC (*SANE Prince of Wales International Centre*) de Oxford.

	Sexo	Circunvolución del cíngulo		Circunvolución temporal superior		Circunvolución del parahipocampo		Paracingulado	
		Izq	Dcho	Izq	Dcho	Izq	Dcho	Izq	Dcho
Controles	M	6	4	6	4	5	3	6	4
	F	3	3	2	4	3	4	4	7
Pacientes	M	4	3	6	3	4	1	6	3
	F	5	3	6	4	5	3	4	3

Tabla M45. Muestras disponibles para el análisis por PCR cuantitativa.

II.9.2 RT-PCR

El protocolo seguido para la conversión de RNA en cDNA se muestra a continuación:

- Descongelación en hielo de la muestra de RNA.
- Dilución en H₂O libre de RNasas de 1 µg de RNA en un volumen total de 11 µL.
- Adición de 1 µL de *Random Primer Hexanucleotides* 300 mM (Promega)
- Incubación durante 5' a 65°C.
- Durante la incubación, preparación de la mezcla:

Componente	Cantidad/muestra
Tampón <i>First Strand</i> 5X	4 µL
DTT; ditiotretitol (0.1M)	1 µL
dNTPs (10 µM)	2 µL
<i>SuperScript™ III Reverse Transcriptase</i> (200 U/µL)	1 µL
Total	8 µL

SuperScript™ III Reverse Transcriptase, DTT y tampón 5X (Invitrogen).

- Centrifugación breve de los tubos, para llevar el contenido al fondo, y colocación de los tubos en hielo.
- Adición de los 8 µL de mezcla preparados a cada tubo.
- Incubación 5' a temperatura ambiente.
- Incubación de 1 hora a 50°C.
- Inactivación durante 5' a 60°C.

II.9.3 Diseño de cebadores para la PCR cuantitativa

El diseño de cebadores que amplificaran exclusivamente cDNA se llevó a cabo utilizando el programa *PerIPrimer* (versión 1.1.14), introduciendo la secuencia correspondiente al mensajero del gen *FOXP2* (Número de acceso: AY144615) y la secuencia de DNA genómico correspondiente procedente de la base de datos del *UCSC Genome Browser*.

Los cebadores para el gen *RPII*, utilizado para normalizar, fueron amablemente cedidos, por el doctor James Close, investigador del centro *SANE POWIC (SANE Prince of Wales Internacional Centre)* de Oxford.

Las secuencias de los cebadores utilizados se muestran a continuación:

Nombre	Secuencia (5'-3')	Tamaño del producto (pb)	Región amplificada
FOXQF FOXQR	GGTGCAACAGTTAGAAATACAG ACCAGATTTAGAGGTTTGGGA	116	Exones 9-11
RPIIF RPIIR	GTGCGGCTGCTTCCATAA GCACCACGTCCAATGACAT	228	Exones 24-26

Tabla M46. Cebadores utilizados para la PCR cuantitativa.

II.9.4 Condiciones de la amplificación por PCR para comprobar conversión del RNA a cDNA

Para cada muestra se comprobó mediante reacción de PCR específica de cDNA que la conversión de RNA a cDNA había funcionado correctamente. El programa de PCR utilizado fue el Programa cDNA 1, con 60°C como temperatura de asociación. La mezcla de reacción fue la mezcla estándar.

II.9.5 qRT-PCR

Las reacciones de PCR cuantitativa se realizaron mediante el uso de *SYBR Green*. El *SYBR Green* es un marcador fluorescente que se une al surco pequeño de las moléculas de DNA de doble cadena aumentando entonces la intensidad de la emisión de fluorescencia. Así, durante la PCR, la polimerasa *DNA AmpliTaq Gold*[®] amplifica la secuencia diana, creándose productos de PCR. El marcador *SYBR Green I* se une entonces a cada nueva copia de DNA de doble cadena. Conforme progresa la PCR, se generan más productos de PCR. Puesto que el marcador *SYBR Green I* se une a todos los DNAs de doble cadena, el resultado es un aumento de la intensidad de fluorescencia proporcional a la cantidad de producto de PCR producida.

En la figura M7 está esquematizado el proceso:

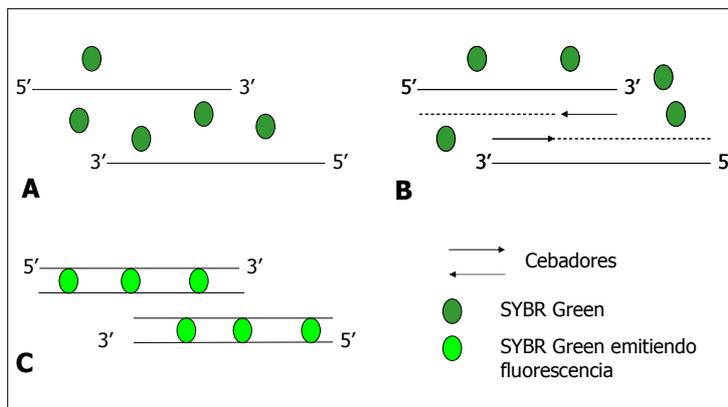


Figura M7. Esquema del funcionamiento del SYBR Green, utilizada en la PCR cuantitativa. A. Cuando el DNA está desnaturado, el marcador *SYBR Green* flota libre en el medio. B. La fase de extensión comienza cuando los cebadores se unen al DNA molde. C. Cuando la polimerización se completa, el *SYBR Green* se une a los productos de PCR sintetizados, emitiendo fluorescencia.

Para las reacciones de PCR cuantitativa se usó el *Power SYBR® Green PCR Master Mix* (Applied Biosystems), que incluye todos los componentes necesarios, a excepción de los cebadores y el DNA.

Las mezclas de reacción, en función de los diferentes experimentos realizados, fueron las siguientes:

Componente	Cantidad	
	Experimentos 1, 2, y 3	Experimento 4
H ₂ O	10 µL	7.8 µL
Cebador F (10 mM)	0.75 µL	0.6 µL
Cebador R (10 mM)	0.75 µL	0.6 µL
<i>Master Mix SYBR</i>	12.5 µL	10 µL
cDNA procedente de la RT-PCR	1 µL	1 µL
Total	25 µL	20µL

Tabla 47. Mezcla de reacción para PCR cuantitativa con *Power SYBR® Green PCR Master Mix* (Applied Biosystems).

Previo el análisis de las muestras procedentes de cerebro, se optimizaron las condiciones de la PCR cuantitativa para cada par de cebadores.

Según el experimento realizado, se usaron 2 o 3 réplicas de cada muestra de cDNA (experimentos 1-2 y 4, dos réplicas; experimento 3, tres réplicas).

Finalmente, el programa de PCR utilizado para todos los experimentos fue el siguiente:

Ciclo	Pasos	Temperatura (°C)	Tiempo	
1	1	25	30 seg	
2	1	95	3 min	
3	1	95	30 seg	X 50 Análisis en tiempo real
	2	60	30 seg	
	3	72	30 seg	
4	1	95	1 min	
5	1	55	1 min	
6	1	55	10 seg	

Tabla M48. Programa de PCR utilizado para los experimentos de PCR cuantitativa con *Power SYBR® Green PCR Master Mix* (Applied Biosystems).

En cada uno de los experimentos, además de con las muestras problema, se llevó a cabo la reacción con diluciones seriadas (1, 1/5, 1/25, 1/125, 1/625) de uno de los cDNAs, con el fin de obtener la recta patrón, a partir de la cual se calcularon los valores de las muestras problema (ver siguiente apartado de análisis y cuantificación de resultados).

Las reacciones de PCR se llevaron a cabo utilizando el sistema de detección de PCR en tiempo real *iCycler iQ* (Bio-Rad Laboratories).

II.9.6 Análisis y cuantificación de resultados

Para el análisis de los resultados obtenidos se utilizó el software asociado al sistema de detección utilizado, *iCycler iQ Optical System Software versión 3.1.7050*, y el programa Excel de Microsoft.

Se utilizó uno de los métodos más comunes de análisis de resultados de PCR cuantitativa: la cuantificación mediante la utilización de curvas estándar.

Para ello, se obtuvieron las curvas estándar correspondientes a cada gen analizado, a partir de las reacciones correspondientes a las diluciones seriadas realizadas. En estas curvas estándar, las concentraciones relativas de las diluciones se expresaron en unidades arbitrarias y se representaron los logaritmos en base 10 frente a los valores de Ct o ciclo umbral, que indican el número fraccional de ciclo en el que la cantidad de diana amplificada alcanza un umbral determinado. A partir de esta representación se obtuvo una recta patrón con la que, para cada uno de los valores de Ct de las muestras problema se extrapoló la concentración inicial relativa de cada una de ellas.

Para normalizar los resultados del gen *FOXP2* se utilizó el gen *RPII*.

Los pasos a seguir, realizados con el programa Excel de Microsoft fueron los siguientes:

- Obtención de las medias aritméticas de los valores de *FOXP2* y *RPII* para cada muestra.
- División de la media obtenida para el gen *FOXP2* entre la obtenida para el gen *RPII*.

La desviación estándar de este cociente se calcula a partir de las desviaciones estándares de los valores de *FOXP2* y *RPII*, siguiendo las siguientes fórmulas:

$$CV_{\text{NORM}} = \sqrt{CV_{\text{FOXP2}}^2 + CV_{\text{RPII}}^2}$$

$$CV = \frac{S}{\bar{X}} = \frac{\text{Desviación estándar}}{\text{Media}}$$

$$\text{Desviación estándar}_{\text{NORM}} = (CV_{\text{NORM}}) (\text{Media}_{\text{NORM}})$$

- Designación de una de las muestras como calibrador.
- División del valor de cada muestra normalizada entre el valor de la muestra normalizada utilizada como calibrador.

En este caso se trata de una división por una constante arbitraria, por lo que el valor cv de este resultado es el mismo que el valor CV_{NORM} . Para obtener la desviación típica se multiplica por el nuevo valor de media.

III. Herramientas informáticas y bases de datos

III.1 Programas informáticos

III.1.1 Programas utilizados en el estudio de asociación

Para el estudio de asociación se ha utilizado una amplia variedad de programas informáticos.

A continuación se muestran todos ellos, junto con un breve resumen de para qué fueron empleados:

- ***SPSS***

Los datos originales de la genotipación fueron introducidos en una base de datos de SPSS, en la que se encuentran también una larga serie de variables demográficas y clínicas de todos los individuos.

El SPSS permitió llevar a cabo tablas de contingencia de los SNPs con diferentes variables y obtener los valores de χ^2 y valores P asociados al correspondiente test exacto de Fisher.

- ***Haploview***

Este programa se utilizó para determinar si las distribuciones de alelos para cada uno de los SNPs estudiados se encuentran en equilibrio de Hardy-Weinberg.

También permitió obtener valores de desequilibrio de ligamiento entre marcadores moleculares bialélicos y realizar representaciones gráficas con ellos.

Además, este programa tiene implementada la herramienta *tagger*. Esta herramienta permite la selección y evaluación de *tagSNPs* de datos genotípicos como los presentes en el *International HapMap Project*. Los usuarios pueden especificar regiones genómicas de interés en las que el programa elige los *tagSNPs* que capturan todas las variantes de interés (Barrett *et al.*, 2005).

- ***SNPstats* (<http://bioinfo.iconcologia.net/snpstats/start.htm>)**

Este programa permitió realizar análisis de asociación entre polimorfismos bialélicos y la esquizofrenia u otras variables clínicas (Sole *et al.*, 2006).

Si la variable respuesta es dicotómica, aplica el método estadístico de regresión logística, y proporciona una estimación de la OR (odds ratio) para cada genotipo respecto al de referencia, que sirve como medida de asociación al cuantificar cuánto más probable es la aparición de una determinada enfermedad en presencia de un genotipo dado respecto a cuando no está presente.

En el caso de variables con más de dos niveles ó estados, el programa utiliza el método estadístico de regresión lineal y proporciona un valor medio junto con el error estándar para la variable respuesta en cada clase, según el modelo de herencia. El valor medio para la clase más frecuente se asume como valor de referencia y se calcula su diferencia con las medias de las otras clases.

Al mismo tiempo que lleva a cabo el análisis de asociación, este programa prueba qué modelo de herencia se ajusta más a las proporciones observadas en los grupos

Material y Métodos

comparados partiendo de la base de que cada genotipo está formado por dos alelos y el riesgo del genotipo depende del número de copias de determinado alelo que contiene. Así, en función del número de copias que se necesitan para modificar el riesgo, los modelos disponibles son los siguientes:

- *modelo codominante*: modelo general en el que cada genotipo tiene un riesgo diferente y no aditivo, donde se compara el heterocigoto y el homocigoto para el alelo de menor frecuencia con el homocigoto para el alelo más frecuente,
- *modelo dominante*: modelo en el que una copia de un alelo es suficiente para modificar el riesgo, y por tanto, genotipos homocigotos y heterocigóticos que la contienen tienen el mismo riesgo
- *modelo recesivo*: en el que se necesitan dos copias para cambiar el riesgo
- *modelo de sobredominancia*: modelo en el que el heterocigoto tiene mayor riesgo
- *modelo aditivo*: en el que cada copia de determinado alelo modifica el riesgo de forma aditiva.

Para ayudar al usuario a seleccionar el mejor modelo, el programa calcula el criterio de información akaike (AIC) y el criterio de información bayesiano (BIC), de manera que el modelo a seleccionar es aquel cuyos valores AIC y BIC son menores.

- **UNPHASED**

El programa UNPHASED permite realizar análisis de haplotipos. Implementa inferencias de máxima probabilidad en el efecto de los haplotipos al tiempo que permite la inclusión de datos perdidos, como aquellos con fase incierta y genotipos perdidos. Proporciona tests de haplotipos tanto a nivel global, buscando diferencias entre los grupos estudiados para todos los haplotipos, como a nivel individual, dando valores de OR, χ^2 y significación para cada haplotipo. Además, se pueden realizar permutaciones que permiten validar la consistencia de los resultados (Dudbridge *et al.*, 2008).

- **Quanto**

Este programa se utilizó para calcular la medida de poder estadístico de nuestras muestras, en función de las frecuencias alélicas obtenidas y el modelo determinado por el programa SNPstats (Gauderman *et al.*, 2002).

III.1.2 Programas de análisis de secuencias

Sequencher

Programa ampliamente utilizado para el análisis de las secuencias obtenidas las reacciones de secuenciación.

DNASTAR

De este programa se utilizaron las herramientas *SeqManII v5.0* y *EditSeq v5.0* para el análisis de las secuencias de los productos de secuenciación del análisis de metilación.

GeneScan-v3.7

Utilizado para el análisis del tamaño de fragmentos de DNA.

ClustalX 1.83 y ClustalW (este último incluido en MEGA4)

El programa *ClustalX 1.83* fue utilizado para el alineamiento inicial de las secuencias de primates. Posteriormente se llevaron modificaciones manuales del alineamiento, en aquellas posiciones no resueltas correctamente por el programa.

Al alineamiento de las secuencias de primates se añadieron las secuencias correspondientes a las especies *Mus musculus* y *Rattus norvegicus*. Este paso se llevó a cabo dentro del programa *ClustalW* implementado en el paquete *MEGA4*, con la opción de mantener los huecos existentes.

III.1.3 Programas de análisis evolutivo**III.1.3.1 Paquete informático Genomatix**

- ***El Dorado***

El programa de anotación genómica *El Dorado* de *Genomatix* contiene información de las secuencias genómicas de diferentes organismos y utiliza información de diferentes fuentes junto con datos generados con algoritmos propios de *Genomatix*. Ofrece información, entre otros, de regiones promotoras y predicciones realizadas con el programa *Promoter Inspector*, también incluido en el paquete *Genomatix*, para los genes anotados.

Del programa *El Dorado* se extrajeron los promotores anotados situados en la región 5' del gen *FOXP2*, de las especies *H. sapiens*, *P. troglodytes*, *M. mulatta*, *M. musculus*.

- ***Promoter Inspector***

Este programa se utilizó para obtener predicciones de regiones promotoras para las especies analizadas en este trabajo no disponibles en *El Dorado* (Scherf *et al.*, 2000).

- ***GEMs Launcher y MatInspector***

Con el fin de determinar la presencia de sitios de unión a factores de transcripción conservados dentro de las secuencias estudiadas se utilizó la herramienta *GEMs Launcher* de *Genomatix*, que proporciona un listado de estos sitios de unión presentes en comparaciones de secuencias múltiples. La búsqueda la lleva a cabo *MatInspector*, también herramienta de *Genomatix*, que utiliza una amplia biblioteca de descripciones de matrices para los sitios de unión a factores de transcripción, para localizar los complementarios de DNA. Proporciona datos de similitud de la secuencia problema con la matriz para el sitio de unión al factor de transcripción así como similitud con únicamente el núcleo o core de ese mismo sitio. Además, asigna un índice de calidad a los aciertos, permitiendo su filtrado y selección en función de la similitud de la matriz con la secuencia testada (Quandt *et al.*, 1995).

III.1.3.2 Otros programas de análisis evolutivo

- ***E-Shadows***

Este programa es una herramienta para llevar a cabo análisis comparativo de secuencias cercanas. Está basado en la consideración de que mutaciones en los diferentes linajes se acumulan independientemente en cada uno de ellos durante la evolución. Analiza alineamientos de secuencias múltiples del programa *ClustalW*, obtenidos de secuencias muy similares en comparaciones pareadas e identifica aquellas regiones que han acumulado pequeñas cantidades de mutaciones a través de la evolución. Para esto utiliza dos aproximaciones: Islas del Modelo Oculto de Harkov (*Hidden Harkov Model Islands, HMMI*) y el Umbral de divergencia (*Divergente Threshold, DT*) para distinguir entre regiones que evolucionan neutralmente y regiones que lo hacen funcionalmente (Ovcharenko *et al.*, 2004).

- ***Mulan***

El programa *Mulan* detecta regiones conservadas evolutivamente en alineamientos y permite implementar lo que se conoce como *phylogenetic shadowing* o ensombrecimiento filogenético, estrategia utilizadas para identificar elementos génicos con baja tasa de mutación en comparaciones de secuencias múltiples de especies muy cercanas (Ovcharenko *et al.*, 2005).

- ***PaupUp***

Programa utilizado para obtener las relaciones filogenéticas entre las secuencias de las especies analizadas. Incorpora herramienta para llevar a cabo el *model test*, que permite determinar el modelo evolutivo más adecuado cuando se lleva a cabo un análisis por máxima verosimilitud.

- ***PhyML online* (<http://atgc.lirmm.fr/phyml/>)**

Programa filogenético basado en el principio de máxima verosimilitud (Guindon and Gascuel, 2003).

- ***MEGA4***

Programa que integra herramientas para llevar a cabo alineamientos manuales y automáticos de secuencias, inferencia de árboles filogenéticos y estima de tasas evolutivas. *MEGA* is an integrated tool for conducting automatic and manual sequence alignment.

III.1.4 Otros programas

- ***Perlprimer***

Programa utilizado para el diseño de cebadores específicos de cDNA.

- ***CpG Island Searcher* (<http://www.cpgislands.com/>)**

Este programa permite definir islas CG a partir de secuencias proporcionadas por el usuario así como representar de forma gráfica todos los dinucleótidos presentes en la secuencia.

- **Excel (Microsoft)**

III.2 Bases de datos y direcciones de Internet

UCSC Genome Browser (<http://genome.ucsc.edu/>)

NCBI (<http://www.ncbi.nlm.nih.gov/>)

Human Genome Duplication Database (<http://projects.tcag.ca/humandup/>)

Hapmap Genome Browser (http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B36/)

SymAtlas (<http://symatlas.gnf.org/SymAtlas/>)

Material y Métodos

Resultados



I. Variaciones estructurales en el gen *FOXP2* y en *HAR1A*

La primera parte de los resultados incluye un análisis de variaciones estructurales de los genes *FOXP2* y *HAR1A* en relación con el desarrollo de alucinaciones auditivas en pacientes psicóticos, así como con otras alteraciones del lenguaje.

Este análisis está centrado principalmente en el gen *FOXP2*, debido a que es el primer gen relacionado directamente con una alteración del lenguaje (Lai *et al.*, 2001). Así, para este gen se ha llevado a cabo un estudio de asociación caso-control en el que se compararon las frecuencias alélicas, genotípicas y haplotípicas de diferentes polimorfismos, entre pacientes y controles, así como en diferentes variables clínicas del espectro psicótico.

Continuando con el análisis del gen, se han buscado posibles expansiones de los tramos de poliglutaminas, ya que expansiones de este tipo de trinucleótidos se encuentran asociadas a diferentes enfermedades mentales.

Por último, se ha llevado a cabo un estudio del gen *FOXP2* en niños con Trastorno Específico del Lenguaje (TEL), en el que se realizó la búsqueda de las mutaciones descritas en región codificante, así como un análisis de la región promotora.

En cuanto al gen *HAR1A*, este gen se caracteriza por incluir una región con evolución acelerada en el linaje humano (Pollard *et al.*, 2006). Dada la propuesta de que la esquizofrenia constituye un desorden específicamente humano (Crow, 2000) y teniendo en cuenta que la evolución acelerada de genes implicados en desarrollo del sistema nervioso podría estar relacionada con el origen del *H. sapiens*, se consideró que un gen de las características del gen *HAR1A* podría constituir un buen gen candidato para la esquizofrenia.

I.1 Estudio de asociación caso control con polimorfismos del gen *FOXP2*

El gen *FOXP2* se localiza en el brazo largo del cromosoma 7 (7q31) como se muestra en la figura R1.

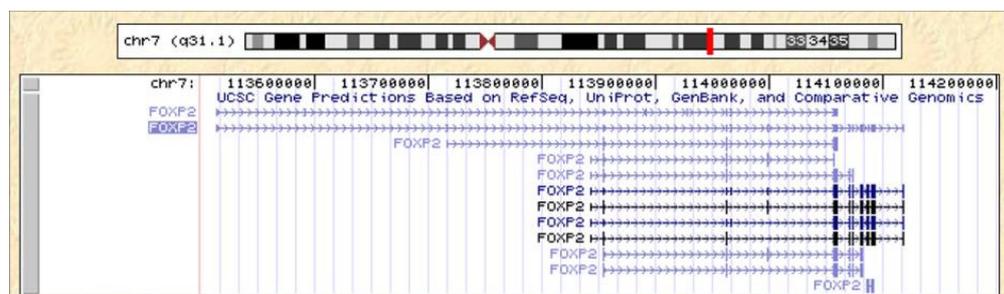


Figura R1. Contexto genómico en el que se localiza el gen *FOXP2*. Figura extraída del *UCSC Genome Browser*.

Resultados

Los SNPs estudiados se distribuyen a lo largo de una región de unas 605 Kb, donde se encuentra el gen, desde las posiciones -4282 hasta +601779 tomando como sitio +1 el primer nucleótido del exón s1. Estas posiciones se basan en la secuencia del gen *FOXP2* denominada "*Homo sapiens full length insert cDNA clone YX52E07*", procedente del *UCSC Genome Browser* (Figura R2). Los SNPs se localizan aguas arriba del gen y en las regiones intrónicas de los 26 exones que lo componen, encontrándose la mayoría de ellos en la región reguladora en 5' y cercanías del exón s1, así como en los intrones de exones no codificantes.

El estudio de asociación se aplicó a una muestra de 293 pacientes (215 con alucinaciones auditivas, AA) y 340 controles. Se analizó un total de 32 SNPs de los cuales 6 resultaron monomórficos en nuestra muestra. En la tabla R1, se muestran estos polimorfismos, así como el número de individuos analizados para confirmar la ausencia de variación. Puesto que los polimorfismos rs1358278 y rs2396722 se encontraron en completo desequilibrio de ligamiento en estudios previos, se decidió genotipar únicamente el segundo de ellos.

SNP	Método genotipación	Individuos analizados
rs6466478	RFLPs	50 controles, 42 pacientes
rs13308496	iPLEX™-MassArray	333 controles, 276 pacientes
rs7784307	RFLPs	18 controles, 46 pacientes
rs10276237	RFLPs	20 controles, 48 pacientes
rs7798050	RFLPs	48 controles, 48 pacientes
rs10254225	iPLEX™-MassArray	333 controles, 276 pacientes

Tabla R1. SNPs monomórficos en la muestra analizada.

Las frecuencias alélicas y genotípicas para el resto de polimorfismos se muestran en las tablas R4 y R5. Todos los polimorfismos, con excepción de rs717233 se encontraron en equilibrio de Hardy-Weinberg. Los valores de P asociados al test de equilibrio de Hardy-Weinberg se muestran en la tabla R2. Las frecuencias genotípicas del polimorfismo rs717233 no se encontraron en equilibrio de Hardy-Weinberg, por lo que este polimorfismo no se tuvo en cuenta a la hora de realizar los análisis de asociación.

	Valores P del test de equilibrio Hardy-Weinberg		SNPs	Valores P del test de equilibrio Hardy-Weinberg	
	Controles	Pacientes		Controles	Pacientes
rs7803667	1.0	0.3744	rs2694941	0.215	0.6228
rs10447760	0.7394	0.2217	rs1852469	0.3018	0.7478
rs6961558	1.0	1.0	rs10255943	1.0	0.7105
rs923875	1.0	0.2088	rs10486026	0.3348	0.9273
rs1597548	1.0	0.9671	rs2396753	0.0638	0.9787
rs10500038	0.3499	1.0	rs17137124	0.0596	0.4863
rs4730626	0.204	0.8688	rs7799652	0.7255	0.5769
rs717233	0.0391	0.4304	rs1456029	0.3627	0.9972
rs1668335	0.7392	0.3942	rs12670585	0.7125	0.5963
rs11771168	0.9577	0.8903	rs1456031	0.3718	1.0
rs1916977	0.1774	0.7229	rs2396765	0.7525	0.5808
rs2396722	0.6377	1.0	rs1456021	0.6308	0.7897
rs2253478	0.7714	0.5998			

Tabla R2. Polimorfismos estudiados y resultados del test de equilibrio de Hardy-Weinberg en los grupos analizados.

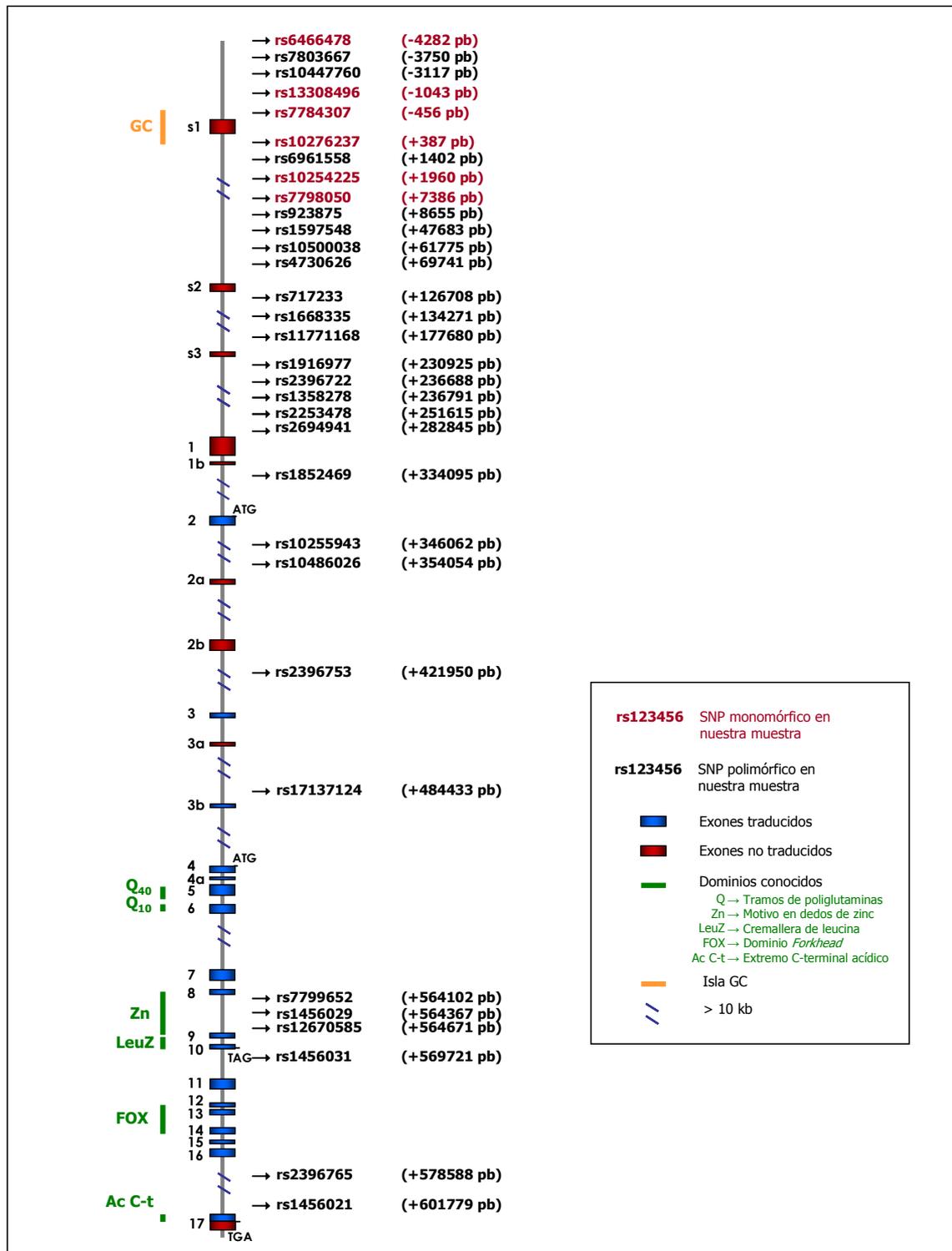


Figura R2. Distribución de todos los polimorfismos estudiados en la región del gen *FOXP2*. Las distancias reflejadas corresponden al número de pares de bases respecto a la primera base del exón s1, tomada como sitio +1. Se incluyen en la figura los dominios conocidos del gen.

Resultados

I.1.1 Análisis del desequilibrio de ligamiento

En las figuras R3 y R4 se muestran los esquemas del desequilibrio de ligamiento (LD) obtenido para controles y pacientes respectivamente. Los valores de LD de r^2 se calcularon mediante el programa *Haploview v4.1* (ver Material y Métodos, epígrafe III.1). Para cada SNP se puede observar este parámetro respecto al resto de los polimorfismos genotipados del gen, siguiendo la diagonal a partir de una de las esquinas de cada caja. Los colores de las cajas indican los siguientes valores de r^2 : $r^2 = 0$, blanco; $0 < r^2 < 1$, distintos tonos de gris; $r^2 = 1$, negro. En la parte superior de cada diagrama se indica la posición relativa de cada uno de los polimorfismos genotipados sobre la región analizada de cada gen.

Como puede observarse, los valores de LD obtenidos en este estudio varían a lo largo de la amplia región analizada. El rango de valores r^2 se encuentra entre valores de 0 y 0.968, tanto para controles como para pacientes.

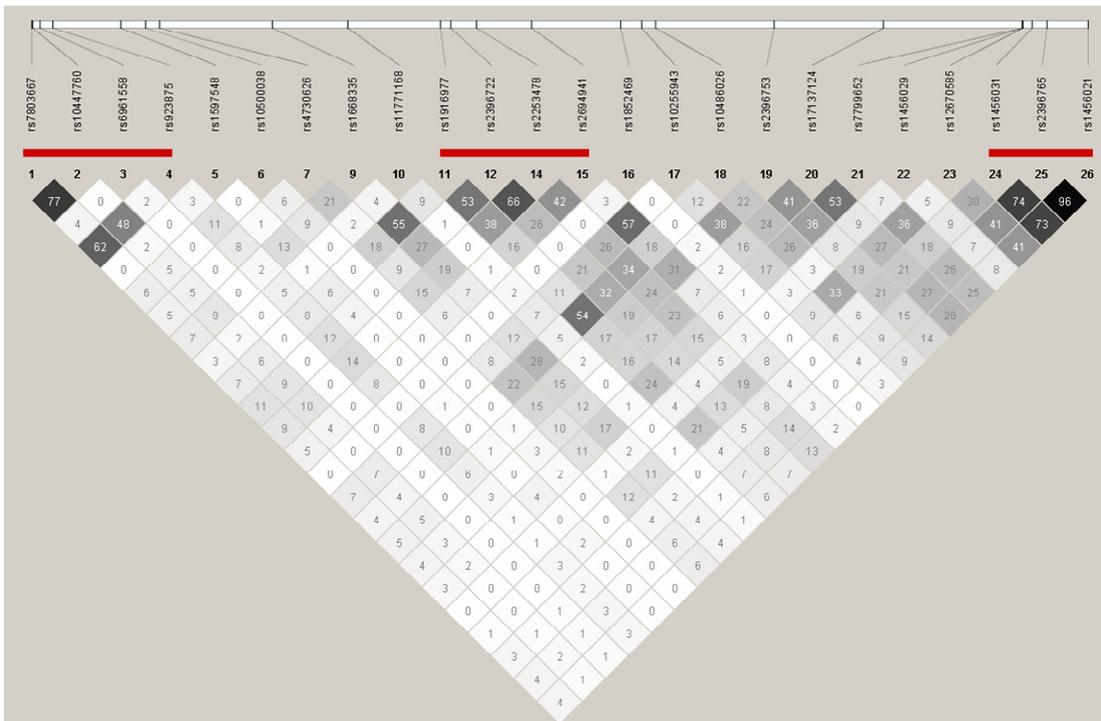


Figura R3. Representación de los valores de desequilibrio de ligamiento (LD) entre los polimorfismos analizados para el grupo de controles. En el gráfico aparecen representados los valores de r^2 sobre 100. El código de color es el siguiente $r^2=0$ blanco, $r^2=1$ negro, valores intermedios en grados de color gris, estimados por el programa *Haploview v4.1*. Las barras rojas representan posibles bloques de LD.

No se observan grandes bloques de LD. Únicamente destacan como posibles bloques los conjuntos formados por los polimorfismos rs7803667/rs10447760/rs6961558/rs923875, que abarcan 12 Kb, rs1916977/rs2396722/ rs2253478/rs2694941, de 50 Kb, y rs1456031/rs2396765/ rs1456021 de 31 Kb (Figuras R3 y R4). Estos patrones de LD son similares entre pacientes y controles, indicando ausencia de grandes diferencias en las frecuencias de recombinación entre los diferentes polimorfismos en ambos grupos muestrales.

En la figura R5 se muestra un gráfico de la misma región extraído del *HapMap*, correspondiente a la población CEU, que es la más cercana a la población muestral utilizada en este estudio. Aunque no se puede hacer una comparación precisa de los resultados obtenidos en este estudio con los obtenidos en el *HapMap*, sí se observa en los esquemas que en ambos casos, en la región cromosómica analizada, los mayores valores de LD se encuentran en los extremos y fragmento central, con zonas de bajo LD que separan esas áreas.

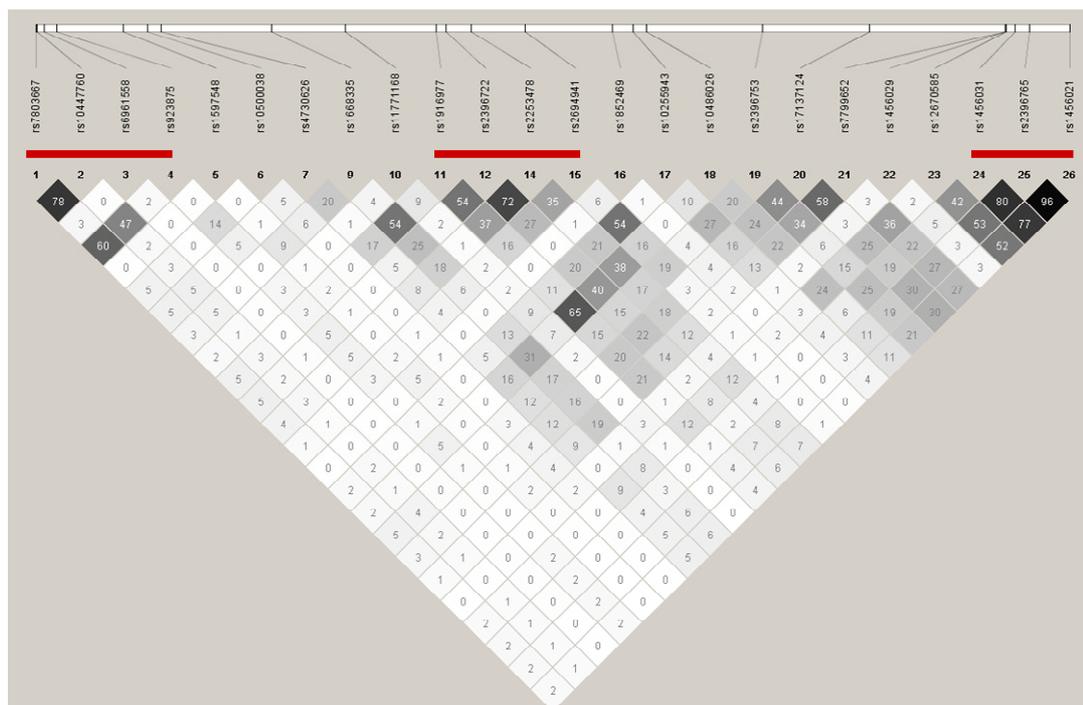


Figura R4. Representación de los valores de LD entre los polimorfismos analizados para el grupo de pacientes. En el gráfico aparecen representados los valores de r^2 sobre 100. El código de color es el siguiente: $r^2=0$ blanco, $r^2=1$ negro, valores intermedios en grados de color gris, estimados por el programa *Haploview v4.1*. Las barras rojas representan posibles bloques de LD.

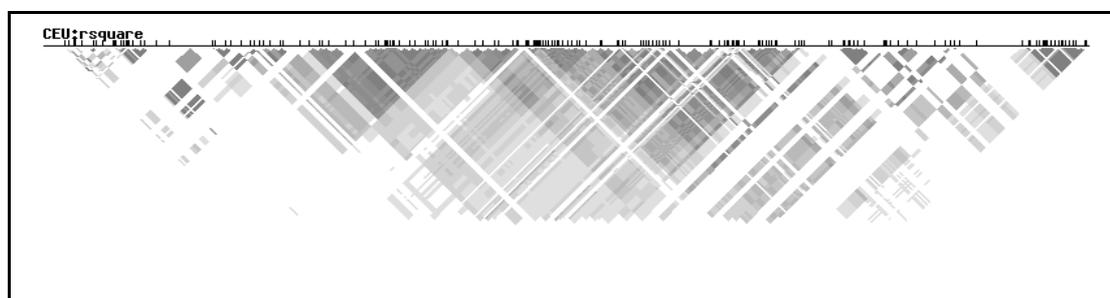


Figura R5. Esquema de los valores r^2 de LD disponibles para la región analizada obtenidos a partir de la base de datos del *HapMap (HapMap Data Rel 23a/ phaseII Mar 2008)*.

Resultados

I.1.2 Distribución de las frecuencias alélicas y genotípicas en pacientes y controles

A continuación se presentan los resultados obtenidos del análisis de la distribución de frecuencias genotípicas y alélicas en los grupos estudiados. Como se hace mención en el apartado de Material y Métodos (apartado II.2.2), algunos de los polimorfismos ya han sido analizados en otros estudios. Sin embargo, únicamente hay dos estudios de asociación de SNPs del gen *FOXP2* en relación con esquizofrenia, ambos de nuestro grupo y con una muestra más pequeña, submuestra del presente estudio (Sanjuán *et al.*, 2005; Sanjuán *et al.*, 2006b).

El estudio de asociación se ha realizado en distintos grupos muestrales, tal y como se indica en la tabla R3 y se especifica en el epígrafe I.1.1 del Material y Métodos. El número de individuos corresponde al número de muestras obtenidas en la genotipación por RFLPs (con excepción de rs6169558). En el caso de muestras genotipadas por *iPLEXTM-MassArray*, los valores pueden ser menores debidos a fallos en la genotipación (las muestras que fallaban en la genotipación por RFLPs se repetían hasta obtener el genotipo). Si es así, el número de individuos aparecerá reflejado en la tabla del correspondiente análisis.

Grupos comparados en el test de asociación	Número de individuos
Control/Paciente	340
	293
Control/Paciente con AA	340
	215
Control/Paciente con esquizofrenia	340
	214

Tabla R3. Grupos muestrales comparados en el test de asociación y número de individuos en cada comparación.

Por claridad, sólo se presentarán en las tablas resultados correspondientes a las comparaciones controles-pacientes y controles-pacientes con AA.

En las tablas R4 y R5 se muestran, las frecuencias genotípicas y alélicas de cada grupo. Además se incluyen los valores χ^2 y valores P asociados a los correspondientes tests de asociación mediante test exacto de Fisher. Los valores MAF (*Minor Allele Frequency*) fueron mayores de 0.02 en todos los polimorfismos analizados.

Al comparar pacientes frente a controles no se encontraron diferencias significativas en las frecuencias genotípicas para ninguno de los SNPs. Respecto a las frecuencias alélicas, se encontraron diferencias significativas únicamente en rs10447760 ($\chi^2=3.993$, $P=0.046$), aunque este valor no se mantiene significativo al corregir por Bonferroni ($P=1$).

SNP	Frecuencias genotípicas			χ^2	P	Frecuencias alélicas		χ^2	P (*)
	TT	TA	AA			T	A		
rs7803667									
Controles	169 (0.5)	142 (0.42)	29 (0.09)	3.145	0.208	0.71	0.29	2.587	0.108
Pacientes	125 (0.43)	139 (0.47)	29 (0.1)			0.66	0.34		
rs10447760	CC	CT	TT			C	T		
Controles	193 (0.57)	125 (0.37)	22 (0.06)	5.596	0.061	0.75	0.25	3.993	0.046 (1.00)
Pacientes	139 (0.47)	133 (0.45)	21 (0.07)			0.70	0.30		
rs6961558	GG	GA	AA			G	A		
Controles (N = 339)	323 (0.95)	16 (0.05)	0	0.666	0.431	0.98	0.02	0.652	0.419
Pacientes (N = 292)	282 (0.97)	10 (0.03)	0			0.98	0.02		
rs923875	AA	AC	CC			A	C		
Controles	125 (0.37)	164 (0.48)	51 (0.15)	1.597	0.455	0.61	0.39	0.831	0.362
Pacientes	94 (0.32)	154 (0.53)	45 (0.15)			0.58	0.42		
rs1597548	CC	CG	GG			C	G		
Controles	300 (0.88)	39 (0.11)	1 (0.002)	1.122	0.840	0.94	0.06	0.492	0.483
Pacientes	263 (0.9)	30 (0.1)	0			0.95	0.05		
rs10500038	CC	CT	TT			G	A		
Controles (N = 319)	213 (0.67)	99 (0.31)	7 (0.02)	1.371	0.510	0.82	0.18	0.7	0.403
Pacientes (N = 265)	171 (0.65)	84 (0.32)	10 (0.04)			0.80	0.20		
rs4730626	GG	GA	AA			G	A		
Controles (N = 315)	201 (0.64)	96 (0.3)	18 (0.06)	2.124	0.349	0.79	0.21	1.871	0.171
Pacientes (N = 262)	178 (0.68)	75 (0.29)	9 (0.03)			0.82	0.18		
rs1668335	GG	GA	AA			G	A		
Controles (N = 315)	163 (0.52)	125 (0.4)	27 (0.09)	0.185	0.913	0.72	0.28	0.011	0.917
Pacientes (N = 263)	139 (0.53)	100 (0.38)	24 (0.09)			0.72	0.28		
rs11771168	CC	CT	TT			C	T		
Controles (N = 319)	192 (0.6)	112 (0.35)	15 (0.05)	0.234	0.897	0.78	0.22	0.095	0.789
Pacientes (N = 265)	164 (0.62)	88 (0.33)	13 (0.05)			0.78	0.22		
rs1916977	AA	AG	GG			A	G		
Controles (N = 314)	183 (0.58)	107 (0.34)	24 (0.08)	0.400	0.824	0.75	0.25	0.019	0.892
Pacientes (N = 263)	152 (0.58)	94 (0.36)	17 (0.06)			0.76	0.24		
rs2396722	TT	TC	CC			T	C		
Controles	142 (0.42)	149 (0.44)	49 (0.14)	0.375	0.838	0.64	0.36	0.032	0.857
Pacientes	121 (0.41)	134 (0.46)	38 (0.13)			0.64	0.36		
rs2253478	GG	GA	AA			G	A		
Controles	128 (0.4)	138 (0.44)	51 (0.16)	1.900	0.402	0.62	0.38	0.003	0.954
Pacientes	100 (0.38)	129 (0.49)	35 (0.13)			0.62	0.38		

Tabla R4. Frecuencias genotípicas y alélicas en controles y pacientes con los valores de χ^2 y P asociados. Se indican los valores de N en aquellos SNPs en los que N<340 en controles y N<293 en pacientes.

^a corresponden a tests en los que en más de una de las clases los valores esperados son menores de 5.

(*) corresponde al valor P corregido por Bonferroni.

Resultados

SNP	Frecuencias genotípicas			χ^2	P	Frecuencias alélicas		χ^2	P (*)
	TT	TA	AA			T	A		
rs2694941									
Controles (N = 315)	110 (0.35)	144 (0.46)	61 (0.19)	0.198	0.909	0.58	0.42	0.001	0.979
Pacientes (N = 261)	89 (0.34)	124 (0.48)	48 (0.18)			0.58	0.42		
rs1852469									
Controles	319 (0.94)	21 (0.06)	0	1.863^a	0.388	0.97	0.03	1.247	0.264
Pacientes	269 (0.92)	23 (0.08)	1 (0.003)			0.96	0.04		
rs10255943									
Controles (N = 314)	154 (0.49)	126 (0.4)	34 (0.11)	0.837	0.672	0.69	0.31	0.001	0.973
Pacientes (N = 263)	125 (0.48)	114 (0.43)	24 (0.09)			0.69	0.31		
rs10486026									
Controles (N = 315)	196 (0.62)	97 (0.31)	22 (0.07)	2.385	0.305	0.78	0.22	0.275	0.600
Pacientes (N = 263)	163 (0.62)	89 (0.34)	11 (0.04)			0.79	0.21		
rs2396753									
Controles	99 (0.29)	186 (0.55)	55 (0.16)	4.860	0.089	0.56	0.44	3.759	0.053
Pacientes	73 (0.25)	153 (0.52)	67 (0.23)			0.51	0.49		
rs17137124									
Controles	84 (0.25)	167 (0.49)	89 (0.26)	1.915	0.383	0.49	0.51	1.825	0.177
Pacientes	86 (0.29)	139 (0.47)	68 (0.23)			0.53	0.47		
rs7799652									
Controles (N = 319)	87 (0.27)	168 (0.53)	64 (0.2)	2.137	0.346	0.54	0.46	1.488	0.223
Pacientes (N = 265)	87 (0.33)	129 (0.49)	49 (0.18)			0.57	0.43		
rs1456029									
Controles (N = 319)	192 (0.6)	109 (0.34)	18 (0.06)	1.685	0.427	0.77	0.23	1.728	0.189
Pacientes (N = 265)	147 (0.55)	98 (0.37)	20 (0.08)			0.74	0.26		
rs12670585									
Controles (N = 315)	156 (0.5)	126 (0.4)	33 (0.1)	1.025	0.608	0.70	0.30	0.573	0.449
Pacientes (N = 262)	134 (0.51)	107 (0.41)	21 (0.08)			0.72	0.28		
rs1456031									
Controles	98 (0.29)	165 (0.49)	77 (0.23)	2.387	0.308	0.53	0.47	2.116	0.146
Pacientes	101 (0.34)	133 (0.45)	59 (0.2)			0.57	0.43		
rs2396765									
Controles (N = 317)	111 (0.35)	149 (0.47)	57 (0.18)	2.059	0.356	0.59	0.41	1.598	0.206
Pacientes (N = 263)	100 (0.38)	127 (0.48)	36 (0.14)			0.62	0.38		
rs1456021									
Controles (N = 334)	116 (0.35)	158 (0.47)	60 (0.18)	3.676	0.157	0.58	0.42	3.085	0.079
Pacientes (N = 278)	109 (0.39)	134 (0.48)	35 (0.13)			0.63	0.37		

Tabla R4. Continuación.

Resultados

SNP	Frecuencias genotípicas			χ^2	P (*)	Frecuencias alélicas		χ^2	P (*)
	TT	TA	AA			T	A		
rs7803667	TT	TA	AA			T	A		
Controles	169 (0.5)	142 (0.42)	29 (0.09)	2.240	0.328	0.71	0.29	2.047	0.153
Pacientes con AA	93 (0.43)	100 (0.47)	22 (0.1)			0.67	0.33		
rs10447760	CC	CT	TT			C	T		
Controles	193 (0.57)	125 (0.37)	22 (0.06)	3.842	0.148	0.75	0.25	2.673	0.102
Pacientes con AA	104 (0.48)	96 (0.45)	15 (0.07)			0.71	0.29		
rs6961558	GG	GA	AA			G	A		
Controles (N = 339)	323 (0.95)	16 (0.05)	0	0.317^b	0.369	0.98	0.02	0.310	0.578
Pacientes con AA (N = 215)	207 (0.96)	8 (0.04)	0			0.98	0.02		
rs923875	AA	AC	CC			A	C		
Controles	125 (0.37)	164 (0.48)	51 (0.15)	0.465	0.784	0.61	0.39	0.275	0.600
Pacientes con AA	73 (0.34)	109 (0.51)	33 (0.15)			0.59	0.41		
rs1597548	CC	CG	GG			C	G		
Controles	300 (0.88)	39 (0.11)	1 (0.002)	0.094^a	0.956	0.94	0.06	0.0001	0.991
Pacientes con AA	189 (0.88)	26 (0.12)	0			0.94	0.06		
rs10500038	CC	CT	TT			G	A		
Controles (N = 319)	213 (0.67)	99 (0.31)	7 (0.02)	3.258	0.206	0.82	0.18	1.729	0.189
Pacientes con AA (N = 200)	126 (0.63)	64 (0.32)	10 (0.05)			0.79	0.21		
rs4730626	GG	GA	AA			G	A		
Controles (N = 315)	201 (0.64)	96 (0.3)	18 (0.06)	1.492	0.494	0.79	0.21	1.586	0.208
Pacientes con AA (N = 200)	137 (0.68)	55 (0.28)	8 (0.04)			0.82	0.18		
rs1668335	GG	GA	AA			G	A		
Controles (N = 315)	163 (0.52)	125 (0.4)	27 (0.09)	0.285	0.867	0.72	0.28	0.003	0.955
Pacientes con AA (N = 200)	102 (0.51)	83 (0.42)	15 (0.08)			0.72	0.28		
rs11771168	CC	CT	TT			C	T		
Controles (N = 319)	192 (0.6)	112 (0.35)	15 (0.05)	0.415	0.822	0.78	0.22	0.329	0.566
Pacientes con AA (N = 200)	126 (0.63)	65 (0.32)	9 (0.04)			0.79	0.21		
rs1916977	AA	AG	GG			A	G		
Controles (N = 314)	183 (0.58)	107 (0.34)	24 (0.08)	0.603	0.740	0.75	0.25	0.062	0.804
Pacientes con AA (N = 200)	116 (0.58)	72 (0.36)	12 (0.06)			0.76	0.24		
rs2396722	TT	TC	CC			T	C		
Controles	142 (0.42)	149 (0.44)	14 (0.14)	0.094	0.956	0.64	0.36	0.063	0.802
Pacientes con AA	91 (0.42)	95 (0.44)	29 (0.13)			0.64	0.36		
rs2253478	GG	GA	AA			G	A		
Controles (N = 317)	128 (0.4)	138 (0.44)	51 (0.16)	0.215	0.904	0.62	0.38	0.143	0.705
Pacientes con AA (N = 199)	82 (0.41)	88 (0.44)	29 (0.15)			0.63	0.37		

Tabla R5. Frecuencias genotípicas y alélicas en controles y pacientes con AA con los valores de χ^2 y P asociados.

Se indican los valores de N en aquellos SNPs en los que N<340 en controles y N<215 en pacientes con AA.

^a corresponde a tests en los que en más de una de las clases los valores esperados son menores de 5.

^b corresponde a tests en los que debido a la falta de clases, el test se realizó a partir de una tabla 2x2 en lugar de 3x2.

(*) corresponde al valor P corregido por Bonferroni.

Resultados

SNP	Frecuencias genotípicas			χ^2	P (*)	Frecuencias alélicas		χ^2	P (*)
	TT	TA	AA			T	A		
rs2694941	TT	TA	AA			T	A		
Controles (N = 315)	110 (0.35)	144 (0.46)	61 (0.19)	0.172	0.915	0.58	0.42	0.049	0.824
Pacientes con AA (N = 198)	66 (0.33)	94 (0.47)	38 (0.19)			0.57	0.43		
rs1852469	AA	AT	TT			A	T		
Controles	319 (0.94)	21 (0.06)	0	2.226^a	0.329	0.97	0.03	1.342	0.246
Pacientes con AA	197 (0.92)	17 (0.08)	1 (0.005)			0.96	0.04		
rs10255943	GG	GA	AA			G	A		
Controles (N = 314)	154 (0.49)	126 (0.4)	34 (0.11)	0.650	0.730	0.69	0.31	0.015	0.904
Pacientes con AA (N = 200)	94 (0.47)	87 (0.44)	19 (0.1)			0.69	0.31		
rs10486026	TT	TC	CC			T	C		
Controles (N = 315)	196 (0.62)	97 (0.31)	22 (0.07)	2.930	0.234	0.78	0.22	0.657	0.418
Pacientes con AA (N = 200)	126 (0.63)	67 (0.34)	7 (0.04)			0.79	0.21		
rs2396753	AA	AC	CC			A	C		
Controles	99 (0.29)	186 (0.55)	55 (0.16)	9.269	0.010 (0.24)	0.56	0.44	8.183	0.004 (0.096)
Pacientes con AA	45 (0.21)	115 (0.53)	55 (0.26)			0.48	0.52		
rs17137124	TT	TC	CC			T	C		
Controles	84 (0.25)	167 (0.49)	89 (0.26)	4.137	0.124	0.49	0.51	4.209	0.040 (0.96)
Pacientes con AA	68 (0.32)	103 (0.48)	44 (0.2)			0.56	0.44		
rs7799652	TT	TG	GG			T	G		
Controles (N = 319)	87 (0.27)	168 (0.53)	64 (0.2)	2.530	0.293	0.54	0.46	2.385	0.123
Pacientes con AA (N = 200)	66 (0.33)	102 (0.51)	32 (0.16)			0.58	0.42		
rs1456029	AA	AG	GG			A	G		
Controles (N = 319)	192 (0.6)	109 (0.34)	18 (0.06)	2.247	0.325	0.77	0.23	2.164	0.141
Pacientes con AA (N = 200)	110 (0.55)	73 (0.36)	17 (0.08)			0.73	0.27		
rs12670585	CC	CT	TT			C	T		
Controles (N = 315)	156 (0.5)	126 (0.4)	33 (0.1)	1.244	0.553	0.70	0.30	0.639	0.424
Pacientes con AA (N = 199)	102 (0.51)	82 (0.41)	15 (0.08)			0.72	0.28		
rs1456031	TT	TC	CC			T	C		
Controles	98 (0.29)	165 (0.49)	77 (0.23)	3.842	0.148	0.53	0.47	3.242	0.072
Pacientes con AA	73 (0.34)	106 (0.49)	36 (0.17)			0.59	0.41		
rs2396765	TT	TC	CC			T	C		
Controles (N = 317)	111 (0.35)	149 (0.47)	57 (0.18)	3.240	0.203	0.59	0.41	1.716	0.192
Pacientes con AA (N = 198)	74 (0.37)	100 (0.51)	24 (0.12)			0.63	0.37		
rs1456021	TT	TG	GG			T	G		
Controles (N = 334)	116 (0.35)	158 (0.47)	60 (0.18)	4.473	0.104	0.58	0.42	3.043	0.081
Pacientes con AA (N = 212)	82 (0.39)	106 (0.5)	24 (0.11)			0.64	0.36		

Tabla R5. Continuación.

Cuando se comparan controles frente a pacientes con AA se obtuvieron diferencias significativas en las frecuencias genotípicas para el polimorfismo rs2396753 ($\chi^2=9.269$, $P=0.010$). Al comparar frecuencias alélicas, además de este polimorfismo ($\chi^2=8.183$, $P=0.004$), se encontraron diferencias para rs17137124 ($\chi^2=4.209$, $P=0.040$). Sin embargo, al igual que en el caso anterior, estos valores no se mantienen significativos tras la corrección múltiple obteniéndose como valores P corregidos $P=0.24$ y $P=0.096$ al comparar frecuencias genotípicas y alélicas del polimorfismo rs2396753, y $P=0.96$ para frecuencias alélicas del polimorfismo rs17137124.

Cuando se compararon controles con pacientes con esquizofrenia (datos no mostrados), los resultados fueron similares. Se obtuvieron diferencias en las frecuencias genotípicas para los polimorfismos rs2396753 ($\chi^2=6.163$, $P=0.047$) y rs10447760 ($\chi^2=7.164$, $P=0.029$), aunque estas diferencias dejan de ser significativas al realizar la corrección de Bonferroni ($P=1$ y $P=0.696$, respectivamente). Respecto a las frecuencias alélicas, se obtuvieron diferencias significativas nuevamente para el polimorfismo rs2396753 ($\chi^2=5.434$, $P=0.020$), y una tendencia para el polimorfismo rs10447760 ($\chi^2=3.705$, $P=0.054$), aunque tras corregir por Bonferroni, se pierde la significación ($P=0.474$ y $P=1$ respectivamente).

El programa *SNPStats* permitió, mediante regresión logística testar qué modelo de herencia se ajustaba más a las proporciones observadas en casos y controles. Al mismo tiempo este programa proporciona una medida de asociación basada en la estimación de la odds ratio (OR), al cuantificar cuánto más probable es la aparición de la enfermedad en presencia de determinado genotipo respecto a cuando no está presente.

A continuación se muestra para cada polimorfismo, qué modelo se ajustó más en las comparaciones realizadas, esto es, qué modelo obtuvo un menor valor para el BIC (Criterio de Información Bayesiana), junto con los valores de OR y valor P asociado (Tablas R6 y R7).

Al comparar pacientes con controles mediante regresión logística con el programa *SNPStats* se observa que hay diferencias significativas para el polimorfismo rs10447760 ($P=0.02$), coincidente con los resultados obtenidos mediante el test exacto de Fisher mediante tablas de contingencia, y también para rs2396753 ($P=0.03$). No obstante, la significación se pierde al llevar a cabo la corrección múltiple ($P=0.48$ y $P=0.72$, respectivamente).

Al comparar controles con pacientes con AA mediante regresión logística se observa que hay diferencias significativas para los polimorfismos rs10447760 ($P=0.05$), rs2396753 ($P=0.003$), rs17137124 ($P=0.04$) y rs1456021 ($P=0.03$). Al igual que en casos anteriores la significación se pierde al llevar a cabo la corrección múltiple ($P=1$, $P=0.072$, $P=0.96$ y $P=0.72$, respectivamente). Los polimorfismos rs2396753 y rs17137124 también mostraron diferencias significativas en el análisis mediante el test exacto de Fisher, aunque no se mantenían tampoco tras las correcciones múltiples.

Resultados

SNPs	Modelo de herencia	Genotipos	Controles	Pacientes	OR	P (*)
rs7803667	Dominante	TT AT-AA	169 (0.50) 171 (0.50)	125 (0.43) 168 (0.57)	1 1.33 (0.97-1.82)	0.08
rs10447760	Dominante	CC CT-TT	193 (0.57) 147 (0.43)	139 (0.47) 154 (0.53)	1 1.45 (1.06-1.99)	0.02 (0.48)
rs6961558	No determinado	GG GA	323 (0.95) 16 (0.05)	282 (0.97) 10 (0.03)	1 0.72 (0.32-1.6)	0.41
rs923875	Dominante	AA AC-CC	125 (0.37) 215 (0.63)	94 (0.32) 199 (0.68)	1 1.23 (0.89-1.71)	0.22
rs1597548	Recesivo	CC-CG GG	339 (0.997) 1 (0.003)	293 (1.0) 0	1 0.00 (0-NA)	0.26
rs10500038	Recesivo	GG-GA AA	312 (0.98) 7 (0.02)	255 (0.96) 10 (0.04)	1 1.75 (0.66-4.66)	0.26
rs4730626	Aditivo				0.82 (0.62-1.10)	0.18
rs1668335	Sobredominancia	GG-AA GA	190 (0.60) 125 (0.40)	163 (0.62) 100 (0.38)	1 0.93 (0.67-1.30)	0.68
rs11771168	Sobredominancia	CC-TT CT	207 (0.65) 112 (0.35)	177 (0.67) 88 (0.33)	1 0.92 (0.65-1.30)	0.63
rs1916977	Recesivo	AA-AG GG	290 (0.92) 24 (0.08)	246 (0.93) 17 (0.07)	1 0.84 (0.44-1.59)	0.58
rs2396722	Recesivo	TT-CT CC	291 (0.86) 49 (0.14)	255 (0.87) 38 (0.13)	1 0.88 (0.56-1.40)	0.6
	Sobredominancia	CC-TT CT	191 (0.56) 149 (0.44)	159 (0.54) 134 (0.46)	1 1.08 (0.79-1.48)	0.63
rs2253478	Sobredominancia	GG-AA GA	179 (0.56) 138 (0.44)	135 (0.51) 129 (0.49)	1 1.24 (0.89-1.72)	0.2
rs2694941	Sobredominancia	TT-AA TA	171 (0.54) 144 (0.46)	137 (0.52) 124 (0.48)	1 1.07 (0.77-1.49)	0.67
rs1852469	Recesivo	AA-TA TT	340 (1.0) 0	292 (0.997) 1 (0.003)	1 NA (0.00-NA)	0.21
rs10255943	Sobredominancia	GG-AA GA	188 (0.60) 126 (0.40)	149 (0.57) 114 (0.43)	1 1.14 (0.82-1.59)	0.43
rs10486026	Recesivo	TT-CT CC	293 (0.93) 22 (0.07)	252 (0.96) 11 (0.04)	1 0.58 (0.28-1.22)	0.14
rs2396753	Recesivo	AA-CA CC	285 (0.84) 55 (0.16)	226 (0.77) 67 (0.23)	1 1.54 (1.03-2.29)	0.03 (0.72)
rs17137124	Dominante	TT CT-CC	84 (0.25) 256 (0.75)	86 (0.29) 207 (0.71)	1 0.79 (0.56-1.12)	0.19
	Aditivo				0.86 (0.69-1.07)	0.18
rs7799652	Dominante	TT TG-GG	87 (0.27) 232 (0.73)	87 (0.33) 178 (0.67)	1 0.77 (0.54-1.09)	0.14
rs1456029	Aditivo				1.19 (0.91-1.55)	0.2
rs12670585	Recesivo	CC-CT TT	282 (0.90) 33 (0.10)	241 (0.92) 21 (0.08)	1 0.74 (0.42-1.32)	0.31
rs1456031	Dominante	TT TC-CC	98 (0.29) 242 (0.71)	101 (0.34) 192 (0.66)	1 0.77 (0.55-1.08)	0.13
rs2396765	Recesivo	TT-CT CC	260 (0.82) 57 (0.18)	227 (0.86) 36 (0.14)	1 0.72 (0.46-1.14)	0.16
rs1456021	Recesivo	TT-GT GG	274 (0.82) 60 (0.18)	243 (0.87) 35 (0.13)	1 0.66 (0.42-1.03)	0.07

Tabla R6. Resultados de los análisis de regresión para el test de asociación entre controles y pacientes. En el valor de la OR con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo. En aquellos casos en los que el menor valor BIC estimado es el mismo para diferentes modelos de herencia se incluyen los resultados correspondientes a dichos modelos.

(*) corresponde al valor P corregido por Bonferroni.

SNPs	Modelo de herencia	Genotipos	Controles	Pacientes con AA	OR	P (*)
rs7803667	Dominante	TT AT-AA	169 (0.50) 171 (0.50)	93 (0.43) 122 (0.57)	1 1.30 (0.92-1.83)	0.14
rs10447760	Dominante	CC CT-TT	193 (0.57) 147 (0.43)	104 (0.48) 111 (0.52)	1 1.40 (0.99-1.97)	0.05 (1)
rs6961558	No determinado	GG GA	323 (0.95) 16 (0.05)	207 (0.96) 8 (0.04)	1 0.78 (0.33-1.86)	0.57
rs923875	Dominante	AA AC-CC	125 (0.37) 215 (0.63)	73 (0.34) 142 (0.66)	1 1.13 (0.79-1.62)	0.5
rs1597548	Recesivo	CC-CG GG	339 (0.997) 1 (0.003)	215 (1.0) 0	1 0.00 (0.00-NA)	0.32
rs10500038	Recesivo	GG-GA AA	312 (0.98) 7 (0.02)	190 (0.95) 10 (0.05)	1 2.35 (0.88-6.27)	0.09
rs4730626	Aditivo				0.83 (0.61-1.13)	0.22
rs1668335	Recesivo	GG-GA AA	288 (0.91) 27 (0.09)	185 (0.925) 15 (0.075)	1 0.86 (0.45-1.67)	0.66
rs11771168	Dominante	CC CT-TT	192 (0.60) 127 (0.40)	126 (0.63) 74 (0.37)	1 0.89 (0.62-1.28)	0.52
	Sobredominancia	CC-TT CT	207 (0.65) 112 (0.35)	135 (0.675) 65 (0.325)	1 0.89 (0.61-1.29)	0.54
	Aditivo				0.91 (0.67-1.24)	0.56
rs1916977	Recesivo	AA-AG GG	290 (0.92) 24 (0.08)	188 (0.94) 12 (0.06)	1 0.77 (0.38-1.58)	0.47
rs2396722	Recesivo	TT-CT CC	291 (0.86) 49 (0.14)	186 (0.87) 29 (0.13)	1 0.93 (0.56-1.52)	0.76
rs2253478	Recesivo	GG-AG AA	266 (0.84) 51 (0.16)	170 (0.85) 29 (0.15)	1 0.89 (0.54-1.46)	0.64
rs2694941	Dominante	TT AT-AA	110 (0.35) 205 (0.65)	66 (0.33) 132 (0.67)	1 1.07 (0.74-1.56)	0.71
	Sobredominancia	TT-AA TA	171 (0.54) 144 (0.46)	104 (52.5%) 94 (47.5%)	1 1.07 (0.75-1.53)	0.7
rs1852469	Recesivo	AA-TA TT	340 (1.0) 0	214 (0.995) 1 (0.005)	1 NA (0-NA)	0.17
rs10255943	Sobredominancia	GG-AA GA	188 (0.60) 126 (0.40)	113 (0.565) 87 (0.435)	1 1.15 (0.80-1.65)	0.45
rs10486026	Recesivo	TT-CT CC	293 (0.93) 22 (0.07)	193 (0.965) 7 (0.035)	1 0.48 (0.20-1.15)	0.09
rs2396753	Aditivo				1.48 (1.14-1.91)	0.003 (0.072)
rs17137124	Aditivo				0.78 (0.61-0.99)	0.04 (0.96)
rs7799652	Aditivo				0.81 (0.62-1.05)	0.11
rs1456029	Aditivo				1.23 (0.93-1.63)	0.15
rs12670585	Recesivo	CC-CT TT	282 (0.90) 33 (0.10)	184 (0.92) 15 (0.08)	1 0.70 (0.37-1.32)	0.26
rs1456031	Aditivo				0.80 (0.63-1.02)	0.07
rs2396765	Recesivo	TT-CT CC	260 (0.82) 57 (0.18)	174 (0.88) 24 (0.12)	1 0.63 (0.38-1.05)	0.07
rs1456021	Recesivo	TT-GT GG	274 (0.82) 60 (0.18)	188 (0.89) 24 (0.11)	1 0.58 (0.35-0.97)	0.03 (0.72)

Tabla R7. Resultados de los análisis de regresión para el test de asociación entre controles y pacientes con AA. En el valor de la OR con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo. En aquellos casos en los que el menor valor BIC estimado es el mismo para diferentes modelos de herencia se incluyen los resultados correspondientes a dichos modelos.

(*) corresponde al valor P corregido por Bonferroni.

Resultados

En cuanto a la comparación mediante regresión logística de las frecuencias genotípicas de controles y pacientes con esquizofrenia (datos no mostrados), al igual que en el caso de las tablas de contingencia, se obtuvieron diferencias significativas para los polimorfismos rs10447760 (modelo aditivo, $P=0.008$) y rs2396753 (modelo de sobredominancia, $P=0.014$), aunque tras la corrección múltiple se pierde la significación ($P=0.192$ y $P=0.336$, respectivamente).

Con el programa *SNPStats* también se analizó la existencia de diferencias en las frecuencias genotípicas de los diferentes polimorfismos entre mujeres y varones, tanto en el grupo control, como en la muestra de pacientes con AA y la muestra total de pacientes. En algunos casos se han encontrado diferencias significativas (valores de $P \leq 0.05$), pero estas diferencias desaparecen al aplicar la corrección de Bonferroni.

1.1.3 Análisis de haplotipos en pacientes y controles

Debido al elevado número de polimorfismos analizados, la estima de haplotipos con todos ellos no se podía aplicar de una forma que permitiera obtener resultados robustos. Por esta razón, se recurrió al análisis por medio de ventanas de 4 SNPs consecutivos, con el programa *UNPHASED v3.12*. No se tuvieron en cuenta los haplotipos con frecuencia menor a 0.03 y se eliminaron del análisis los polimorfismos con $MAF < 0.05$ en la muestra utilizada: los SNPs rs6169558, rs1597548, y rs1852469.

De manera excepcional, se analizaron dos haplotipos compuestos por 5 y 7 SNPs, para los que se habían encontrado diferencias significativas entre pacientes con AA y controles un estudio anterior con menor tamaño muestral (Sanjuán *et al.*, 2006b). Estos haplotipos fueron: rs7803667/ rs10447760/ rs923875/ rs2396722/ rs2396753 y rs7803667/ rs10447760/ rs923875/ rs2396722/ rs2396753/ rs17137124/ rs1456031. En el estudio anterior se utilizó el polimorfismo rs1358278 en lugar de rs2396722. Puesto que ambos polimorfismos se encuentran en completo LD, la utilización de rs2396722 no afecta a los resultados.

En el análisis por ventanas, al comparar tanto pacientes frente a controles, pacientes con AA frente a controles como pacientes con esquizofrenia frente a controles, no se obtuvieron diferencias significativas para ninguno de los test globales aplicados a las combinaciones de 4 SNPs consecutivos.

Considerados los test individuales, se obtuvieron diferencias significativas en algunos de ellos, mostrados en las tablas R8 y R9 (datos de comparación entre pacientes con esquizofrenia y controles no mostrados). Sin embargo, la significación no se mantuvo al corregir por Bonferroni dentro de los haplotipos obtenidos de la combinación de 4 SNPs correspondiente.

SNPs del haplotipo	Test Global			Test individual				
	df	χ^2	P	Haplotipo	Frec. Pacientes	Frec. controles	χ^2	P (*)
rs7803667 rs10447760 rs923875 rs10500038	4	4.153	0.3857	T-C-A-G	0.373	0.434	4.136	0.0420 (0.21)
rs7799652 rs1456029 rs12670585 rs1456031	6	11.36	0.0779	G-G-T-C	0.030	0.008	4.651	0.0310 (0.22)
rs12670585 rs1456031 rs2396765 rs1456021	4	7.869	0.0965	C-T-T-T	0.550	0.484	4.426	0.0354 (0.177)

Tabla R8. Análisis de haplotipos de pacientes respecto a controles. Límite para la frecuencia de haplotipos raros = 0.03. Haplotipos obtenidos en ventanas de 4 SNPs.

(*) corresponde al valor P corregido por Bonferroni para el número de haplotipos.

SNPs del haplotipo	Test Global			Test individual				
	df	χ^2	P	Haplotipo	Frec. Pacientes con AA	Frec. Controles	χ^2	P (*)
rs10255943 rs10486026 rs2396753 rs17137124	5	8.061	0.1529	G-T-C-T	0.166	0.129	4.018	0.0450 (0.27)
rs10486026 rs2396753 rs17137124 rs7799652	6	7.495	0.2775	T-C-T-T	0.414	0.355	4.195	0.0405 (0.28)
rs2396753 rs17137124 rs7799652 rs1456029	8	11.12	0.1952	A-C-G-A	0.271	0.351	5.7	0.0170 (0.15)
rs12670585 rs1456031 rs2396765 rs1456021	4	5.25	0.2626	C-T-T-T	0.550	0.4836	3.892	0.0485 (0.24)

Tabla R9. Análisis de haplotipos de pacientes con alucinaciones auditivas respecto a controles. Límite para la frecuencia de haplotipos raros = 0.03. Haplotipos obtenidos en ventanas de 4 SNPs.

(*) corresponde al valor P corregido por Bonferroni para el número de haplotipos.

En cuanto al análisis de haplotipos de la combinación de SNPs anteriormente especificada, en las tablas R10 y R11 se muestran los resultados obtenidos para la comparación pacientes con AA frente a controles. Sombreados en color azul se muestran los haplotipos para los que se obtenía diferencias significativas en el estudio anterior (Sanjuán *et al.*, 2006b).

Tanto para los haplotipos de 5 SNPs como los de 7, se obtuvieron diferencias significativas para los tests globales: $P=0.00678$ y $P=0.0064$ respectivamente. En los tests individuales, se observaron diferencias entre pacientes con AA y controles para los mismos haplotipos que en el estudio realizado con menor número de muestras: el haplotipo rs7803667T/ rs10447760C/ rs923875A/ rs2396722C/ rs2396753C ($P=0.0059$), el haplotipo rs7803667T/ rs10447760C/ rs923875A/ rs2396722C/

Resultados

rs2396753A (P=0.0109), y el haplotipo rs7803667T/ rs10447760C/ rs923875A/ rs2396722C/ rs2396753C/ rs17137124T/ rs1456031T (P=0.0249) aunque al corregir por Bonferroni no se mantenían (P=0.059, P=0.0981, P= 0.149, respectivamente). Así, pues, aunque estos resultados muestran la misma tendencia que los obtenidos anteriormente, no mantienen la significación obtenida.

rs7803667	rs10447760	rs923875	rs2396722	rs2396753	rs17137124	rs1456031	Controles	Pacientes con AA	Odds-Ratio (IC 95%)	χ^2	P (*)
T	C	A	C	A	C	C	0.187	0.153	1	2.868	0.0904
A	T	C	T	C	T	T	0.114	0.135	1.441 (0.875-2.373)	2.835	0.0922
T	C	A	T	C	T	T	0.117	0.139	1.465 (0.885-2.423)	0.9728	0.324
T	C	A	T	A	C	C	0.053	0.034	0.789 (0.341-1.832)	1.801	0.1795
T	C	A	C	A	C	T	0.054	0.030	0.689 (0.254-1.868)	0.861	0.3536
T	C	A	T	C	T	C	0.034	0.028	1.019 (0.402-2.582)	0.014	0.9068
A	T	C	T	A	C	C	0.036	0.006	0.205 (0.025-1.654)	4.361	0.0368
T	C	A	C	C	T	T	0.022	0.068	3.838 (1.606-9.17)	5.025	0.0249 (0.224)
T	C	A	T	A	T	C	0.028	0.032	1.383 (0.552-3.463)	0.036	0.849
Otros haplotipos							0.355	0.375			

Tabla R10. Análisis de haplotipos al comparar pacientes con AA con controles. Límite para la frecuencia de haplotipos raros = 0.03. Test de asociación global: $\chi^2=21.3$, df=8, P=0.0064.

(*) Valor de P corregido por Bonferroni para el número de haplotipos comparados.

rs7803667	rs10447760	rs923875	rs2396722	rs2396753	Controles	Pacientes con AA	Odds-Ratio (IC 95%)	χ^2	P (*)
T	C	A	C	A	0.269	0.174	1	6.479	0.0109 (0.098)
T	C	A	T	C	0.173	0.193	1.718 (1.088-2.713)	0.491	0.4833
A	T	C	T	C	0.150	0.170	1.752 (1.137-2.698)	2.287	0.1305
T	C	A	T	A	0.130	0.130	1.549 (0.925-2.595)	0.576	0.4479
A	T	C	T	A	0.067	0.036	0.844 (0.373-1.911)	2.789	0.0949
T	C	C	T	A	0.039	0.036	1.427 (0.612-3.327)	0.955	0.3285
T	C	A	C	C	0.034	0.079	3.583 (1.703-7.54)	7.579	0.0059 (0.053)
T	C	C	T	C	0.033	0.017	0.791 (0.236-2.653)	0.551	0.4562
A	T	C	C	A	0.016	0.046	4.505 (1.445-14.05)	2.809	0.0937
Otros haplotipos					0.089	0.119			

Tabla R11. Análisis de haplotipos al comparar pacientes con AA con controles. Límite para la frecuencia de haplotipos raros = 0.03. Test de asociación global: $\chi^2=21.14$, df=8, P=0.00678.

(*) Valor de P corregido por Bonferroni para el número de haplotipos comparados.

Los resultados de la comparación de la distribución de frecuencias alélicas y genotípicas en los grupos muestrales analizados no muestran evidencias de diferencias significativas entre ambos grupos cuando se aplican las correcciones múltiples más conservativas (corrección de Bonferroni). Sin embargo, a lo largo de los análisis hay un dato que conviene recalcar, que consiste en la asociación del polimorfismo rs2396753 con las AA. Al comparar pacientes con controles, este polimorfismo no mostraba diferencias significativas entre ambos grupos, lo cual probablemente se debiera a la presencia de no alucinadores en la muestra, sin embargo, al comparar únicamente pacientes con AA frente a controles, se obtienen diferencias significativas, con una distribución de frecuencias alélicas y genotípicas claramente diferentes entre ambos grupos, a pesar de no soportar la corrección por Bonferroni (figura R6).

Respecto a los resultados obtenidos para el análisis de haplotipos en pacientes o pacientes con AA frente a controles, el análisis por ventanas de SNPs no mostró presencia de diferencias significativas entre los grupos estudiados. Este tipo de análisis permite que los haplotipos que se obtienen sean más robustos, pero también supone una pérdida de información sobre los haplotipos compuestos por SNPs que no se encuentran en una misma ventana. Esto se ve reflejado en el análisis de un haplotipo específico para el que se había obtenido diferencias significativas en un estudio anterior, que se analizó de forma independiente debido a que la combinación de SNPs utilizados no se encontraba en una ventana común. Para estos haplotipos, las diferencias globales de haplotipos entre pacientes con AA y controles resultaron significativas, y en el caso de haplotipos individuales se mantuvo la tendencia, si bien no la significación.

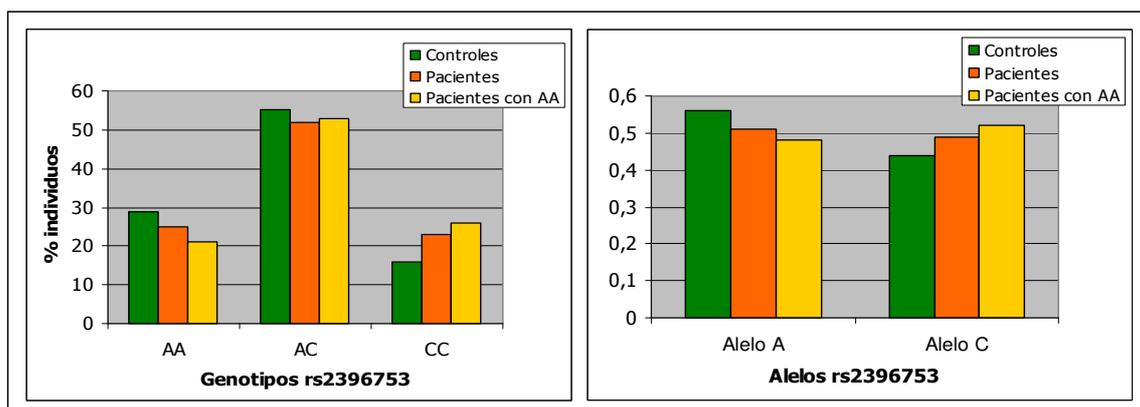


Figura R6. Frecuencias genotípicas y alélicas del polimorfismo rs2396753 en pacientes, pacientes con AA y controles.

Resultados

1.1.4 Medida del poder estadístico

En un estudio de asociación, es importante conocer el poder estadístico de los datos que se utilizan. La estima del poder estadístico depende de diferentes factores: el tamaño muestral, el modelo de herencia, la frecuencia de la enfermedad en la población, las frecuencias alélicas y la OR obtenida. Esta estima debe realizarse antes del inicio de la genotipación para ver el poder de predicción que tienen los SNPs que se van a utilizar. Sin embargo, en nuestro caso, puesto que para gran parte de los SNPs se desconocían datos de frecuencias genotípicas en poblaciones similiares o no había ninguna información al respecto, este análisis se llevo a cabo después, encaminado a obtener el poder estadístico de los SNPs seleccionados en lugar de tomarlo como base para la selección de los mismos.

Para cuantificar el poder estadístico, se utilizó el programa *Quanto v1.2.3*.

En las siguientes tablas se observa para cada polimorfismo el poder estadístico asociado al test, teniendo en cuenta las frecuencias alélicas y el modelo de herencia seleccionado con el programa *SNPStats*. Para determinar el poder estadístico se consideraron los valores de OR obtenidos con el mismo programa.

SNPs	Modelo de herencia	Poder estadístico en el test controles (N=340) / pacientes (N=293)
rs7803667	Dominante	0.4312
rs10447760	Dominante	0.6422
rs6961558	Aditivo	0.1282
	Dominante	0.1262
	Recesivo	0.0508
rs923875	Dominante	0.22
rs1597548	Recesivo	nd
rs10500038	Recesivo	0.2927
rs4730626	Aditivo	0.2850
rs1668335	Sobredominancia	0.0884
rs11771168	Sobredominancia	0.0929
rs1916977	Recesivo	0.0802
rs2396722	Recesivo	0.0822
	Sobredominancia	0.1010
rs2253478	Sobredominancia	0.4647
rs2694941	Sobredominancia	0.0913
rs1852469	Recesivo	nd
rs10255943	Sobredominancia	0.1626
rs10486026	Recesivo	0.2528
rs2396753	Recesivo	0.6230
rs17137124	Dominante	0.2654
	Aditivo	0.2664
rs7799652	Dominante	0.3328
rs1456029	Aditivo	0.2641
rs12670585	Recesivo	0.1737
rs1456031	Dominante	0.3275
rs2396765	Recesivo	0.3070
rs1456021	Recesivo	0.4557

Tabla R12. Estimas del poder estadístico de la comparación de controles frente a pacientes, para las OR obtenidas mediante el programa *SNPStats*. Aquellos casos en los que no se pudo calcular el poder estadístico debido a no disponer de OR se muestran como nd.

SNPs	Modelo de herencia	Poder estadístico en el test controles (N=340) / pacientes con AA (N=215)
rs7803667	Dominante	0.3234
rs10447760	Dominante	0.4864
rs6961558	Aditivo	0.1022
	Dominante	0.1009
	Recesivo	0.050
rs923875	Dominante	0.1031
rs1597548	Recesivo	nd
rs10500038	Recesivo	0.5663
rs4730626	Aditivo	0.2230
rs1668335	Recesivo	0.0642
rs11771168	Dominante	0.0992
	Sobredominancia	0.1201
	Aditivo	0.0951
rs1916977	Recesivo	0.1049
rs2396722	Recesivo	0.0588
rs2253478	Recesivo	0.0745
rs2694941	Dominante	0.0654
	Sobredominancia	0.0844
rs1852469	Recesivo	nd
rs10255943	Sobredominancia	0.1860
rs10486026	Recesivo	0.3270
rs2396753	Aditivo	0.8860
rs17137124	Aditivo	0.5179
rs7799652	Aditivo	0.3944
rs1456029	Aditivo	0.3069
rs12670585	Recesivo	0.1912
rs1456031	Aditivo	0.4346
rs2396765	Recesivo	0.4438
rs1456021	Recesivo	0.5769

Tabla R13. Estimaciones del poder estadístico de la comparación de controles frente a pacientes con AA, para OR obtenidos mediante el programa *SNPStats*. Aquellos casos en los que no se pudo calcular el poder estadístico debido a no disponer de OR se muestran como nd.

El poder estadístico de nuestro estudio varía considerablemente en el rango de ORs estimadas para los SNPs. En general, la mayoría de los valores obtenidos son bajos, por lo que la potencia de detectar asociación en nuestra muestra es relativamente baja. Cabe destacar que el polimorfismo para el cual se detectaron mayores diferencias entre los grupos estudiados, el SNP rs2396753, es uno de los cuales para los que se obtiene mayor poder estadístico.

Resultados

1.1.5 Análisis de variables clínicas

Como se especifica en el apartado I.1.1.1.3 de la sección de Material y Métodos, además de analizar las frecuencias alélicas y genotípicas de los grupos muestrales ya mencionados, se analizaron otras variables clínicas relacionadas, tanto con las alucinaciones auditivas, en las que se centra este trabajo, como en otros aspectos del desarrollo de la esquizofrenia.

1.1.5.1 Variables clínicas con dos estados

El análisis de estas variables se llevó a cabo del mismo modo que para las comparaciones entre controles-pacientes con AA, controles-pacientes y controles-esquizofrénicos.

A continuación se muestran los análisis relativos a las variables con dos estados:

Variable	Estado de la variable	Número de individuos
Presencia de AA	Con historia de AA	215
	Sin historia de AA	77
Cronicidad de las AA	Presencia de cronicidad	80
	Ausencia de cronicidad	135
Inicio de la enfermedad	Inicio temprano (hasta los 20 años)	110
	Inicio tardío (a partir de los 28 años)	89

Tabla R14. Variables clínicas dicotómicas analizadas.

Con el fin de proporcionar mayor claridad a la exposición de resultados, únicamente se presentan en formato de tabla los resultados correspondientes a la presencia o ausencia de AA, debido al interés central de este trabajo en las mismas y en su relación con el lenguaje (Tabla R15).

Al comparar pacientes con AA, con pacientes sin éste síntoma mediante la obtención de tablas de contingencia, se encontraron diferencias significativas en ambos grupos para las frecuencias de los polimorfismos rs2253478 ($\chi^2=6.487$, $P=0.039$), rs2396753 ($\chi^2=6.611$, $P=0.037$) y rs1456031 ($\chi^2=6.611$, $P=0.037$), aunque estas diferencias se pierden al llevar a cabo la corrección por Bonferroni ($P=0.936$, $P=0.888$, $P=0.888$, respectivamente). Respecto a las frecuencias alélicas, únicamente se encontraron diferencias en el segundo de ellos, rs2396753 ($\chi^2=6.605$, $P=0.010$). Nuevamente, al corregir por Bonferroni se pierde la significación ($P=0.24$).

Este polimorfismo es el mismo para el que se había obtenido diferencias al comparar pacientes con AA frente a controles. En este caso se observa que los valores obtenidos para los pacientes sin AA (frecuencias alélicas de 0.60 para el alelo A y 0.40 para el alelo C), se acercan más a los obtenidos en controles (0.56 y 0.44) que a los obtenidos en el conjunto de los pacientes (0.51 y 0.49) o pacientes con AA (0.48 y 0.52). Sin

embargo, debido al pequeño número de individuos incluido en el grupo de pacientes sin AA, los resultados obtenidos de esta comparación deben manejarse con precaución.

Respecto a la cronicidad de las AA (datos no mostrados), al llevar a cabo la comparación mediante tablas de contingencia se obtuvieron diferencias significativas en las frecuencias genotípicas en los polimorfismos rs1456031 ($\chi^2=9.824$, $P=0.007$), rs2396765 ($\chi^2=7.484$, $P=0.023$) y rs1456021 ($\chi^2=9.748$, $P=0.007$). Al corregir por Bonferroni, estas diferencias dejan de ser significativas ($P=0.168$, $P=0.552$ y $P=0.168$, respectivamente). Estos tres polimorfismos también muestran diferencias significativas al comparar las frecuencias alélicas: rs1456031 ($\chi^2=8.948$, $P=0.0028$), rs2396765 ($\chi^2=5.487$, $P=0.0191$) y rs1456021 ($\chi^2=8.083$, $P=0.0045$), a los que se añade rs6961558 ($\chi^2=4.831$, $P=0.0279$). Sin embargo, nuevamente, al corregir por Bonferroni se pierde la significación ($P=0.067$, $P=0.459$, $P=0.107$ y $P=0.669$, respectivamente).

Al comparar pacientes con inicio temprano de la enfermedad (primer episodio antes de los 20 años) con pacientes con inicio tardío (primer episodio después de cumplir los 28 años), no se encontraron diferencias significativas para ninguno de los polimorfismos en las frecuencias genotípicas o alélicas (datos no mostrados).

Al igual que en las comparaciones entre los diferentes grupos estudiados, para las variables clínicas con dos estados el programa *SNPStats* permitió también, mediante regresión logística, testar qué modelo de herencia se ajustaba más a las proporciones observadas en los grupos.

En la tabla R16 se muestra para cada polimorfismo qué modelo se ajustó más en las comparaciones realizadas, esto es, qué modelo obtiene un menor valor para BIC, con las OR y valores de P asociados.

Al comparar pacientes con AA frente a pacientes sin este síntoma por regresión logística, se obtuvieron diferencias significativas tanto para los polimorfismos que ya presentaban tendencias en el estudio mediante tablas de contingencia, rs2253478 ($P=0.01$), rs2396753 ($P=0.008$) y rs1456031 ($P=0.03$), como para los polimorfismos rs10500038 ($P=0.02$) y rs1668335 ($P=0.04$). No obstante la significación se pierde al aplicar la corrección por Bonferroni ($P=0.24$, $P=0.19$, $P=0.72$, $P=0.48$ y $P=0.96$, respectivamente).

Respecto a la cronicidad de las AA (datos no mostrados), cuando se compararon pacientes con AA crónicos respecto a pacientes con AA no crónicos se obtuvo diferencias significativas para los polimorfismos para los cuales se obtuvieron también en las tablas de contingencia, rs2396765 (modelo dominante $P=0.0056$), rs1456031 (modelo dominante $P=0.002$) y rs1456021 (modelo dominante $P=0.0018$), así como para el polimorfismo rs10500038 (modelo recesivo $P=0.035$). Al corregir por Bonferroni, únicamente rs1456031 ($P=0.048$) y rs1456021 ($P=0.043$) mantienen las diferencias significativas. Para rs2396765 y rs10500038, los valores corregidos son $P=0.134$ y $P=0.84$, respectivamente.

En cuanto a la comparación mediante regresión logística entre pacientes con inicio temprano e inicio tardío de la enfermedad (datos no mostrados), no se obtuvieron diferencias significativas para ninguno de los polimorfismos analizados.

Resultados

SNP	Frecuencias genotípicas			χ^2	P (*)	Frecuencias alélicas		χ^2	P
	TT	TA	AA			T	A		
rs7803667									
Pacientes sin AA	31 (0.4)	39 (0.51)	7 (0.09)	0.398	0.839	0.66	0.34	0.044	0.835
Pacientes con AA	93 (0.43)	100 (0.47)	22 (0.1)			0.67	0.33		
rs10447760									
Pacientes sin AA	34 (0.44)	37 (0.48)	6 (0.08)	0.410	0.836	0.68	0.31	0.342	0.559
Pacientes con AA	104 (0.48)	96 (0.45)	15 (0.07)			0.71	0.29		
rs6961558									
Pacientes sin AA (N = 76)	74 (0.97)	2 (0.03)	0	0.201^c	0.738	0.99	0.01	0.197	0.657
Pacientes con AA (N = 215)	207 (0.96)	8 (0.04)	0			0.98	0.02		
rs923875									
Pacientes sin AA	21 (0.27)	44 (0.57)	12 (0.16)	1.238	0.546	0.56	0.44	0.558	0.455
Pacientes con AA	73 (0.34)	109 (0.51)	33 (0.15)			0.59	0.41		
rs1597548									
Pacientes sin AA	73 (0.95)	4 (0.05)	0	2.927^c	0.062	0.97	0.03	2.768	0.096
Pacientes con AA	189 (0.88)	26 (0.12)	0			0.94	0.06		
rs10500038									
Pacientes sin AA (N = 64)	44 (0.69)	20 (0.31)	0	3.458^a	0.196	0.84	0.16	1.771	0.183
Pacientes con AA (N = 200)	126 (0.63)	64 (0.32)	10 (0.05)			0.79	0.21		
rs4730626									
Pacientes sin AA (N = 61)	41 (0.67)	19 (0.31)	1 (0.02)	0.986^a	0.599	0.83	0.17	0.019	0.892
Pacientes con AA (N = 200)	137 (0.68)	55 (0.28)	8 (0.04)			0.82	0.18		
rs1668335									
Pacientes sin AA (N = 62)	37 (0.6)	17 (0.27)	8 (0.13)	4.704	0.097	0.73	0.27	0.126	0.722
Pacientes con AA (N = 200)	102 (0.51)	83 (0.42)	15 (0.08)			0.72	0.28		
rs11771168									
Pacientes sin AA (N = 64)	38 (0.59)	22 (0.34)	4 (0.06)	0.456^a	0.804	0.77	0.23	0.416	0.519
Pacientes con AA (N = 200)	126 (0.63)	65 (0.32)	9 (0.04)			0.79	0.21		
rs1916977									
Pacientes sin AA (N = 62)	36 (0.58)	21 (0.34)	5 (0.08)	0.371^a	0.895	0.75	0.25	0.052	0.820
Pacientes con AA (N = 200)	116 (0.58)	72 (0.36)	12 (0.06)			0.76	0.24		
rs2396722									
Pacientes sin AA	30 (0.39)	38 (0.49)	9 (0.12)	0.628	0.781	0.64	0.36	0.03	0.862
Pacientes con AA	91 (0.42)	95 (0.44)	29 (0.13)			0.64	0.36		
rs2253478									
Pacientes sin AA (N = 64)	18 (0.28)	40 (0.62)	6 (0.09)	6.487	0.039 (0.936)	0.59	0.41	0.641	0.423
Pacientes con AA (N = 199)	82 (0.41)	88 (0.44)	29 (0.15)			0.63	0.37		

Tabla R15. Frecuencias genotípicas y alélicas de pacientes con AA y pacientes sin AA con los valores de χ^2 y P asociados. ^a corresponden a tests en los que una de las clases los valores esperados son menores a 5.

^b corresponden a tests en los que una de las clases los valores esperados son menores a 2.

^c corresponde a tests en los que debido a la falta de clases, el test χ^2 se realizó sobre una tabla 2x2 en lugar de 3x2.

(*) corresponde al valor P corregido por Bonferroni.

Resultados

SNP	Frecuencias genotípicas			χ^2	P (*)	Frecuencias alélicas		χ^2	P (*)
	TT	TA	AA			T	A		
rs2694941									
Pacientes sin AA (N = 62)	23 (0.37)	29 (0.47)	10 (0.16)	0.440	0.809	0.60	0.40	0.451	0.502
Pacientes con AA (N = 198)	66 (0.33)	94 (0.47)	38 (0.19)			0.57	0.43		
rs1852469									
Pacientes sin AA	71 (0.92)	6 (0.08)	0	0.361^b	1.000	0.96	0.04	0.076	0.783
Pacientes con AA	197 (0.92)	17 (0.08)	1 (0.005)			0.96	0.04		
rs10255943									
Pacientes sin AA (N = 62)	30 (0.48)	27 (0.44)	5 (0.08)	0.126	0.954	0.70	0.30	0.088	0.766
Pacientes con AA (N = 200)	94 (0.47)	87 (0.44)	19 (0.1)			0.69	0.31		
rs10486026									
Pacientes sin AA (N = 62)	37 (0.6)	21 (0.34)	4 (0.6)	1.068^a	0.636	0.77	0.23	0.562	0.454
Pacientes con AA (N = 200)	126 (0.63)	67 (0.34)	7 (0.04)			0.79	0.21		
rs2396753									
Pacientes sin AA	27 (0.35)	38 (0.49)	12 (0.16)	6.611	0.037 (0.888)	0.60	0.40	6.605	0.010 (0.24)
Pacientes con AA	45 (0.21)	115 (0.53)	55 (0.26)			0.48	0.52		
rs17137124									
Pacientes sin AA	18 (0.23)	36 (0.47)	23 (0.3)	3.512	0.169	0.47	0.53	3.55	0.059
Pacientes con AA	68 (0.32)	103 (0.48)	44 (0.2)			0.56	0.44		
rs7799652									
Pacientes sin AA (N = 64)	21 (0.33)	27 (0.42)	16 (0.25)	3.021	0.222	0.54	0.46	0.837	0.360
Pacientes con AA (N = 200)	66 (0.33)	102 (0.51)	32 (0.16)			0.58	0.42		
rs1456029									
Pacientes sin AA (N = 64)	36 (0.56)	25 (0.39)	3 (0.05)	1.030^a	0.630	0.76	0.24	0.322	0.571
Pacientes con AA (N = 200)	110 (0.55)	73 (0.36)	17 (0.08)			0.73	0.27		
rs12670585									
Pacientes sin AA (N = 62)	32 (0.52)	24 (0.39)	6 (0.1)	0.343^a	0.837	0.71	0.29	0.037	0.848
Pacientes con AA (N = 199)	102 (0.51)	82 (0.41)	15 (0.08)			0.72	0.28		
rs1456031									
Pacientes sin AA	28 (0.36)	27 (0.35)	22 (0.29)	6.611	0.037 (0.888)	0.54	0.46	1.028	0.311
Pacientes con AA	73 (0.34)	106 (0.49)	36 (0.17)			0.59	0.41		
rs2396765									
Pacientes sin AA (N = 64)	26 (0.41)	27 (0.42)	11 (0.17)	1.754	0.435	0.62	0.38	0.033	0.854
Pacientes con AA (N = 198)	74 (0.37)	100 (0.51)	24 (0.12)			0.63	0.37		
rs1456021									
Pacientes sin AA (N = 65)	27 (0.42)	28 (0.43)	10 (0.15)	1.266	0.536	0.63	0.37	0.016	0.901
Pacientes con AA (N = 212)	82 (0.39)	106 (0.5)	24 (0.11)			0.64	0.36		

Tabla R15. Continuación.

Resultados

SNPs	Modelo de herencia	Genotipos	Pacientes con AA	Pacientes sin AA	OR	P (*)
rs7803667	Dominante	TT	93 (0.43)	31 (0.40)	1	0.65
		AT-AA	122 (0.57)	46 (0.60)	1.13 (0.67-1.92)	
rs10447760	Dominante	CC	104 (0.48)	34 (0.44)	1	0.52
	Aditivo	CT-TT	111 (0.52)	43 (0.56)	1.18 (0.70-2.00)	
rs6961558	No determinado	GG	207 (0.96)	74 (0.97)	1	0.65
		GA	8 (0.04)	2 (0.03)	0.70 (0.15-3.37)	
rs923875	Dominante	AA	73 (0.34)	21 (0.27)	1	0.28
		AC-CC	142 (0.66)	56 (0.73)	1.37 (0.77-2.44)	
rs1597548	No determinado	CC	189 (0.88)	73 (0.95)	1	0.07
		CG	26 (0.12)	4 (0.05)	0.40 (0.13-1.18)	
rs10500038	Recesivo	GG-GA	190 (0.95)	64 (1.0)	1	0.02 (0.48)
		AA	10 (0.05)	0	0.00 (0.00-NA)	
rs4730626	Recesivo	GG-GA	192 (0.96)	60 (0.98)	1	0.34
		AA	8 (0.04)	1 (0.02)	0.40 (0.05-3.26)	
rs717233	Sobredominancia	TT-CC	126 (0.63)	46 (0.72)	1	0.21
		TC	73 (0.37)	18 (0.28)	0.68 (0.36-1.25)	
rs1668335	Sobredominancia	GG-AA	117 (0.585)	45 (0.73)	1	0.04 (0.96)
		GA	83 (0.415)	17 (0.27)	0.53 (0.29-0.99)	
rs11771168	Aditivo				1.17 (0.73-1.87)	0.53
rs1916977	Recesivo	AA-AG	188 (0.94)	57 (0.92)	1	0.57
		GG	12 (0.06)	5 (0.08)	1.37 (0.46-4.07)	
rs2396722	Sobredominancia	TT-CC	120 (0.56)	39 (0.51)	1	0.44
		TC	95 (0.44)	38 (0.49)	1.23 (0.73-2.07)	
rs2253478	Sobredominancia	GG-AA	111 (0.56)	24 (0.375)	1	0.01 (0.24)
		GA	88 (0.44)	40 (0.325)	2.10 (1.18-3.75)	
rs2694941	Aditivo				0.87 (0.58-1.31)	0.51
rs1852469	Recesivo	AA-TA	214 (0.995)	77 (1.0)	1	0.43
		TT	1 (0.005)	0	0.00 (0.00-NA)	
rs10255943	Recesivo	GG-GA	181 (0.905)	57 (0.92)	1	0.73
		AA	19 (0.095)	5 (0.08)	0.84 (0.30-2.34)	
rs10486026	Recesivo	TT-CT	193 (0.965)	58 (0.935)	1	0.33
		CC	7 (0.035)	4 (0.065)	1.90 (0.54-6.72)	
rs2396753	Aditivo				0.60 (0.40-0.88)	0.008 (0.19)
rs17137124	Aditivo				1.41 (0.98-2.03)	0.07
rs7799652	Recesivo	TT-GT	168 (0.84)	48 (0.75)	1	0.11
		GG	32 (0.16)	16 (0.25)	1.75 (0.89-3.46)	
rs1456029						
rs12670585	Recesivo	CC-CT	184 (0.925)	56 (0.90)	1	0.6
		TT	15 (0.075)	6 (0.97)	1.31 (0.49-3.55)	
rs1456031	Recesivo	TT-TC	179 (0.83)	55 (0.71)	1	0.03 (0.72)
		CC	36 (0.17)	22 (0.29)	1.99 (1.08-3.66)	
rs2396765	Sobredominancia	TT-CC	98 (0.495)	37 (0.58)	1	0.25
		TC	100 (0.505)	27 (0.42)	0.72 (0.40-1.26)	
rs1456021	Sobredominancia	TT-GG	106 (0.50)	37 (0.57)	1	0.33
		TG	106 (0.50)	28 (0.43)	0.76 (0.43-1.32)	

Tabla R16. Resultados de los análisis de regresión para el test de asociación de pacientes con AA y pacientes sin AA. En el valor de la OR con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo. En aquellos casos en los que el menor valor BIC estimado es el mismo para diferentes modelos de herencia se incluyen los resultados correspondientes a dichos modelos.

(*) corresponde al valor P corregido por Bonferroni.

I.1.5.2 Análisis de haplotipos de variables clínicas con dos estados

Al comparar pacientes con AA frente a pacientes sin AA, se obtuvieron diferencias significativas para el test global correspondiente a la combinación de los polimorfismos rs10447760/rs923875/rs10500038/rs4730626 (P=0.0357).

Respecto a los test individuales, se obtuvieron diferencias significativas en algunos de ellos, mostrados en la tabla R17. Al corregir por Bonferroni dentro de los haplotipos obtenidos de la combinación de 4 SNPs correspondiente, la significación se mantuvo para dos de ellos: rs923875C/rs10500038G/rs4730626A/rs1668335G (P=0.037) y rs10255943A/rs10486026T/rs2396753A/rs17137124C (P=0.018), a pesar de que los test globales de estas combinaciones no resultaron significativos.

Al comparar pacientes con AA crónicos frente a pacientes con AA no crónicos, se obtuvieron diferencias significativas para el test global correspondiente a la combinación de los polimorfismos rs12670585/rs1456031/rs2396765/rs1456021 (P=0.004).

Respecto a los test individuales, se obtuvieron diferencias significativas en algunos de ellos. Sin embargo, la significación no se mantuvo al corregir por Bonferroni dentro de los haplotipos obtenidos de la combinación de 4 SNPs correspondiente.

Al comparar pacientes con inicio temprano de la enfermedad frente a pacientes con inicio tardío no se obtuvieron diferencias significativas para ninguno de los test globales aplicados a las combinaciones de 4 SNPs consecutivos.

Además, al igual que en los casos anteriores, respecto a los test individuales, aunque se obtuvieron diferencias significativas en algunos, la significación no se mantuvo al corregir por Bonferroni.

SNPs del haplotipo	Test Global			Test individual				
	df	χ^2	P	Haplotipo	Frec. Pacientes con AA	Frec. Pacientes sin AA	χ^2	P (*)
rs10447760 rs923875 rs10500038 rs4730626	5	11.94	0.0357	C-C-G-A	0.017	0.062	5.288	0.0215
rs923875 rs10500038 rs4730626 rs1668335	8	15.32	0.0531	C-G-A-G	0.009	0.032	8.012	0.0046 (0.037)
rs10500038 rs4730626 rs1668335 rs11771168	6	11.57	0.0722	A-G-G-C	0.1577	0.0436	5.475	0.0193
rs10255943 rs10486026 rs2396753 rs17137124	6	8.62	0.1961	A-T-A-C	0.0145	0.0727	8.797	0.003 (0.018)

Tabla R17. Análisis de haplotipos al comparar pacientes con AA y pacientes sin AA. Límite para la frecuencia de haplotipos raros = 0.03. Haplotipos obtenidos en ventanas de 4 SNPs.

(*) Valores de P que se mantienen significativos después de la corrección por Bonferroni para el número de haplotipos analizados.

Resultados

I.1.5.3 Variables clínicas con más de dos estados

Como se menciona en el apartado I.1.1.1.4 del Material y Métodos el análisis de asociación también se llevó a cabo con los ítems comprendidos en la escala PSYRATS, así como la puntuación total de esta escala, y dos ítems de la escala Manchester.

Los ítems a los que nos referimos son los siguientes:

Escala PSYRATS de AA	Escala Manchester
<ul style="list-style-type: none"> ▪ Frecuencia de aparición de las voces ▪ Duración ▪ Localización de las voces ▪ Intensidad (volumen) ▪ Grado de convicción del origen de las voces ▪ Cantidad de contenido negativo de las voces ▪ Frecuencia de contenido negativo ▪ Frecuencia con la que producen ansiedad ▪ Repercusión en la vida diaria de las voces ▪ Control sobre las voces 	<ul style="list-style-type: none"> ▪ Incoherencia del lenguaje ▪ Pobreza del lenguaje

Al tratarse de variables clínicas con múltiples estados (Valores de 0 a 4 para los ítems de las escalas, y suma de los correspondientes a la escala PSYRATS para la puntuación total de ésta), se utilizó la regresión lineal como método estadístico de análisis.

En las tablas R18, R19 y R20 se muestran únicamente los polimorfismos que resultaron significativos ($P < 0.05$) para cada una de las variables estudiadas.

Los resultados correspondientes a la **puntuación total de la escala PSYRATS** muestran que únicamente se encontró asociación de esta variable con 4 SNPs: rs6961558, rs10500038, rs2396765 y rs1456021, aunque la significación se pierde al corregir por Bonferroni (tabla R18). En el caso del polimorfismo rs6961558, la obtención de OR significativas probablemente sea debida a la baja frecuencia del alelo menor y ausencia de uno de los genotipos.

SNP	Modelo de herencia	Genotipos	Frecuencia absoluta	Respuesta Media (error estándar)	Diferencias (IC 95%)	P
rs6961558	nd	GG GA	207 8	16.76 (1.08) 0.75 (0.75)	0 -16.01 (-26.82- -5.21)	0.0041
rs10500038	aditivo	-	-	-	4.19 (0.50-7.88)	0.027
rs2396765	aditivo	-	-	-	3.44 (0.12-6.76)	0.043
rs1456021	aditivo	-	-	-	3.21 (0.01-6.40)	0.05

Tabla R18. Análisis de asociación entre genotipos y la puntuación total de la escala PSYRATS para las alucinaciones auditivas.

En los casos en los que el programa no determina el modelo de herencia debido a que sólo hay presentes dos de los tres posibles genotipos, esto aparece indicado como nd. En el valor de las diferencias con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo.

En el caso de las **variables de la escala PSYRATs**, los polimorfismos para los que se obtuvo una tendencia en la puntuación total vuelven a encontrarse asociados con algunas de las variables (tabla R20). En concreto, rs6961558 se encuentra asociado con todos los ítems, aunque nuevamente es probable que esto se deba a la baja frecuencia de uno de los dos alelos. rs10500038 muestra tendencia de diferencias significativas en los ítems de Frecuencia, Duración, Intensidad, Contenido Negativo, Frecuencia de Negativo, Repercusión y Control. Los SNPs rs1456021 y rs239665 muestran tendencia para los ítems de Frecuencia de Contenido Negativo, Frecuencia de Ansiedad e Intensidad de la Ansiedad. Además de estos SNPs, rs1668335 mostró algunos valores significativos para los ítems Frecuencia de Ansiedad y Repercusión y el SNP rs17137124 para el ítem Repercusión. Sin embargo, para todos ellos tras la corrección por Bonferroni no se obtiene ningún polimorfismo con diferencias significativas.

En el caso de las **variables de la escala Manchester** analizadas, los polimorfismos rs7803667, rs10447760, rs923875, rs4730626, rs2253478 y rs1456021 resultaron significativos para el ítem Incoherencia del Lenguaje, y rs2396722, rs2253478 y rs4730676 significativos para el ítem Pobreza del Lenguaje. Es de destacar que tras la corrección por Bonferroni se mantiene la significación para el polimorfismo rs2253478 con la variable Pobreza del lenguaje (P corregida =0.038).

SNP	Modelo de herencia	Genotipos	Frecuencia absoluta	Respuesta Media (error estándar)	Diferencias (IC 95%)	P (*)
<i>Incoherencia del lenguaje</i>						
rs7803667	Dominante	TT	113	0.45 (0.08)	0.00	0.0034
		AT-AA	150	0.83 (0.09)	0.38 (0.13-0.64)	
rs10447760	Aditivo	-	-	-	0.23 (0.03-0.44)	0.027
rs923875	Dominante	AA	84	0.44 (0.09)	0.00	0.016
		AC-CC	179	0.78 (0.08)	0.34 (0.07-0.61)	
rs4730626	Sobredominancia	GG-AA	171	0.82 (0.09)	0.00	0.022
		AG	72	0.47 (0.12)	-0.35 (-0.65- -0.05)	
rs2253478	Sobredominancia	GG-AA	129	0.86 (0.1)	0.00	0.044
		AG	115	0.57 (0.1)	-0.29 (-0.56- -0.01)	
rs1456021	Recesivo	AA-AG	227	0.68 (0.07)	0.00	0.047
		GG	18	1.22 (0.36)	0.54 (0.01-1.07)	
<i>Pobreza del lenguaje</i>						
rs2396722	Sobredominancia	TT-CC	143	0.93 (0.09)	0.00	0.024
		TC	120	0.64 (0.08)	-0.29 (-0.54- -0.04)	
rs2253478	Sobredominancia	GG-AA	129	1.08 (0.1)	0.00	0.0016 (0.038)
		GA	115	0.65 (0.08)	-0.43 (-0.69- -0.16)	
rs4730676	Sobredominancia	GG-AA	171	0.99 (0.08)	0.00	0.0068
		GA	72	0.6 (0.11)	-0.40 (-0.68- -0.11)	

Tabla R19. Resultados del análisis de asociación genotipos-variables de la escala Manchester en los casos en los que el valor de $P < 0.05$.

En los casos en los que el programa no determina el modelo de herencia debido a que sólo hay presentes dos de los tres posibles genotipos, esto aparece indicado como nd. En el valor de las diferencias con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo.

(*) Valores de P que se mantienen significativos después de la corrección por Bonferroni para el número de polimorfismos analizados.

Resultados

SNP	Modelo de herencia	Genotipos	Frecuencia absoluta	Respuesta Media (error estándar)	Diferencias (IC 95%)	P
<i>Frecuencia</i>						
rs6961558	nd	GG GA	207 8	1.69 (0.12) 0 (0)	0 -1.69 (-2.87- -0.50)	0.0057
rs10500038	Aditivo	-	-	-	0.45 (0.05-0.85)	0.03
<i>Duración</i>						
rs6961558	nd	GG GA	207 8	1.65 (0.12) 0 (0)	0 -1.65 (-2.83- -0.48)	0.0063
rs10500038	Aditivo	-	-	-	0.59 (0.19-0.98)	0.0041
<i>Localización</i>						
rs6961558	nd	GG GA	207 8	1.48 (0.11) 0 (0)	0 -1.48 (-2.60- -0.36)	0.01
<i>Intensidad</i>						
rs6961558	nd	GG GA	207 8	1.23 (0.09) 0.25 (0.25)	0 -0.98 (-1.87- -0.09)	0.032
rs10500038	Recesivo	GG-AG AA	190 10	1.15 (0.09) 2.1 (0.43)	0 0.95 (0.13-1.76)	0.023
<i>Convicción</i>						
rs6961558	nd	GG GA	207 8	1.72 (0.12) 0.12 (0.12)	0 -1.59 (-2.79- -0.40)	0.0097
<i>Negativo</i>						
rs6961558	nd	GG GA	207 8	1.53 (0.11) 0.12 (0.12)	0 -1.40 (-2.55- -0.25)	0.018
rs10500038	Aditivo	-	-	-	0.42 (0.03-0.80)	0.036
<i>Frecuencia negativo</i>						
rs6961558	nd	GG GA	207 8	1.56 (0.11) 0.12 (0.12)	0.00 -1.44 (-2.58- -0.29)	0.015
rs10500038	Aditivo	-	-	-	0.53 (0.15-0.91)	0.0071
rs2396765	Aditivo	-	-	-	0.42 (0.08-0.77)	0.018
rs1456021	Aditivo	-	-	-	0.38 (0.04-0.72)	0.029
<i>Frecuencia ansiedad</i>						
rs6961558	nd	GG GA	207 8	1.35 (0.11) 0 (0)	0.00 -1.35 (-2.45- -0.25)	0.017
rs1668335	Sobredominancia	GG-AA GA	117 83	1.54 (0.16) 1.05 (0.16)	0 -0.49 (-0.93- -0.05)	0.032
rs2396765	Aditivo	-	-	-	0.43 (0.09-0.76)	0.014
rs1456021	Aditivo	-	-	-	0.43 (0.11-0.75)	0.0094
<i>Intensidad ansiedad</i>						
rs6961558	nd	GG GA	207 8	1.33 (0.11) 0 (0)	0.00 -1.33 (-2.41- -0.26)	0.016
rs2396765	Aditivo	-	-	-	0.43 (0.10-0.76)	0.011
rs1456021	Aditivo	-	-	-	0.43 (0.12-0.75)	0.0077
<i>Repercusión</i>						
rs6961558	nd	GG GA	206 8	1.43 (0.1) 0 (0)	0.00 -1.43 (-2.45- -0.41)	0.0065
rs10500038	Aditivo	-	-	-	0.39 (0.05-0.73)	0.026
rs1668335	Sobredominancia	GG-AA GA	116 83	1.56 (0.14) 1.11 (0.15)	0 -0.45 (-0.86- -0.05)	0.03
rs17137124	Sobredominancia	TT-CC TC	112 102	1.57 (0.14) 1.16 (0.14)	0 -0.41 (-0.80- -0.03)	0.038
<i>Control</i>						
rs6961558	nd	GG GA	207 8	1.82 (0.13) 0.12 (0.12)	0.00 -1.70 (-2.96- -0.44)	0.0089
rs10500038	Recesivo	GG-AG AA	190 10	1.68 (0.13) 3.1 (0.53)	0.00 1.42 (0.27-2.56)	0.016

Tabla R20. Resultados del análisis de asociación genotipos-variables de la escala PSYRATS para las alucinaciones auditivas en los casos en los que los valores de $P < 0.05$.

En los casos en los que el programa no determina el modelo de herencia debido a que sólo hay presentes dos de los tres posibles genotipos, esto aparece indicado como nd. En el valor de las diferencias con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo.

Al comparar la distribución de las frecuencias alélicas y genotípicas entre los grupos definidos por variables de dos estados, únicamente se encontraron evidencias de diferencias significativas para la comparación entre pacientes crónicos y no crónicos de los polimorfismos rs1456031 y rs1456021. En estos SNPs se mantuvo la significación para el análisis mediante regresión logística tras llevar a cabo la corrección múltiple. No obstante, este resultado debe ser tomado con prudencia, ya que los resultados no son iguales respecto a los obtenidos por test de Fisher, en cuyo caso no se mantenía significación tras corrección múltiple, y además, los tamaños muestrales comparados son pequeños. Un dato a favor de la consistencia de estos datos es que en el análisis de haplotipos se obtuvo diferencias significativas para el haplotipo que contiene ambos SNPs: rs12670585/rs1456031/ rs2396765/ rs1456021 ($P=0.004$).

Para la presencia o ausencia de AA y el inicio temprano o tardío de la enfermedad no se obtuvieron diferencias significativas en la distribución de frecuencias alélicas, o genotípicas.

En el caso de los haplotipos, además del ya mencionado, se obtuvieron diferencias significativas después de la corrección de Bonferroni entre pacientes con AA y pacientes sin AA en los tests individuales correspondientes a rs923875C/rs10500038G/rs4730626A/rs1668335G y rs10255943A/ rs10486026T/rs2396753A/rs17137124C (P corregidas = 0.037 y 0.018, respectivamente), a pesar de que los test globales de estas combinaciones no resultaron significativos.

El análisis de la distribución de frecuencias genotípicas respecto a variables de las escalas PSYRATS y Manchester mostró diferencias significativas tras la corrección múltiple para el polimorfismo rs2253478 con la variable Pobreza del lenguaje (P corregida =0.0384).

I.2 Estudio de asociación caso-control con polimorfismos del gen *HAR1A*

El gen *HAR1A* se localiza en el extremo del brazo largo del cromosoma 20 (20q13.33) como se muestra en la figura R7. Se encuentra organizado en dos exones. El exón 1 incluye una región de 118 pb con 18 cambios específicos de la especie humana que constituye el elemento con tasa de sustitución más acelerada del linaje humano, denominado elemento HAR1, de *Human Accelerated Region 1* (Pollard *et al.*, 2006). Este exón solapa con parte del gen *HAR1R*.

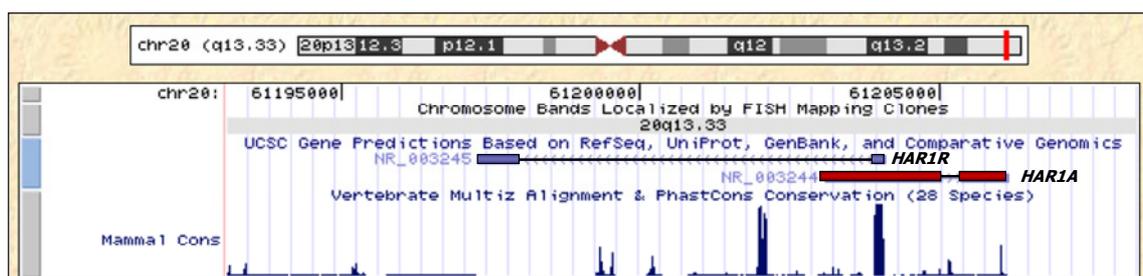


Figura R7. Contexto genómico en el que se localiza el gen *HAR1A*. En la parte inferior de la figura aparece reflejado el grado de conservación con otros mamíferos. Figura extraída del *UCSC Genome Browser*, con modificaciones.

Resultados

Los SNPs estudiados se distribuyen a lo largo de la región donde se encuentra el gen, cubriendo desde la posición -1017 hasta +4068, respecto del inicio del exón 1, como se puede ver en la figura R8. No se encontró ningún polimorfismo descrito en la región acelerada descrita como HAR1 (Pollard *et al.*, 2006).

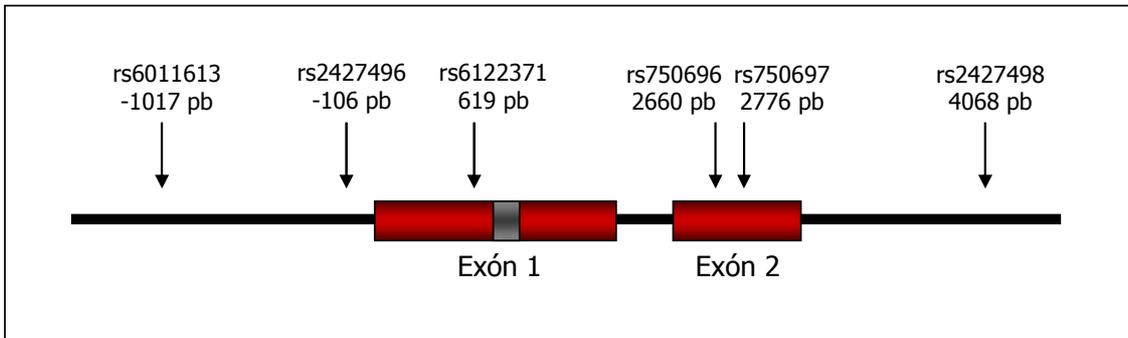


Figura R8. Distribución de los polimorfismos estudiados en la región del gen *HAR1A*. Las distancias reflejadas corresponden al número de pares de bases respecto a la primera posición en el exón 1. La región HAR1 aparece en color gris sobre el exón 1.

Después de eliminar las muestras que fallaron en la genotipación, se analizaron un total de 333 controles y 276 pacientes.

Tanto en el grupo de controles como en el de pacientes, las frecuencias genotípicas se encuentran en equilibrio de Hardy-Weinberg, cuyos valores aparecen reflejados en la tabla R21.

SNPs	Valores P del test de equilibrio Hardy-Weinberg	
	Controles	Pacientes
rs6011613	0.72	0.19
rs2427496	0.55	0.8
rs6122371	0.44	0.4
rs750696	0.65	1
rs750697	0.14	0.22
rs2427498	0.27	0.18

Tabla R21. Resultados del test de equilibrio de Hardy-Weinberg para los SNPs estudiados en controles y pacientes.

1.2.1 Análisis del desequilibrio de ligamiento

A pesar de estar distribuidos a lo largo de poco más de 5 kilobases, no se encontró un elevado desequilibrio de ligamiento entre los SNPs (rango r^2 desde 0.018 a 0.538). Esto se observa en la figura R9, donde se muestran diagramas con los valores r^2 para los principales grupos analizados. De los SNPs analizados, únicamente rs750696 constituye un tagSNP. Los tagSNPs son SNPs que se caracterizan por ser capaces de capturar la mayor parte de la variabilidad de una región.

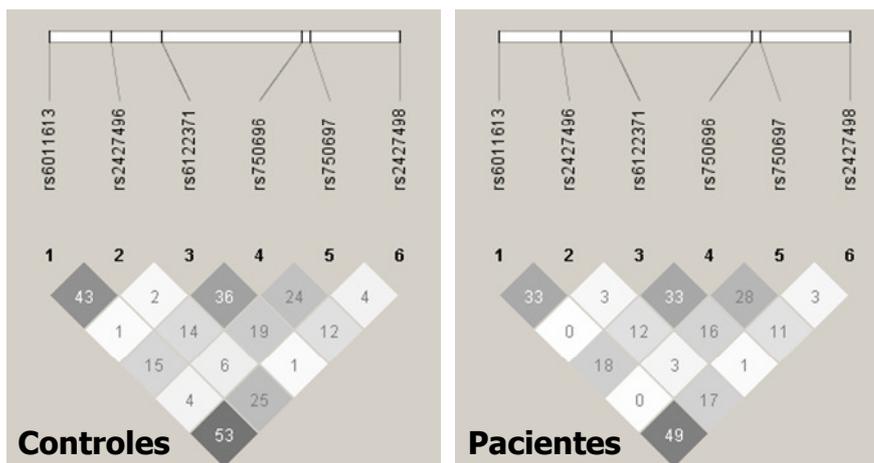


Figura R9. Diagramas de desequilibrio de ligamiento para los grupos de controles y pacientes. Los valores de r^2 , aparecen reflejados sobre 100.

En la figura R10 se muestra un gráfico de la misma región extraído del *HapMap*, correspondiente a la población CEU. En la región, únicamente hay disponible información de 4 SNPs, 3 de los cuales (rs6011613, rs750696 y rs2427498) se analizaron en este estudio. Para estos, los valores de LD de la base de datos son similares a los obtenidos en las poblaciones de este trabajo.

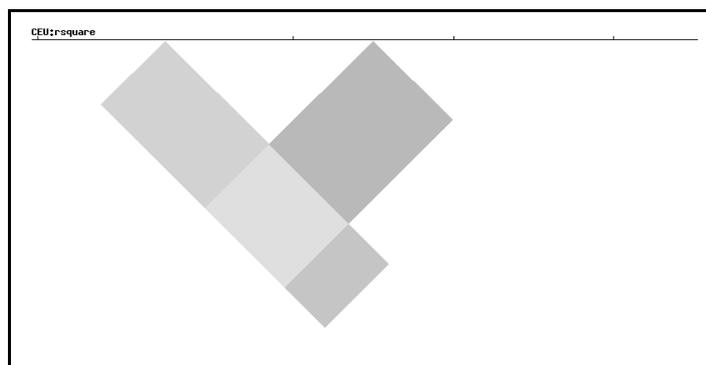


Figura R10. Esquema de los valores r^2 de LD disponibles para la región analizada obtenidos a partir de la base de datos del *HapMap* (*HapMap Data Rel 23a/ phaseII Mar 2008*).

1.2.2 Análisis de frecuencias genotípicas y alélicas de polimorfismos del gen *HAR1A* en pacientes y controles

A continuación, se presentan los resultados obtenidos para el análisis de frecuencias genotípicas y alélicas en los grupos estudiados. Al tratarse del primer estudio de asociación caso-control realizado con polimorfismos de este gen, no hay datos previos de otros estudios con los que comparar, a excepción del publicado por nuestro grupo de investigación (Tolosa *et al*, 2008).

Al igual que en el caso del gen *FOXP2*, las comparaciones realizadas en este estudio y el número de individuos en cada uno de los grupos son las siguientes:

Resultados

Grupos comparados en el test de asociación	Número de individuos
Control/Paciente	333
	276
Control/Paciente con AA	333
	212
Control/Paciente con esquizofrenia	333
	209

Tabla R22. Grupos muestrales comparados en el test de asociación y número de individuos en cada comparación.

Por claridad sólo se presentarán en este apartado los resultados correspondientes a las comparaciones controles-pacientes y controles-pacientes con AA.

En las tablas R23 y R24, se muestran las frecuencias genotípicas y alélicas de cada grupo, así como los valores de χ^2 y P asociada a éste mediante el test exacto de Fisher. Los SNPs estudiados mostraron valores de MAF >10%. Para aquellos polimorfismos de los que había información en las bases de datos de las frecuencias genotípicas o alélicas (rs6011613, rs2427498, rs750696, rs750697), éstas no mostraron diferencias significativas respecto a las obtenidas en el grupo de controles de nuestro estudio.

SNP	Frecuencias genotípicas			χ^2	P	Frecuencias alélicas		χ^2	P
	CC	CT	TT			C	T		
rs6011613									
Controles	220 (0.66)	100 (0.30)	13 (0.04)	0.808	0.673	0.81	0.19	0.361	0.541
Pacientes	179 (0.65)	82 (0.30)	15 (0.05)			0.8	0.2		
rs2427496									
Controles	235 (0.71)	88 (0.26)	10 (0.03)	2.307	0.314	0.84	0.16	1.876	0.171
Pacientes	206 (0.75)	66 (0.24)	4 (0.01)			0.87	0.13		
rs6122371									
Controles	93 (0.28)	173 (0.52)	67 (0.20)	0.021	0.994	0.54	0.46	0.001	0.972
Pacientes	76 (0.28)	145 (0.52%)	55 (0.20)			0.54	0.46		
rs750696									
Controles	110 (0.33)	167 (0.50)	56 (0.17)	0.610	0.734	0.58	0.42	0.468	0.494
Pacientes	87 (31.5%)	136 (49.3%)	53 (19.2%)			0.56	0.44		
rs750697									
Controles	192 (0.58)	115 (0.34)	26 (0.08)	0.447	0.808	0.75	0.25	0.475	0.491
Pacientes	152 (0.55)	100 (0.36)	24 (0.09)			0.73	0.27		
rs2427498									
Controles	241 (0.72)	88 (0.26)	4 (0.01)	4.002	0.127	0.86	0.14	0.615	0.433
Pacientes	212 (0.77)	57 (0.21)	7 (0.02)			0.87	0.13		

Tabla R23. Frecuencias genotípicas y alélicas en controles y pacientes con los valores de χ^2 y P asociados.

No se encontraron diferencias significativas en las frecuencias genotípicas y alélicas de ninguno de los SNPs estudiados al realizar las tres comparaciones entre grupos, controles frente a pacientes, controles frente a pacientes con alucinaciones y controles frente a pacientes con esquizofrenia.

También se analizó la existencia de diferencias en las frecuencias genotípicas de los diferentes polimorfismos entre mujeres y varones, tanto en el grupo control, como en la muestra de pacientes con AA y la muestra total de pacientes. En algunos casos se han encontrado diferencias significativas (valores de $P \leq 0.05$), pero estas diferencias desaparecen al aplicar la corrección de Bonferroni.

SNP	Frecuencias genotípicas			χ^2	P	Frecuencias alélicas		χ^2	P
	CC	CT	TT			C	T		
rs6011613									
Controles	220 (0.66)	100 (0.30)	13 (0.04)	0.575	0.7510	0.81	0.19	0.422	0.516
Pacientes con AA	136 (0.64)	65 (0.31)	11 (0.05)			0.79	0.21		
rs2427496									
Controles	235 (0.71)	88 (0.26)	10 (0.03)	5.285^a	0.073	0.84	0.16	3.647	0.056
Pacientes con AA	162 (0.76)	49 (0.23)	1 (0.01)			0.88	0.12		
rs6122371									
Controles	93 (0.28)	173 (0.52)	67 (0.20)	0.348	0.848	0.54	0.46	0.330	0.566
Pacientes con AA	55 (0.26)	111 (0.52)	46 (0.22)			0.52	0.48		
rs750696									
Controles	110 (0.33)	167 (0.50)	56 (0.17)	0.539	0.776	0.58	0.42	0.518	0.472
Pacientes con AA	65 (0.31)	107 (0.50)	40 (0.19)			0.56	0.44		
rs750697									
Controles	192 (0.58)	115 (0.34)	26 (0.08)	0.580	0.756	0.75	0.25	0.003	0.950
Pacientes con AA	119 (0.56)	79 (0.37)	14 (0.07)			0.75	0.25		
rs2427498									
Controles	241 (0.72)	88 (0.26)	4 (0.01)	2.772^a	0.281	0.86	0.14	0.453	0.502
Pacientes con AA	162 (0.76)	45 (0.21)	5 (0.02)			0.87	0.13		

Tabla R24. Frecuencias genotípicas y alélicas en controles y pacientes con AA con los valores de χ^2 y P asociados.

^a corresponden a tests en los que en una de las clases los valores esperados son menores de 5.

El programa *SNPStats* además, permitió, mediante regresión logística testar qué modelo de herencia se ajustaba más a las proporciones observadas en los grupos. En las siguientes tablas se muestra para cada polimorfismo qué modelo se ajustó más en las comparaciones realizadas, esto es, qué modelo obtiene un menor valor para BIC (Criterio de información Bayesiana), con las OR y valores de P asociados.

Los resultados obtenidos mediante regresión logística coinciden con los observados a partir de los test exactos de Fisher, no obteniéndose diferencias significativas para las frecuencias de ninguno de los polimorfismos analizados tras corregir por Bonferroni.

Resultados

SNPs	Modelo de herencia	Genotipos	Controles	Pacientes	OR	P
rs6011613	Recesivo	CC-CT	320 (0.96)	261 (0.95)	1	0.37
		TT	13 (0.04)	15 (0.05)	1.41 (0.66-3.03)	
rs2427496	Recesivo	CC-CT	323 (0.97)	272 (0.99)	1	0.19
		TT	10 (0.03)	4 (0.01)	0.48 (0.15-1.53)	
rs6122371	Dominante	AA	93 (0.28)	76 (0.275)	1	0.91
		AC-CC	240 (0.72)	200 (0.725)	1.02 (0.71-1.46)	
	Recesivo	AA-AC	266 (0.80)	221 (0.80)	1	0.95
		CC	67 (0.20)	55 (0.20)	0.99 (0.66-1.47)	
Sobredominancia	AA-CC	160 (0.48)	131 (0.475)	1	0.89	
	AC	173 (0.52)	145 (0.525)	1.02 (0.74-1.41)		
	Aditivo				1 (0.80-1.27)	0.97
rs750696	Recesivo	CC-CT	277 (0.83)	223 (0.81)	1	0.45
		TT	56 (0.17)	53 (0.19)	1.18 (0.78-1.78)	
rs750697	Dominante	GG	192 (0.58)	152 (0.55)	1	0.52
		AG-AA	141 (0.42)	124 (0.45)	1.11 (0.81-1.53)	
	Aditivo				1.09 (0.85-1.39)	0.51
rs2427498	Sobredominancia	CC-TT	245 (0.74)	219 (0.79)	1	0.09
		CT	92 (0.26)	57 (0.21)	0.72 (0.50-1.06)	

Tabla R25. Resultados de los análisis de regresión para el test de asociación de controles y pacientes. En aquellos casos en los que el menor valor BIC estimado es el mismo para diferentes modelos de herencia se incluyen los resultados correspondientes a dichos modelos.

SNPs	Modelo de herencia	Genotipos	Controles	Pacientes con AA	OR	P (*)
rs6011613	Recesivo	CC-CT TT	320 (0.96) 13 (0.04)	201 (0.95) 11 (0.05)	1 1.35 (0.59-3.07)	0.48
rs2427496	Recesivo	CC-CT TT	323 (0.97) 10 (0.03)	211 (0.995) 1 (0.005)	1 0.15 (0.02-1.20)	0.02 (0.12)
rs6122371	Aditivo				1.08 (0.84-1.38)	0.56
rs750696	Aditivo				1.10 (0.86-1.41)	0.47
rs750697	Sobredominancia	GG-AA GA	218 (0.655) 115 (0.345)	133 (0.67) 79 (0.37)	1 1.13 (0.79-1.61)	0.52
		CC-TT CT	245 (0.74) 88 (0.26)	167 (0.79) 45 (0.21)	1 0.75 (0.50-1.13)	0.17

Tabla R26. Resultados de los análisis de regresión para el test de asociación de controles y pacientes con AA. Los SNPs cuyas frecuencias genotípicas difieren significativamente entre grupos se encuentran sombreados. (*) Valor de P corregido por Bonferroni.

I.2.3 Análisis de haplotipos

A continuación se muestran los resultados obtenidos del análisis de haplotipos. En este caso, a diferencia del gen *FOXP2*, al estudiar 6 SNPs, sí se pudo obtener el haplotipo completo. Al igual que en el caso de frecuencias genotípicas y alélicas, se presentan en este apartado únicamente las comparaciones controles-pacientes y controles-pacientes con AA. En todos los casos, la frecuencia mínima de un haplotipo necesaria para considerarlo en el test fue de 0.03 para alguno de los dos grupos comparados.

Al comparar pacientes con controles (tablas R27 Y R28) no se obtuvieron diferencias significativas ni para el test global ni para ninguno de los haplotipos individuales.

Cuando se compararon controles frente a pacientes con AA, no se obtuvieron diferencias globales, aunque sí para uno de los haplotipos individuales. Sin embargo, esta diferencia (P=0.022) no se mantuvo al llevar a cabo la corrección por el número de haplotipos (P=0.198).

rs6011613	rs2427496	rs6122371	rs750696	rs750697	rs2427498	Controles	Pacientes	Odds-Ratio (IC 95%)	χ^2	P
C	C	A	T	G	C	0.371	0.377	0.936 (0.488-1.796)	0.184	0.668
C	C	C	C	A	C	0.185	0.184	0.915 (0.459-1.824)	0.0001	0.989
C	C	C	C	G	C	0.119	0.104	0.8047 (0.396-1.634)	0.767	0.381
T	T	A	C	G	T	0.079	0.067	0.776 (0.364-1.654)	1.434	0.231
C	C	C	T	G	C	0.047	0.059	1.157 (0.519-2.582)	0.350	0.554
C	C	A	C	A	C	0.036	0.039	1	0.041	0.840
C	T	C	C	G	C	0.034	0.027	0.720 (0.298-1.795)	0.707	0.400
T	C	C	C	G	T	0.033	0.042	1.159 (0.494-2.721)	0.773	0.379
Otros haplotipos						0.096	0.101			

Tabla R27. Análisis de haplotipos al comparar controles (N=333) y pacientes (N=276).
Test de asociación global: $\chi^2=3.157$, df=7, P=0.8701

rs6011613	rs2427496	rs6122371	rs750696	rs750697	rs2427498	Controles	Pacientes con AA	Odds-Ratio (IC 95%)	χ^2	P (*)
C	C	A	T	G	C	0.371	0.371	1.091 (0.522-2.278)	0.075	0.784
C	C	C	C	A	C	0.185	0.172	1.017 (0.467-2.215)	0.219	0.639
C	C	C	C	G	C	0.119	0.121	1.109 (0.505-2.438)	0.004	0.948
T	T	A	C	G	T	0.079	0.063	0.877 (0.375-2.051)	1.876	0.181
C	C	C	T	G	C	0.047	0.067	1.569 (0.651-3.78)	0.948	0.330
C	C	A	C	A	C	0.036	0.033	1	0.133	0.716
C	T	C	C	G	C	0.034	0.026	0.819 (0.295-2.276)	0.944	0.331
T	C	C	C	G	T	0.033	0.048	1.551 (0.614-3.92)	1.672	0.196
T	C	C	C	A	C	0.015	0.037	2.681 (0.894-8.038)	5.257	0.022 (0.198)
Otros haplotipos						0.096	0.062			

Tabla R28. Análisis de haplotipos al comparar controles (N=333) y pacientes con AA (N=212).

Test de asociación global: $\chi^2=9.303$, df=8, P=0.317

Los haplotipos cuyas frecuencias difieren significativamente entre grupos se encuentran sombreados.

(*) Valor de P corregido por el número de haplotipos.

Resultados

No se encontraron diferencias significativas al comparar controles con pacientes con esquizofrenia.

En resumen, los resultados de la comparación de la distribución de frecuencias alélicas, genotípicas y haplotípicas en los grupos muestrales analizados (controles, pacientes, pacientes con AA y pacientes con esquizofrenia) no muestran evidencias de diferencias significativas entre ambos grupos cuando se aplican las correcciones múltiples más conservativas.

1.2.4 Medida del poder estadístico

Para el estudio de asociación del gen *HAR1A* también se utilizó el programa *Quanto*, para calcular el poder estadístico de los datos obtenidos. En las tablas R29 y R30 se observa para cada polimorfismo el poder estadístico asociado al test, teniendo en cuenta las frecuencias alélicas y el modelo de herencia seleccionado con el programa *SNPStats*.

SNPs	Modelo de herencia	Poder estadístico en el test controles (N=333)/pacientes (N=276)
rs6011613	Recesivo	0.1363
rs2427496	Recesivo	0.2219
rs6122371	Dominante	0.0514
	Recesivo	0.0503
	Sobredominancia	0.05
	Aditivo	0.05
rs750696	Recesivo	0.1252
rs750697	Dominante	0.0978
	Aditivo	0.1005
rs2427498	Sobredominancia	0.4605

Tabla R29. Estimaciones del poder estadístico de la comparación de controles frente a pacientes, para OR genéticas del intervalo de confianza 95%.

SNPs	Modelo de herencia	Poder estadístico en el test controles (N=333)/pacientes con AA (N=212)
rs6011613	Recesivo	0.1052
rs2427496	Recesivo	0.5599
rs6122371	Aditivo	0.0948
rs750696	Aditivo	0.1183
rs750697	Sobredominancia	0.1391
rs2427498	Sobredominancia	0.3293

Tabla R30. Estimaciones del poder estadístico de la comparación de controles frente a pacientes con AA, para OR genéticas del intervalo de confianza 95%.

Al igual que en el estudio con el gen *FOXP2*, el poder estadístico varía considerablemente en el rango de ORs estimadas para los SNPs. La mayoría de los valores obtenidos son bajos, por lo que la capacidad de detectar asociación en nuestra muestra es baja. Así, no es de extrañar que no se hayan obtenido resultados positivos.

1.2.5 Análisis de variables clínicas

1.2.5.1 Variables clínicas de dos estados

Al igual que en el caso del gen *FOXP2*, las variables clínicas con dos estados para las cuales se analizaron las frecuencias genotípicas y alélicas de los polimorfismos estudiados son las que se muestran en la tabla R31.

El análisis de estas variables se llevó a cabo del mismo modo que para las comparaciones entre controles-pacientes, controles-esquizofrénicos y controles-pacientes con AA. Con el fin de proporcionar mayor claridad a la exposición de resultados, únicamente se presentan en formato de tabla los resultados correspondientes a la presencia o ausencia de alucinaciones auditivas (Tabla R32).

Variable	Estado de la variable	Número de individuos
Presencia de alucinaciones auditivas	Con historia de alucinaciones auditivas	212
	Sin historia de alucinaciones auditivas	63
Cronicidad de las alucinaciones auditivas	Presencia de cronicidad	79
	Ausencia de cronicidad	133
Inicio de la enfermedad	Inicio temprano (hasta los 20 años)	104
	Inicio tardío (a partir de los 28 años)	85

Tabla R31. Variables clínicas dicotómicas analizadas.

Al comparar pacientes con AA y pacientes sin AA, no se encontraron diferencias significativas para las frecuencias genotípicas y alélicas.

Respecto a la cronicidad de las alucinaciones, no se encontraron diferencias significativas en las frecuencias genotípicas. Para las frecuencias alélicas, únicamente se encontraron diferencias significativas en el polimorfismo rs6011613 ($P=0.036$). Sin embargo, el valor de P no se mantiene significativo tras la corrección por Bonferroni ($P=0.216$).

Por último, al comparar pacientes con inicio temprano de la enfermedad con pacientes con inicio tardío, no se encontraron diferencias significativas para las frecuencias genotípicas ni para las alélicas.

Los modelos de herencia que más se ajustan a las proporciones observadas se calcularon mediante el programa *SNPStats*, y los correspondientes a la comparación entre pacientes con AA y pacientes sin AA se muestran en la tabla R33. Para estos, no se detectaron diferencias significativas en las frecuencias de ninguno de los polimorfismos mediante el análisis por regresión logística, resultado coincidente con el obtenido mediante los test exactos de Fisher.

Los mismos resultados se obtuvieron al tener en cuenta la cronicidad de las alucinaciones auditivas y el inicio temprano o tardío de la enfermedad.

Resultados

SNP	Frecuencias genotípicas			χ^2	P	Frecuencias alélicas		χ^2	P
	CC	CT	TT			C	T		
rs6011613	CC	CT	TT			C	T		
Pacientes sin AA	42 (0.67)	17 (0.27)	4 (0.06)	0.388^a	0.830	0.8	0.2	0.023	0.868
Pacientes con AA	136 (0.64)	65 (0.31)	11 (0.05)			0.79	0.21		
rs2427496	CC	CT	TT			C	T		
Pacientes sin AA	43 (0.68)	17 (0.27)	3 (0.05)	6.883^b	0.037	0.82	0.18	3.233	0.072
Pacientes con AA	162 (0.76)	49 (0.23)	1 (0.005)			0.88	0.12		
rs6122371	AA	AC	CC			A	C		
Pacientes sin AA	21 (0.33)	33 (0.52)	9 (0.14)	2.294	0.317	0.6	0.4	2.141	0.143
Pacientes con AA	55 (0.26)	111 (0.52)	46 (0.22)			0.52	0.48		
rs750696	CC	CT	TT			C	T		
Pacientes sin AA	21 (0.33)	29 (0.46)	13 (0.21)	0.383	0.828	0.56	0.44	0.008	0.928
Pacientes con AA	65 (0.31)	107 (0.50)	40 (0.19)			0.56	0.44		
rs750697	GG	GA	AA			G	A		
Pacientes sin AA	33 (0.52)	21 (0.33)	9 (0.14)	3.757	0.159	0.69	0.31	1.628	0.202
Pacientes con AA	119 (0.56)	79 (0.37)	14 (0.07)			0.75	0.25		
rs2427498	CC	CT	TT			C	T		
Pacientes sin AA	49 (0.78)	12 (0.19)	2 (0.03)	0.250^a	0.957	0.87	0.13	0.006	0.936
Pacientes con AA	162 (0.76)	45 (0.21)	5 (0.02)			0.87	0.13		

Tabla R32. Frecuencias genotípicas y alélicas pacientes con AA y pacientes sin AA con los valores de χ^2 y P asociados.

^a corresponden a tests en los que en una de las clases los valores esperados son menores a 5.

^b corresponden a tests en los que en una de las clases los valores esperados son menores a 2.

SNPs	Modelo de herencia	Genotipos	Pacientes con AA	Pacientes sin AA	OR	P (*)
rs6011613	Sobredominancia	CC-TT CT	147 (0.69) 65 (0.31)	46 (0.73) 17 (0.27)	1 0.84 (0.45-1.57)	0.57
rs2427496	Recesivo	CC-CT TT	211 (0.995) 1 (0.005)	60 (0.95) 3 (0.05)	1 10.55 (1.08-103.28)	0.03 (0.18)
rs6122371	Aditivo				0.73 (0.48-1.10)	0.13
rs750696	Sobredominancia	CC-TT CT	105 (0.495) 107 (0.505)	34 (0.54) 29 (0.46)	1 0.84 (0.48-1.47)	0.54
rs750697	Recesivo	GG-AG AA	198 (0.93) 14 (0.07)	54 (0.86) 9 (0.14)	1 0.84 (0.47-1.52)	0.57
rs2427498	Recesivo	CC-CT TT	207 (0.98) 5 (0.02)	61 (0.97) 2 (0.03)	1 1.36 (0.26-7.17)	0.72
	Sobredominancia	CC-TT CT	167 (0.79) 45 (0.21)	51 (0.81) 12 (0.19)	1 0.87 (0.43-1.78)	0.71

Tabla R33. Resultados de los análisis de regresión para el test de asociación de pacientes con AA y pacientes sin AA.

Los SNPs cuyas frecuencias genotípicas difieren significativamente entre grupos se encuentran sombreados.

(*) Valor de P corregido por Bonferroni.

1.2.5.2 Análisis de haplotipos de variables clínicas con dos estados

Al comparar pacientes con AA frente a pacientes sin AA, se obtuvo diferencias globales con un valor $P=0.0104$. En esta comparación surgen dos haplotipos con frecuencias significativamente diferentes en ambos grupos: rs6011613-C/rs2427496-C/rs6122371-C/rs750696-C/rs750697-G/rs2427498-C con una P asociada de 0.016 y rs6011613-T/rs2427496-T/rs6122371-C/rs750696-C/rs750697-G/rs2427498-C con una P asociada de 0.003. El primero de ellos podría constituir un haplotipo de riesgo para la presencia de AA en pacientes psicóticos, sin embargo, el hecho de que no existan diferencias significativas para este haplotipo entre controles y pacientes con AA sugiere que la vulnerabilidad se daría dentro del contexto psicótico.

Este haplotipo en concreto fue el que dio lugar a diferencias significativas ($P=0.002$) en un estudio previo realizado en nuestro grupo, con prácticamente la misma muestra (Tolosa *et al.*, 2008). Las diferencias entre los valores de P de ambos estudios deben ser debidas a la pequeña variación en el tamaño muestral, porque las frecuencias obtenidas para el haplotipo en los dos grupos son las mismas o a la modificación del valor umbral para haplotipos raros (0.05 en el artículo y 0.03 en la presente tesis). En cualquier caso, en este estudio no resiste la corrección por Bonferroni. Respecto al segundo haplotipo, las frecuencias en ambos grupos podrían indicar que se trata de un haplotipo protector. No obstante, debido al pequeño tamaño de la muestra de pacientes sin AA, y a la baja frecuencia del haplotipo en cuestión, es más probable que se trate de un error de tipo I.

Respecto a la cronicidad de las AA y el inicio temprano o tardío de la enfermedad, no se obtuvieron diferencias significativas al realizar las comparaciones correspondientes.

rs6011613	rs2427496	rs6122371	rs750696	rs750697	rs2427498	Pacientes con AA	Pacientes sin AA	Odds-Ratio (IC 95%)	χ^2	P (*)
C	C	A	T	G	C	0.367	0.401	1.419 (0.508-3.964)	0.338	0.561
C	C	C	C	A	C	0.173	0.217	1.24 (0.419-3.671)	1.216	0.270
C	C	C	C	G	C	0.123	0.048	3.975 (1.095-14.43)	5.765	0.016 (0.16)
T	T	A	C	G	T	0.063	0.079	1.245 (0.369-4.204)	0.425	0.515
C	C	C	T	G	C	0.068	0.035	3.019 (0.739-12.33)	1.466	0.226
T	C	C	C	G	T	0.048	0.024	3.124 (0.661-14.76)	1.428	0.232
T	C	C	C	A	C	0.036	0.007	7.653 (0.663-88.32)	2.47	0.116
C	C	A	C	A	C	0.033	0.052	1	0.972	0.324
C	T	C	C	G	C	0.026	0.032	1.241 (0.274-5.63)	0.357	0.550
T	T	C	C	G	C	0.003	0.042	0.099 (0.002-4.372)	9.036	0.003 (0.03)
Otros haplotipos						0.06	0.063			

Tabla R34. Análisis de haplotipos al comparar pacientes con AA (N=212) frente a pacientes sin AA (N=63). Test de asociación global: $\chi^2=21.55$, $df=9$, $P=0.0104$. (*) Valor de P corregido por el número de haplotipos. Los haplotipos cuyas frecuencias difieren significativamente entre grupos se encuentran sombreados.

Resultados

I.2.5.3 Variables clínicas con más de dos estados

Al igual que para el gen *FOXP2* el análisis de las variables de las escalas PSYRATS y Manchester, previamente indicadas en la tabla I.1.1.1.4 de la sección de Material y Métodos se llevó a cabo mediante el programa *SNPStats*, por medio de regresión lineal. Ninguno de los polimorfismos estudiados del gen *HAR1A* resultó significativo en los análisis de los diferentes ítems de la escala PSYRATS, incluido el de puntuación total de la escala. En cuanto a los dos ítems estudiados de la escala Manchester, se encontraron diferencias significativas en la distribución de genotipos del SNP rs750697 con el ítem de Incoherencia del Lenguaje.

SNP	Modelo de herencia	Genotipos	Frecuencia absoluta	Respuesta Media (error estándar)	Diferencias (IC 95%)	P (*)
rs750697	Sobredominancia	GG-AA	155	0.84 (0.09)	0.00	0.0059 (0.035)
		GA	91	0.45 (0.1)	-0.39 (-0.66 - -0.11)	

Tabla R35. Resultados del análisis de asociación genotipos-variables de la escala Manchester en los casos en los que el valor de $P < 0.05$. Resultados correspondientes al ítem Incoherencia del Lenguaje. En el valor de las diferencias con intervalo de confianza del 95%, el color verde indica que el genotipo es protector, y el color rojo indica genotipos de riesgo. (*) Valores de P que se mantienen significativos después de la corrección por Bonferroni para el número de polimorfismos analizados.

En resumen, para el gen *HAR1A*, el análisis de las frecuencias alélicas y genotípicas frente a variables clínicas con dos estados no reveló diferencias significativas para la presencia o ausencia de alucinaciones, cronicidad de las AA o inicio temprano o tardío de la enfermedad. En cambio, el análisis de haplotipos indicó diferencias globales entre las frecuencias haplotípicas de pacientes con AA y pacientes sin AA, aunque no se pudo replicar la significación obtenida en un estudio previo para la presencia de un posible haplotipo de riesgo dentro del fenotipo psicótico (rs6011613C/rs2427496C/rs6122371C/rs750696C/rs750697G/rs2427498C).

El análisis de frecuencias genotípicas frente a variables de las escalas PSYRATS y Manchester, únicamente mostró diferencias significativas para el SNP rs750697 con el ítem de Incoherencia en el Lenguaje (P corregida = 0.035).

I.3 Cambio en el número de repeticiones de trinucleótidos en el gen *FOXP2*

Debido a la ausencia de resultados positivos que tuvieran un fuerte impacto en la vulnerabilidad a psicosis, y especialmente a AA, dentro del estudio de asociación con polimorfismos del gen *FOXP2*, se decidió analizar otro tipo de variante estructural, como son las repeticiones de trinucleótidos. Como se ha mencionado en la introducción, uno de los dominios característicos del gen *FOXP2* es la presencia de dos tramos de poliglutaminas de 40 y 10 residuos, localizados en los exones 5 y 6, respectivamente (Figura R11). Expansiones en este tipo de repeticiones están asociadas a diferentes enfermedades, fundamentalmente de tipo neurológico por lo que el análisis de la variación de las repeticiones de poliglutaminas en el gen *FOXP2* en controles frente a pacientes psicóticos constituía un complemento al estudio de asociación.

Durante la genotipación del polimorfismo rs6961558 encontramos una deleción de nueve nucleótidos en un tramo con repeticiones CGG localizado en el intrón s1, y cercano al inicio de transcripción del exón s1 (Figura R11). Se decidió analizar esta región también, debido a la posible variabilidad en repeticiones entre pacientes y controles.

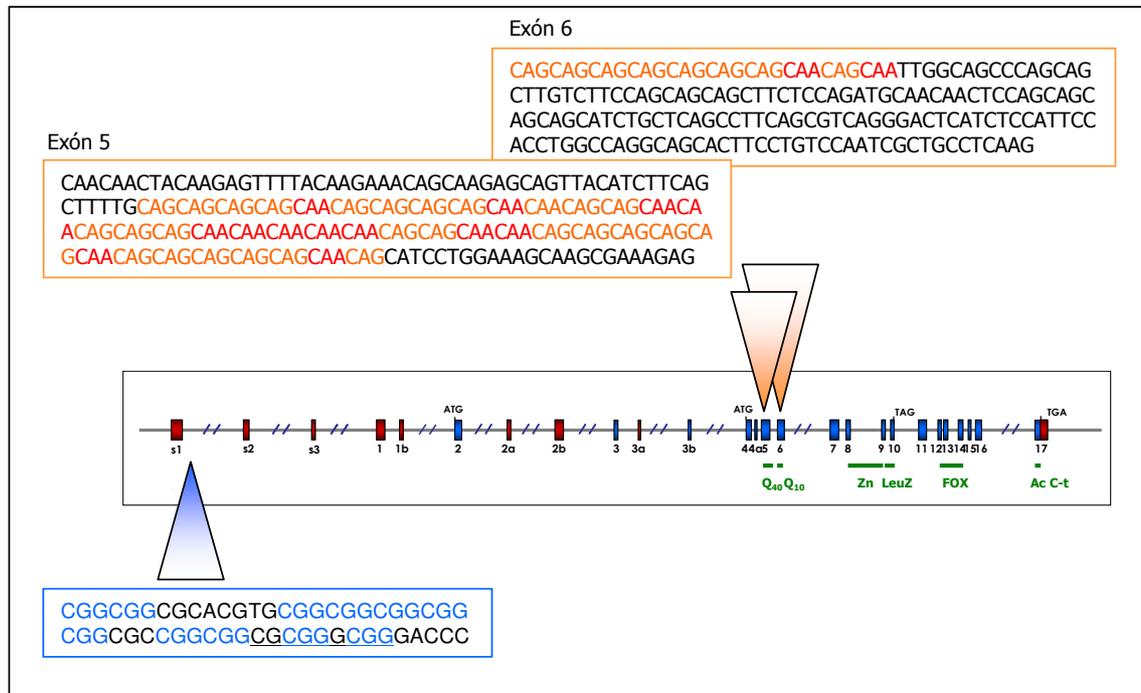


Figura R11. Localización de los tramos de repeticiones de trinucleótidos analizados en el gen *FOXP2*. En color naranja o rojo se muestran los tramos de trinucleótidos de las glutaminas y en color azul las repeticiones analizadas en el intrón s1. El fragmento deleciónado en uno de los pacientes se muestra subrayado.

La búsqueda de posibles expansiones o deleciones en los dos tramos de poliglutaminas se llevó a cabo mediante amplificación por PCR con un cebador marcado con un fluorocromo y posterior análisis de los fragmentos de DNA amplificados por electroforesis capilar (ver apartado II.4 de la sección de Material y Métodos). En las figuras R12 y R13 se muestran ejemplos de los resultados obtenidos para las regiones analizadas.

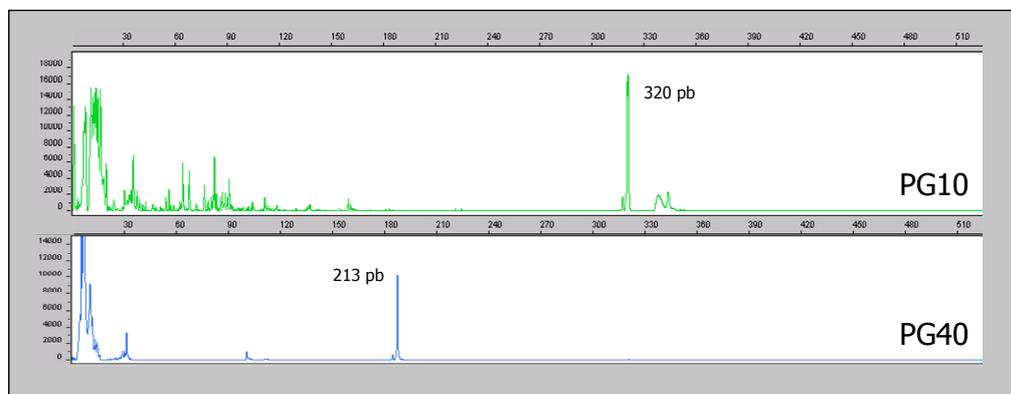


Figura R12. Análisis de fragmentos de los tramos de poliglutaminas. El tamaño de los fragmentos obtenidos por PCR se muestra en el pico correspondiente.

Resultados

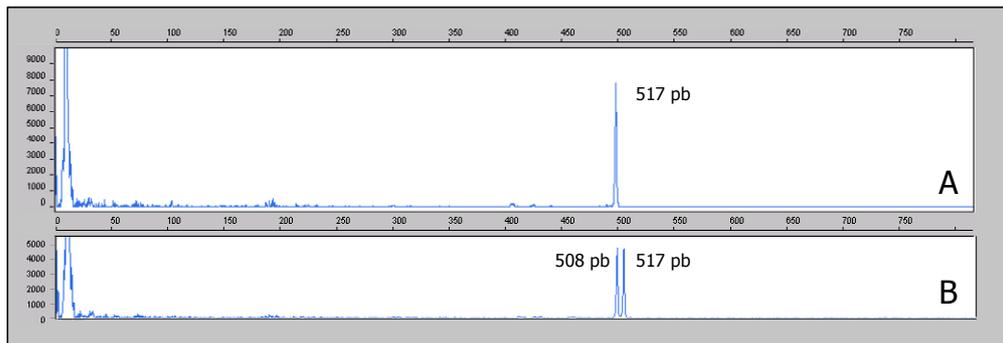


Figura R13. Análisis de fragmentos del tramo de repeticiones localizado en el intrón s1. La muestra A corresponde a un individuo sin variación en el tamaño del fragmento y la muestra B corresponde al individuo en el que se detectó una delección de 3 trinucleótidos CGG en heterocigosis.

Las muestras de pacientes psicóticos y de controles utilizadas fueron las siguientes:

Fragmento analizado	Número de muestras	
	Controles	Pacientes
Tramo de 40 repeticiones de poliglutaminas	151	228
Tramo de 10 repeticiones de poliglutaminas	147	122
Tramo de repeticiones CGG en intrón s1	112	172

Tabla R36. Muestras analizadas en el análisis de fragmentos de DNA por electroforesis capilar.

Tras el análisis detallado de los resultados obtenidos, no se encontró variación en la longitud de los fragmentos obtenidos en ninguno de los individuos paciente o control analizados, con excepción del paciente en el que se había identificado inicialmente la delección de 9 nucleótidos en el tramo de repeticiones localizado en el intrón s1.

I.4 Análisis del gen *FOXP2* en niños con TEL

El análisis del gen *FOXP2* en niños con Trastorno Específico del Lenguaje (TEL) cubrió dos partes. En primer lugar, se buscaron las tres mutaciones puntuales descritas hasta el momento en región codificante, así como posibles expansiones en los tramos de poliglutaminas. En segundo lugar, y dada la baja presencia de mutaciones en región codificante para este gen, se procedió a analizar la región promotora, buscando mutaciones no presentes en población control.

I.4.1 Búsqueda de mutaciones en región codificante

Se buscaron las mutaciones en región codificante del gen *FOXP2* descritas hasta la fecha: mutación R553H en el exón 14, directamente relacionada con alteraciones en el

lenguaje en la familia KE (Lai *et al*, 2001); un cambio no sinónimo en el exón 2 y la mutación R328X en exón 7 encontradas en pacientes con dispraxia verbal (MacDermot *et al*, 2005) y posibles expansiones del tramo de poliglutaminas en los exones 5 y 6 (Figura R14).

La búsqueda de la mutación en el exón 14 se llevó a cabo mediante análisis de RFLPs. (apartado II.3 de Material y Métodos). En el caso de los exones 2 y 7 se amplificó por PCR, y se secuenciaron los productos obtenidos. La determinación de posibles expansiones de poliglutaminas se realizó tal y como se ha descrito en el apartado anterior, por PCR con uno de los cebadores marcado con un fluorocromo y análisis de los fragmentos marcados obtenidos por electroforesis capilar (ver apartado II.4 de Material y Métodos).

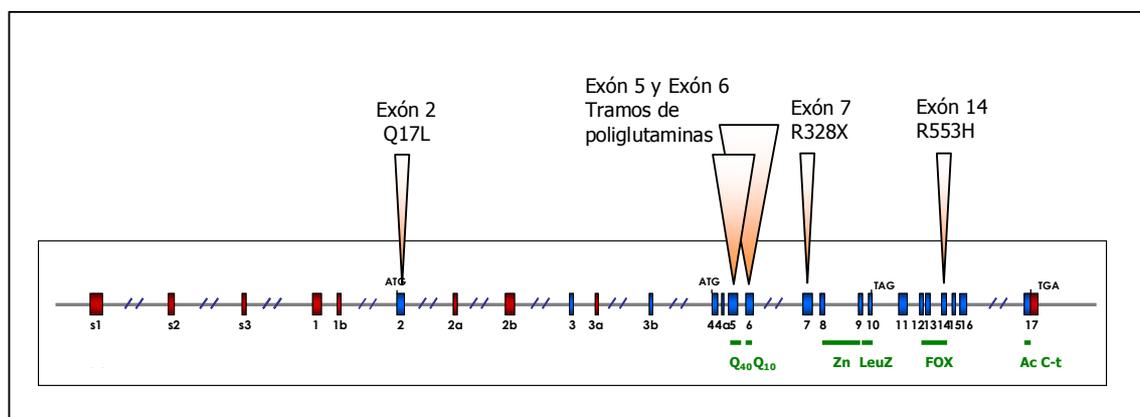


Figura R14. Localización de los cambios analizados en región codificante del gen *FOXP2*.

En los 10 individuos con TEL analizados no se encontró ninguna de las mutaciones en región codificante descritas hasta la fecha ni expansiones en los tramos de poliglutaminas.

1.4.2 Búsqueda de mutaciones en la región promotora

Utilizando la estrategia del paseo cromosómico con cebadores se obtuvieron las secuencias correspondientes a 6.4 kilobases de la región promotora del gen *FOXP2* situada en los alrededores del exón s1 de 10 niños con TEL. En concreto se obtuvo el fragmento incluido entre -4672 y +1619, respecto a la primera base del exón s1.

En la tabla R37 se muestran los polimorfismos descritos en la base de datos dbSNP del NCBI, su localización respecto al inicio de la transcripción del gen *FOXP2* y los genotipos correspondientes para cada paciente estudiado.

En la tabla R38 se muestran las variaciones encontradas en los pacientes que no aparecen en la base de datos del NCBI, así como su presencia en una muestra independiente de 10 controles.

Resultados

SNP	Localización respecto exón S1 (pb)	genotipos	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
rs6466478	- 4282	A/G	A	A	A	A	A	A	A	A	A	A
rs34578815	- 4093	-/A	A	A	A	A	A	A	A	A	A	A
rs7803667	- 3750	A/T	A	A	T	T/A	T	A	T	T/A	T/A	T/A
rs10447760	- 3117	C/T	C	T	C	C/T	C	C/T	C	C/T	C	C/T
rs1845393	- 2418	A/C	C	C	C	C	C	C	C	C	G	C
rs34594627	- 2246	-/T	T	T	T	T	T	T	T	T	T	T
rs11305675	- 1745	-/G	G	G	G	G	G	G	G	G	G	G
rs36094980	- 1463	-/T	T	T	T	T	T	T	T	T	T	T
rs13308496	- 1043	G/T	T	T	T	T	T	T	T	T	T	T
rs13308163	- 989	C/T	C	C	C	C	C	C	C	C	C	C
rs7784307	- 456	A/C	C	C	C	C	C	C	C	C	C	C
rs10539256	-69	-/ACAC	-	ACAC	ACAC	- /ACAC	ACAC	-	ACAC	-	ACAC	- /ACAC
rs13309866	+ 315	A/C	C	C	C	C	C	C	C	C	C	C
rs13309912	+ 355	C/G	C	C	C	C	C	C	C	C	C	C
rs13310148	+ 666	C/G	G	G	G	G	G	G	G	G	G	G
rs35182478	+ 1143	-/C	C	C	-	C/-	-	C	-	C	C/-	C/-
rs35017137	+ 1145	G/T	G	G	T	G/T	T	G	T	G	G/T	G/T
rs6961558	+ 1492	A/G	G	G	G	G	G	G	G	G	G	G
rs6961633	+ 1616	C/T	C	C	C	C	C	C	C	C	C	C

Tabla R37. Estados de los polimorfismos presentes en la región analizada en los niños con TEL analizados.

Localización	Base de datos	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	Presencia en controles
- 3954 pb	G	G	G	A/G	A/G	A/G	G	G	G	A/G	A/G	Sí
- 1048 pb	C	C/G	C	C	C	C	C	C	C	C	C	No
+ 147 pb	C	C/-	C	C	C	C	C	C	C	C	C	No

Tabla R38. Variaciones encontradas en los niños con TEL ausentes en la base de datos.

El cambio localizado en posición -3954 se encontró en controles, por lo que probablemente refleje un SNP presente en la población, y no descrito todavía. En el caso de los otros dos polimorfismos, debido a su ausencia en controles queda por determinar si se trata de SNPs de baja frecuencia o cambios específicos del individuo N1.

II. Análisis evolutivo de la región promotora del gen *FOXP2* en diferentes especies de primates

Se planteó realizar un análisis evolutivo de la región promotora del gen *FOXP2* por varias razones. En primer lugar, nos propusimos analizar la variabilidad de la región 5' por su posible incidencia en esquizofrenia, complementando el análisis de los polimorfismos de la región que se habían analizado en el estudio caso-control del apartado anterior. Dada la alta conservación de la región codificante, estrictamente sometida a selección purificadora, variaciones en la región reguladora del gen podría afectar a su expresión. En segundo lugar, estudiar la variabilidad presente en primates, con el fin de determinar la existencia de hechos diferenciales respecto a la especie humana. Por último, se planteó la búsqueda de regiones conservadas que pudieran constituir elementos funcionales, así como la determinación bioinformática del promotor. Este trabajo continua y complementa el trabajo presentado por Sirena Soriano en su tesis de máster (Soriano, 2007).

II.1 Obtención de las secuencias

Utilizando la estrategia del paseo cromosómico con cebadores se obtuvieron las secuencias correspondientes a aproximadamente 6.4 kilobases de la región promotora del gen *FOXP2* situada en los alrededores del exón s1 en 9 especies de primates:

- *Gorilla gorilla*
- *Macaca mulatta*
- *Macaca nigra*
- *Macaca nemestrina*
- *Macaca fascicularis*
- *Macaca arctoides*
- *Saguinus labiatus*
- *Erythrocebus patas*
- *Lagothrix lagotricha*

Como se comentó en el apartado de Material y Métodos (epígrafe I.2) el DNA de estas especies procedía del *Coriel Institute for Medical Research*, con excepción del de *Gorilla gorilla*, que fue cedido por el Dr. Jaume Bertranpetit.

En el caso de *Macaca mulatta*, no se consiguió obtener un contig completo mediante las amplificaciones de PCRs solapantes. El hueco de 157 pb se cerró a partir de la secuencia obtenida de la base de datos del *UCSC Genome Browser* (Chr3:151,448,158-151,448,457 *Jan 2006 Assembly*).

También partir de bases de datos (NCBI, ver tabla M31 de la sección de Material y Métodos) se extrajeron las secuencias correspondientes a la misma región cromosómica de las siguientes especies:

- *Homo sapiens*
- *Pan troglodytes*
- *Mus musculus*
- *Rattus norvegicus*

Resultados

Con las secuencias de las 11 especies de primates se obtuvo un contig de 6394 pb que cubría la secuencia correspondiente al fragmento comprendido entre las posiciones -4640 y +1618 de la secuencia humana procedente de la base de datos del *UCSC Genome Browser* (Figura R15). Dentro de este contig, las secuencias de todas las especies estaban completas, con excepción de la secuencia de *Lagothrix lagotricha*, que sólo llegaba a la posición +1434.

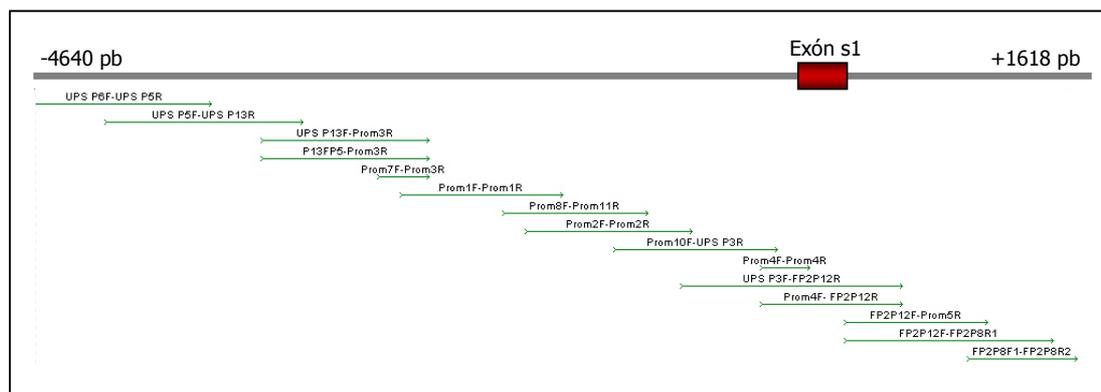


Figura R15. Contig de la región promotora del gen *FOXP2* obtenido por paseo cromosómico. Detalles sobre que fragmentos son específicos de especie en el apartado II.6 de Material y Métodos .

Las secuencias se alinearon inicialmente mediante el programa *ClustalX 1.83*. Posteriormente, se realizaron modificaciones manuales del alineamiento, en aquellas posiciones no resueltas correctamente por el programa.

Al alineamiento obtenido de las secuencias de primates se añadieron las secuencias correspondientes a las especies *Mus musculus* y *Rattus norvegicus*. Este paso se llevó a cabo dentro del programa *ClustalW* implementado en el paquete *MEGA4*, con la opción de mantener los huecos existentes.

La composición nucleotídica de las secuencias se muestra en la tabla R39. Entre primates las proporciones entre los diferentes nucleótidos son similares. Respecto a las especies de roedores analizadas se observa un ligero aumento en las bases TA y un descenso en CG.

	T	C	A	G	Total
<i>Homo sapiens</i>	23.4	27.4	22.2	27.0	6074
<i>Pan troglodytes</i>	23.5	27.5	22.2	26.8	6074
<i>Gorilla gorilla</i>	23.5	27.5	22.2	26.9	6102
<i>Macaca fascicularis</i>	23.6	27.3	22.5	26.6	6064
<i>Macaca arctoides</i>	23.6	27.3	22.5	26.7	6060
<i>Macaca nigra</i>	23.6	27.2	22.4	26.7	6046
<i>Macaca nemestrina</i>	23.6	27.2	22.4	26.8	6040
<i>Macaca mulatta</i>	23.6	27.2	22.4	26.7	6050
<i>Erythrocebus patas</i>	23.5	27.3	22.5	26.7	6050
<i>Saguinus labiatus</i>	24.1	26.9	22.3	26.7	6033
<i>Lagothrix lagotricha</i>	23.8	27.1	22.6	26.6	5988
<i>Mus musculus</i>	27.0	25.5	23.3	24.2	6200
<i>Rattus norvegicus</i>	25.7	25.0	24.0	25.3	6057

Tabla R39. Composición nucleotídica de las secuencias analizadas.

II.2 Análisis de los estados de los polimorfismos humanos en la secuencias de primates analizadas

La mayoría de los polimorfismos de la región estudiada consisten en cambios de base, el resto son inserciones o deleciones de una base, con la excepción de rs10539256 que se trata de una inserción/delección de 4 pb (tabla R40). El análisis del estado de los polimorfismos humanos en las secuencias de primates analizadas reveló que los SNPs parecen fijados en primates y, en la mayoría de los casos, las especies de primates no humanos comparten uno de los estados presentes en humanos, en concreto el correspondiente al alelo de mayor frecuencia en humanos (en los SNPs de los que se pudo obtener datos de frecuencia de bases de datos). Los casos en los que no comparten uno de estos estados se encuentran en las especies más alejadas filogenéticamente de *Homo sapiens* (*Saguinus labiatus* y *Lagothrix lagotricha*) o, como se observa para el polimorfismo rs11305675, se encuentran dentro de una misma rama filogenética. Para este SNP, se observa una adenina en la rama que comparten las especies del género *Macaca* con *Erythrocebus*, en lugar de la inserción o delección de una guanina presente en *Homo sapiens*, lo que indica la existencia de una mutación producida tras la separación de estas especies del resto de primates.

La mayor variabilidad se encuentra en el polimorfismo rs10539256, que en la especie humana consiste en la delección/inserción de la secuencia ACAC dentro de una secuencia con 22 repeticiones AC seguidas interrumpidas por dos AG. El carácter de microsatélite de esta región explica la elevada variabilidad observada entre las especies de primates, donde se observa desde una repetición AC menos en *Gorilla gorilla* que en *Homo sapiens* hasta 8 repeticiones en *Macaca fascicularis*. El elevado grado de polimorfismo también se observa a nivel intraespecífico, ya que trabajando con una única muestra de DNA por especie de primate aparece polimórfico en dos de las especies analizadas, *Macaca nemestrina* y *Saguinus labiatus*.

II.3 Búsqueda de regiones conservadas

La búsqueda de regiones conservadas en las secuencias analizadas se llevó a cabo utilizando el método de *phylogenetic shadowing*, que permite identificar elementos funcionales en especies cercanas, por ejemplo, en primates.

Las comparaciones estándar por pares de secuencias entre secuencias con el 95% o más de identidad no permiten la discriminación entre regiones que evolucionan rápida o lentamente debido a la baja densidad de mutaciones. El *phylogenetic shadowing* compara varias secuencias cercanas al mismo tiempo y combina las mutaciones de todas las secuencias en un único perfil de conservación. Si las mutaciones tienen lugar independientemente en diferentes linajes, la distribución será diferente en las secuencias. Al comparar estas secuencias muy cercanas, se considera que un nucleótido de la secuencia de referencia, ha divergido (o ensombrecido, de donde viene el concepto de *shadowing*) en el conjunto de secuencias, si ese nucleótido no coincide con el que ocupa la misma posición en el resto de especies del alineamiento múltiple. La densidad de nucleótidos divergidos se espera que sea menor en regiones funcionales, con menor tasa de evolución debido a la presión selectiva.

Resultados

SNP	Localización respecto exón S1	<i>H. sapiens</i>	<i>P. troglodytes</i>	<i>G. gorilla</i>	<i>M. mulatta</i>	<i>M. nigra</i>	<i>M. arctoides</i>	<i>M. fascicularis</i>	<i>M. nemestrina</i>	<i>E. patas</i>	<i>S. labiatus</i>	<i>L. lagotricha</i>
rs6466478	- 4282	A/G	A	A	A	A	A	A	A	A	A	A
rs34578815	- 4093	-/A	A	A	A	A	A	A	A	A	A	A
rs7803667	- 3750	A/T	T	T	T	T	T	T	T	T	T	T
rs10447760	- 3117	C/T	C	C	C	C	C	C	C	C	-	-
rs1845393	- 2418	A/C	C	C	C	C	C	C	C	C	C	G
rs34594627	- 2246	-/T	-	-	-	-	-	-	-	-	-	-
rs11305675	- 1745	-/G	G	G	A	A	A	A	A	A	G	G
rs36094980	- 1463	-/T	-	-	-	-	-	-	-	-	-	-
rs13308496	- 1043	G/T	T	T	T	T	T	T	T	T	T	T
rs13308163	- 989	C/T	C	C	C	C	C	C	C	C	G	A
rs7784307	- 456	A/C	C	C	C	C	C	C	C	C	C	C
rs10539256	-69	- /ACAC	AC	-AC	ACAC	6 ACs	6 ACs	8 ACs	-/AC	AC	-/AC	AC
rs13309866	+ 315	A/C	C	C	C	C	C	C	C	C	C	C
rs13309912	+ 355	C/G	C	C	C	C	C	C	C	C	C	C
rs10276237	+387	G/T	G	G	G	G	G	G	G	G	G	G
rs13310148	+ 666	C/G	G	G	G	G	G	G	G	G	A	A
rs35182478	+ 1143	-/C	C	C	C	C	C	C	C	C	C	C
rs35017137	+ 1145	G/T	T	T	T	T	T	T	T	T	T	T
rs6961558	+ 1492	A/G	G	G	G	G	G	G	G	G	G	G
rs6961633	+ 1616	C/T	C	C	C	C	C	C	C	C	C	nd

Tabla R40. Estados de los polimorfismos presentes en *Homo sapiens* en los diferentes primates analizados.

eShadow y *Mulan* son dos herramientas que permiten implementar *phylogenetic shadowing*.

La herramienta ***eShadow*** permite el análisis detallado de la conservación y la optimización de los parámetros de conservación para la detección de elementos funcionales en una filogenia dada. Analiza alineamientos de secuencias múltiples procedentes del *ClustalW* en comparaciones por pares de secuencias e identifica regiones que acumulan pequeñas cantidades de mutaciones a lo largo de la evolución. Para distinguir entre regiones funcionales y regiones que evolucionan neutralmente, está equipado con dos métodos de análisis de la conservación: el de Islas HMM (*Hidden Markov Model Islands*), basadas en el modelo oculto de Markov, de acumulación de la distribución de desapareamientos y el del Umbral de Divergencia (*Divergence Threshold*), que utiliza parámetros estándar de longitud/porcentaje de identidad en el análisis de la conservación multispecífico.

La extracción de los elementos conservados se representa en un gráfico donde se muestran los patrones de conservación (figuras R16 y R17).

Mediante la utilización de *eShadow* se identificaron las siguientes regiones con baja tasa de mutación al analizar el alineamiento múltiple de las 13 especies estudiadas:

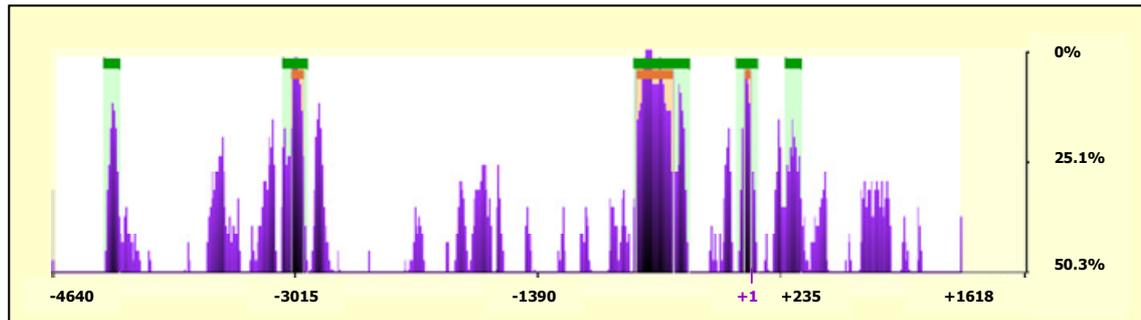


Figura R16. Grado de conservación relativo a la región promotora en las 13 especies analizadas mediante la herramienta *eShadow*. Parámetros por defecto del programa. Barras verdes: DT (umbral de divergencia 20%) Barras naranjas: Islas HMM. Las posiciones del eje de abscisas muestran la distancia respecto al inicio del exón s1 en *H. sapiens*. En el eje de ordenadas se muestra el grado de variación.

Si tenemos en cuenta la posición de las islas HMM, método más restrictivo de los dos utilizados, respecto a la primera base del exón s1 de la secuencia humana se obtuvo que las regiones con menor tasa de mutación se localizan entre las posiciones: (-3050, -2962), (-743, -486) y (-21, +35).

Las especies *Rattus norvegicus* y *Mus musculus* se encuentran muy alejadas de los primates, por lo que para observar qué regiones representan mayor conservación dentro éstos, se repitió el análisis sin estas especies. En este caso, al analizar el alineamiento múltiple con las 11 especies de primates, lo que se obtuvo fue:

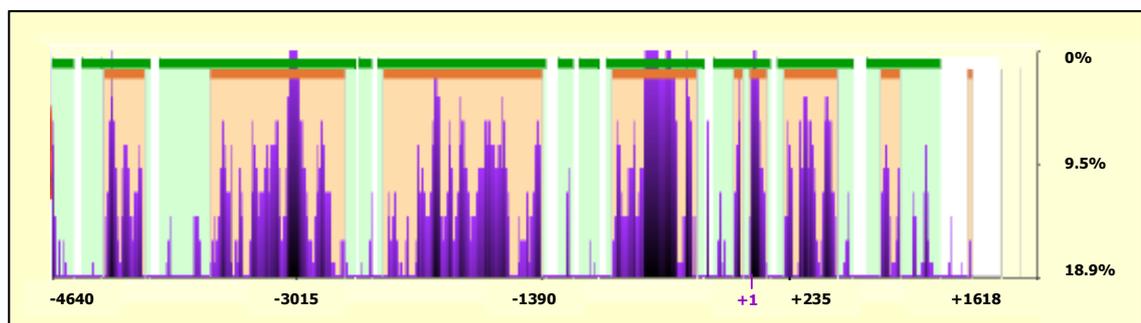


Figura R17. Grado de conservación relativo a la región promotora en las 11 especies de primates analizadas mediante la herramienta *eShadow*. Parámetros por defecto del programa. Barras verdes: DT (umbral de divergencia 20%) Barras naranjas: Islas HMM. Las posiciones del eje de abscisas muestran la distancia respecto al inicio del exón s1 en *H. sapiens*. En el eje de ordenadas se muestra el grado de variación.

En este caso, como era de esperar dada la cercanía de las especies a nivel evolutivo, se obtiene un mayor número de bloques de conservación. La posición de las islas HMM respecto a la primera base del exón s1 en *Homo sapiens* fue la siguiente: (-4303, -4025), (-3595, -2700), (-2440, -1401), (-936, -86), (-52, +70), (+183, +539), (+820, +947), (+1390, +1434).

La herramienta **Mulan** permite detectar regiones conservadas evolutivamente (ECRs, *Evolutionary Conserved Regions*) en los alineamientos que genera a partir de las secuencias que se le incorporan, implementando también *phylogenetic shadowing*.

Resultados

Mediante este programa, en primer lugar, se comparó individualmente la secuencia humana con el resto de secuencias analizadas y se representó la similitud en modo gráfico (figura R18). En estas comparaciones, como era de esperar, se obtuvo mayor grado de conservación al comparar *Homo sapiens* con las secuencias de primates que con las correspondientes a *Mus musculus* y *Ratus norvegicus*. En la región analizada, la especie que más se asemeja a nivel de identidad a *Homo sapiens* es *Gorilla gorilla*, pese a que filogenéticamente la más cercana sea *Pan troglodytes*.

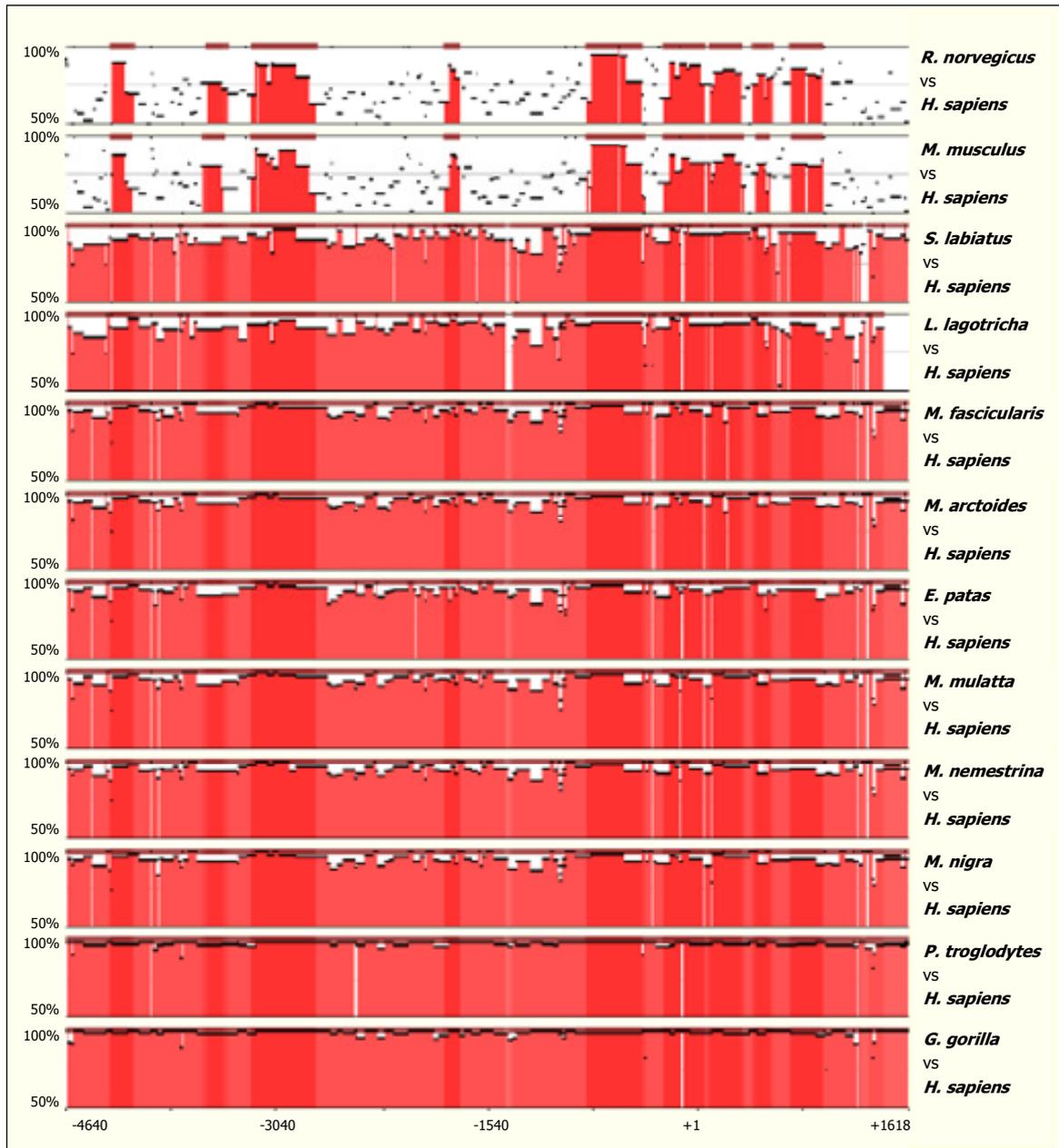


Figura R18. Comparación por pares de secuencias de la secuencia de *H. sapiens* con el resto de especies analizadas. La intensidad del color depende del número de especies que comparten la región (cuanto más oscuro mayor número). Las comparaciones aparecen por orden descendente de similitud. Parámetros utilizados del programa *Mulan*: Longitud mínima ECR 100 pb, similitud mínima ECR 80%. Las posiciones del eje de abscisas muestran la distancia respecto al inicio del exón s1 en *H. sapiens*. En el eje de ordenadas se muestra la medida de la conservación.

En segundo lugar, se analizaron los alineamientos de las 13 especies y se obtuvieron los bloques de conservación entre especies. La implementación práctica del *phylogenetic shadowing* en Mulan está basada en la diferenciación de los nucleótidos ensombrecidos respecto a los completamente conservados (que son exactamente los mismos en todas las secuencias del alineamiento) y el tratamiento de ellos como un conjunto de apareamientos o desapareamientos que se proyectan en la secuencia de referencia (Ovcharenko *et al.*, 2005). Una ventana de 100 pb se desliza a lo largo del conjunto de nucleótidos y permite la representación de los nucleótidos conservados en un eje vertical. Después de escanear todas las posiciones de la secuencia de referencia, se crea un perfil en el que se muestran los ECR detectados con un límite umbral determinado definido por el usuario (figuras R19, R20 y R21).

Al comparar las 13 especies analizadas (Figura R19) se encontraron los siguientes bloques de conservación entre especies: (-3127, -2930), (-775, -401) y (-59, +78).

Nuevamente, con el fin de buscar secuencias conservadas en el linaje de los primates se eliminaron *Rattus norvegicus* y *Mus musculus* del análisis. En este caso se obtuvieron bloques de conservación que ocupan prácticamente la totalidad de la longitud analizada (Figura R20).

Para poder delimitar bloques discretos en secuencias tan cercanas filogenéticamente, se aumentó la similitud mínima necesaria a un 90% (Figura R21). Con estas condiciones se obtuvieron los siguientes bloques: (-4303, -4060), (-3562, -3427), (-3336, -2922), (-2211, -1604), (-1526, -1395), (-957, -391), (-60, +78) y (+264, +548).

De forma resumida, se obtienen tres regiones en las que el grado de similitud entre las 13 especies analizadas es mayor del 80%.

En cualquiera de los análisis realizados por ambos métodos se observa un elevado grado de conservación en torno a (-3127, -2930), (-775, -401) y (-59, +78), por lo que presumiblemente en esa región podría haber elementos reguladores importantes. Este elevado grado de conservación hace pensar que estas regiones deben de contener elementos reguladores importantes, al menos en mamíferos. Este método se ha utilizado otras veces para detectar elementos funcionales (Boffelli *et al.*, 2003).

En cuanto a las secuencias correspondientes a los primates, aumentando la similitud mínima de las regiones conservadas evolutivamente, se obtienen 8 bloques que pueden contener regiones de interés.

Resultados

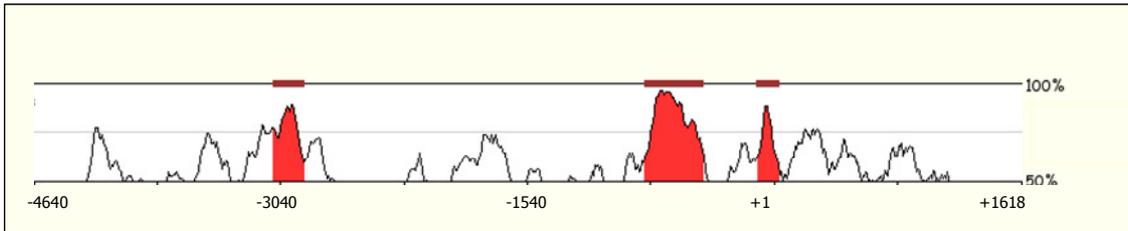


Figura R19. *Phylogenetic Shadowing*, implementado con el programa *Mulan* para el alineamiento múltiple con las 13 especies estudiadas. Longitud mínima ECR 100 pb, similitud mínima ECR 80%. Las posiciones del eje de abscisas muestran la distancia respecto al inicio del exón s1 en *H. sapiens*. En el eje de ordenadas se muestra la medida de la conservación.

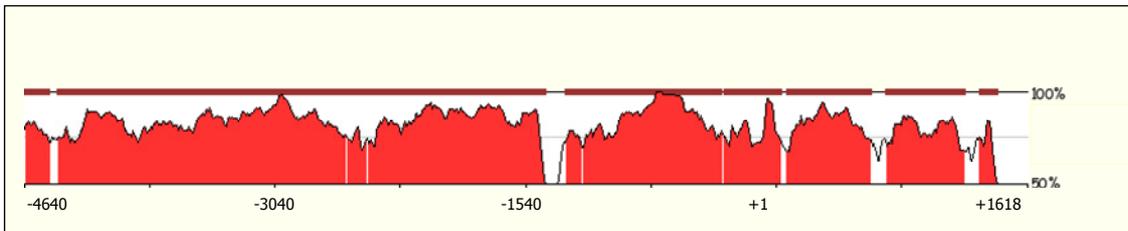


Figura R20. *Phylogenetic Shadowing*, implementado con el programa *Mulan* para el alineamiento múltiple con las 11 especies de primates estudiadas. Longitud mínima ECR 100 pb, similitud mínima ECR 80%. Las posiciones del eje de abscisas muestran la distancia respecto al inicio del exón s1 en *H. sapiens*. En el eje de ordenadas se muestra la medida de la conservación.

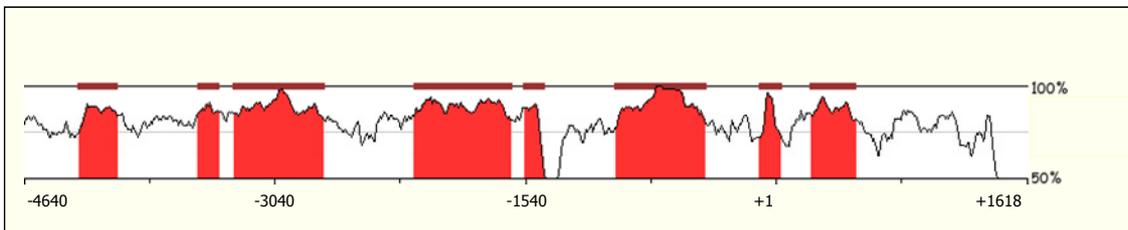


Figura R21 *Phylogenetic Shadowing*, implementado con el programa *Mulan* para el alineamiento múltiple con las 11 especies de primates estudiadas. Longitud mínima ECR 100 pb, similitud mínima ECR 90%. Las posiciones del eje de abscisas muestran la distancia respecto al inicio del exón s1 en *H. sapiens*. En el eje de ordenadas se muestra la medida de la conservación.

Además de los resultados obtenidos mediante los programas *e-Shadow* y *Mulan*, nos parece importante destacar que dentro del mayor pico de conservación obtenido por ambos métodos (-775, -401), se observó en los alineamientos una región muy conservada en la región situada entre -749 y -537 respecto a la primera posición del exón s1 de la secuencia humana. Esta región, que se muestra en la figura R22 llamó la atención debido a la existencia de un único cambio en los alineamientos de primates.

	-749		-729		-709		-689												
<i>Homo sapiens</i>	A	A	A	T	T	T	A	C	T	T	G								
<i>Pan troglodytes</i>								
<i>Gorilla gorilla</i>								
<i>Macaca fascicularis</i>								
<i>Macaca arctoides</i>								
<i>Macaca nigra</i>								
<i>Macaca nemestrina</i>								
<i>Macaca mulatta</i>								
<i>Erythrocebus patas</i>								
<i>Saguinus labiatus</i>								
<i>Lagothrix lagotricha</i>								
<i>Mus musculus</i>	.	A	.	A	.	.	A								
<i>Rattus norvegicus</i>	.	A	.	A	.	.	A								
	-674		-654		-634		-614												
<i>Homo sapiens</i>	C	A	G	C	A	G	C	T	C	C	A	A	T	T	G	C	C	A	A
<i>Pan troglodytes</i>
<i>Gorilla gorilla</i>
<i>Macaca fascicularis</i>	T
<i>Macaca arctoides</i>	T
<i>Macaca nigra</i>	T
<i>Macaca nemestrina</i>	T
<i>Macaca mulatta</i>	T
<i>Erythrocebus patas</i>	T
<i>Saguinus labiatus</i>	T
<i>Lagothrix lagotricha</i>	T
<i>Mus musculus</i>	G
<i>Rattus norvegicus</i>	CTGG
	-599		-579		-559		-539												
<i>Homo sapiens</i>	A	C	A	G	C	A	T	A	C	G									
<i>Pan troglodytes</i>								
<i>Gorilla gorilla</i>								
<i>Macaca fascicularis</i>								
<i>Macaca arctoides</i>								
<i>Macaca nigra</i>								
<i>Macaca nemestrina</i>								
<i>Macaca mulatta</i>								
<i>Erythrocebus patas</i>								
<i>Saguinus labiatus</i>								
<i>Lagothrix lagotricha</i>								
<i>Mus musculus</i>								
<i>Rattus norvegicus</i>								

Figura R22. Alineamiento de la secuencia comprendida entre las posiciones -749 y -537 respecto al inicio del exón s1 de la secuencia humana.

II.4 Delimitación de la región promotora

A partir del programa de anotación genómica *El Dorado* de *Genomatix* se extrajeron los promotores anotados situados en la región 5' del gen *FOXP2*, de las especies *H. sapiens*, *P. troglodytes*, *M. mulatta* y *M. musculus*.

Para el resto de las especies analizadas, no disponibles en *El Dorado*, se utilizó el programa *PromoterInspector*, para obtener predicciones de regiones promotoras. Este programa identifica regiones que contienen promotores, basándose en el contexto genómico de los promotores para la RNA polimerasa II.

En la tabla R41 se muestran los promotores extraídos de *El Dorado*, así como las predicciones obtenidas para cada una de las especies. En todos los primates se obtuvieron 3 predicciones dentro de la región analizada. En la figura R23 se muestra la posición de los promotores extraídos, así como la de las predicciones obtenidas, respecto al primer exón del gen.

Resultados

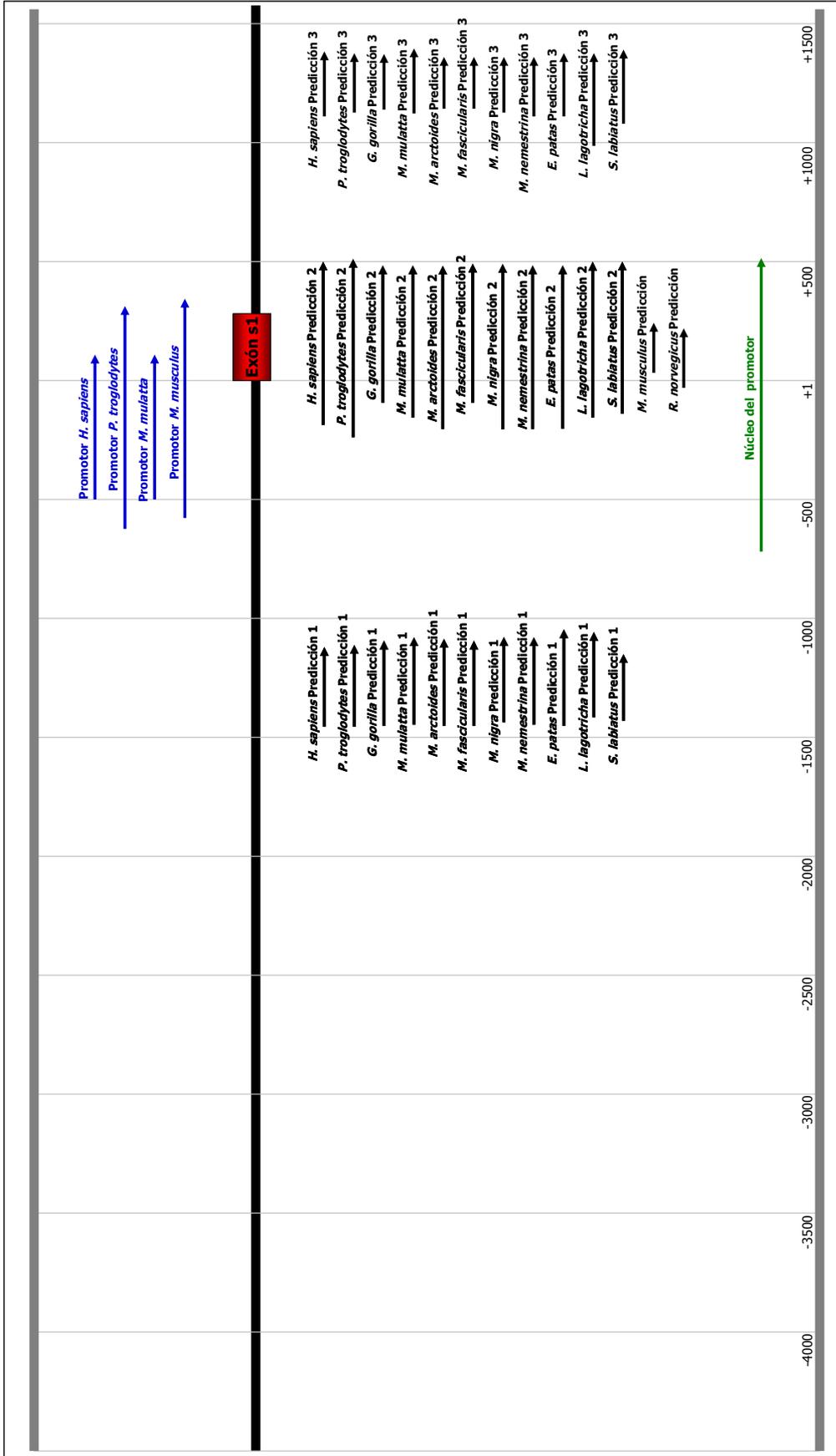


Figura R23. Localización de los promotores extraídos de *El Dorado* (marcado en azul) y las predicciones de promotores obtenidas con el *PromoterInspector* de *Genomatix* (marcado en negro). En verde se muestra la región que proponemos como núcleo del promotor. Las distancias respecto a la primera base del exón s1 se muestran en pares de bases.

Promotores extraídos						
Especie			Promotor	Tamaño en pb		
<i>H. sapiens</i>			-500 → 102	602		
<i>P. troglodytes</i>			-653 → 306	959		
<i>M. mulatta</i>			-500 → 102	602		
<i>M. musculus</i>			-628 → 301	929		
Predicciones						
Especie	Predicción 1	Tamaño en pb	Predicción 2	Tamaño en pb	Predicción 3	Tamaño en pb
<i>H. sapiens</i>	-1480 → -1157	324	-184 → 508	692	1173 → 1444	272
<i>P. troglodytes</i>	-1488 → -1165	324	-248 → 508	756	1189 → 1444	256
<i>G. gorilla</i>	-1488 → -1141	348	-116 → 488	604	1181 → 1396	216
<i>M. mulatta</i>	-1476 → -1129	348	-160 → 488	647	1169 → 1440	216
<i>M. arctoides</i>	-1484 → -1137	348	-288 → 488	716	1183 → 1398	216
<i>M. fascicularis</i>	-1488 → -1141	348	-116 → 488	604	1181 → 1396	216
<i>M. nemestrina</i>	-1471 → -1124	348	-211 → 489	700	1154 → 1397	244
<i>M. nigra</i>	-1476 → -1129	348	-224 → 488	712	1177 → 1396	220
<i>E. patas</i>	-1465 → -1098	368	-213 → 491	704	1156 → 1419	264
<i>L. lagotricha</i>	-1452 → -1109	344	-156 → 500	656	997 → 1392	396
<i>S. labiatus</i>	-1442 → -1207	236	-138 → 498	636	1135 → 1430	295
<i>M. musculus</i>			137 → 328	191		
<i>R. norvegicus</i>			-29 → 213	242		

Tabla R41. Predicciones de los promotores presentes en la región analizada para las diferentes especies. Las posiciones que se muestran corresponden a aquellas respecto a la primera base del exón s1 en la secuencia humana. La denominación de predicción 1, predicción 2 o predicción 3 se hizo para diferenciarlas entre sí.

De los datos de conservación obtenidos en el apartado anterior, así como de las predicciones de promotores en la región cromosómica analizada en las diferentes especies, deducimos la existencia de una región promotora central muy conservada en mamíferos que podría constituir el núcleo del promotor, constituida por un fragmento de aproximadamente 1200 bases adyacente al exón s1 (posiciones entre -749, +507). Este fragmento incluiría la región de elevado grado de conservación (-749, -537) así como la predicción 2 de las especies de primates y las únicas predicciones de promotores en *M. musculus* y *R. norvegicus* (figura R23). Las predicciones 1 y 3 que se encuentran en primates podrían contener elementos reguladores específicos de este linaje, que contribuyeran a la regulación de la expresión. El hecho de que estas regiones no muestren un grado de conservación tan alto entre primates y roedores como el mostrado en la región cercana al exón s1 apoya esta hipótesis.

II.4.1 Sitios de unión de factores de transcripción en la región promotora

Para determinar la presencia de sitios conservados de unión a factores de transcripción se llevó a cabo una búsqueda de sitios de unión a factores de transcripción comunes en secuencias múltiples mediante la herramienta *GEMs Launcher* de *Genomatix*. La búsqueda se lleva a cabo con *MatInspector*, también herramienta de *Genomatix*, que dispone de una genoteca con matrices para sitios de unión de factores de transcripción.

Resultados

Inicialmente, se aplicó una búsqueda en la secuencia completa de los sitios de unión presentes en las 13 especies estudiadas. Para evitar falsos positivos en lo posible, se aplicaron los límites más restrictivos, como la presencia del sitio de unión en todas las especies y el máximo filtrado de aciertos por similitud de la matriz del sitio de unión de factor de transcripción correspondiente. De este modo se obtuvo una representación de los sitios de unión a factores de transcripción a lo largo de la secuencia analizada, mostrándose cada uno de ellos en diferente color en función de la familia a la que pertenecen (figura R24).

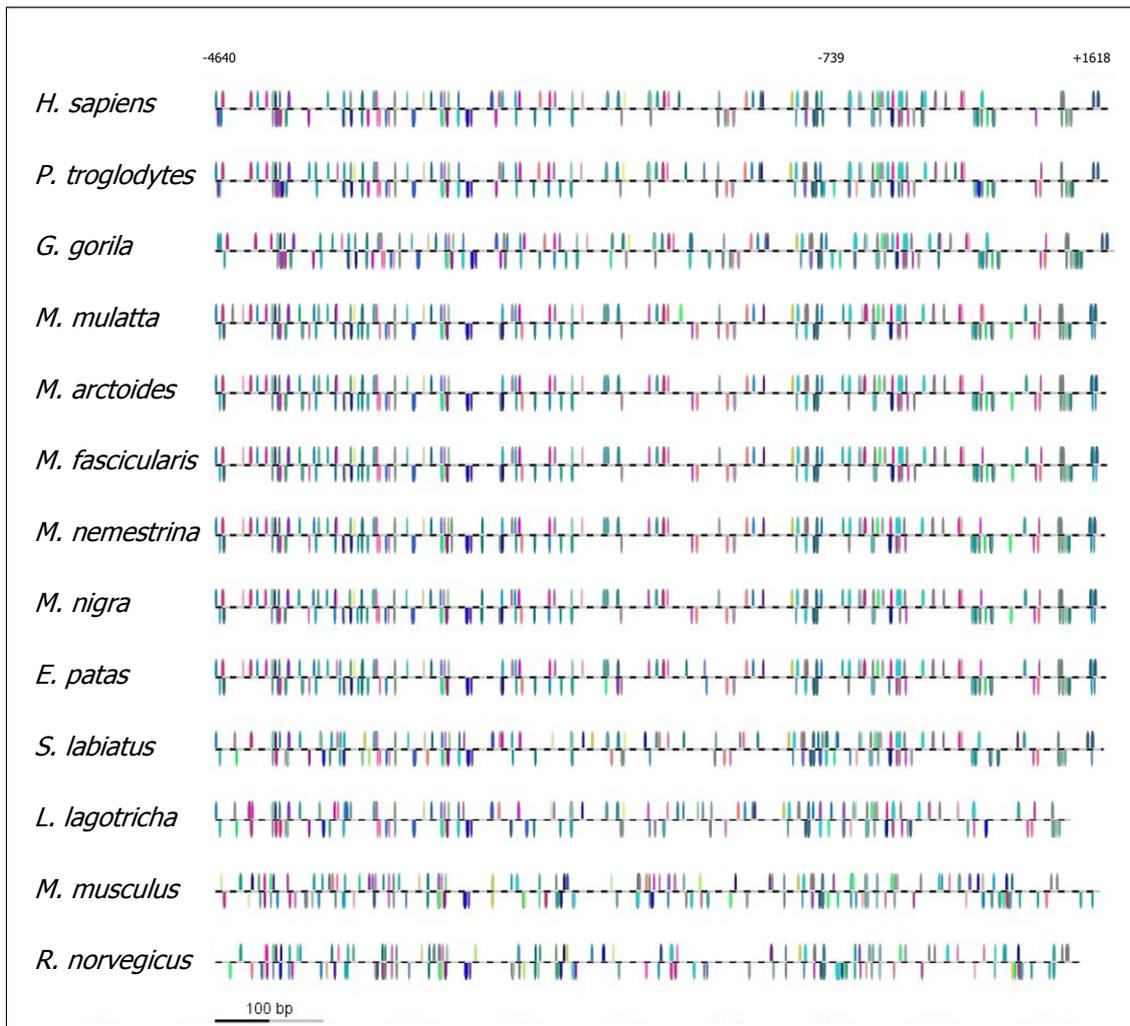


Figura R24. Resultados de la búsqueda de sitios de unión a factores de transcripción presentes en todas las especies en el alineamiento completo. Parámetros seleccionados: presencia en todas las especies, similitud de la matriz 0.05.

A continuación, se analizó la región que proponemos como núcleo del promotor del gen *FOXP2*, comprendida entre -749 y 507. Inicialmente, al igual que al considerar la secuencia completa, se buscaron los sitios comunes presentes en todas las secuencias con los límites más restrictivos (tabla R42, figura R25).

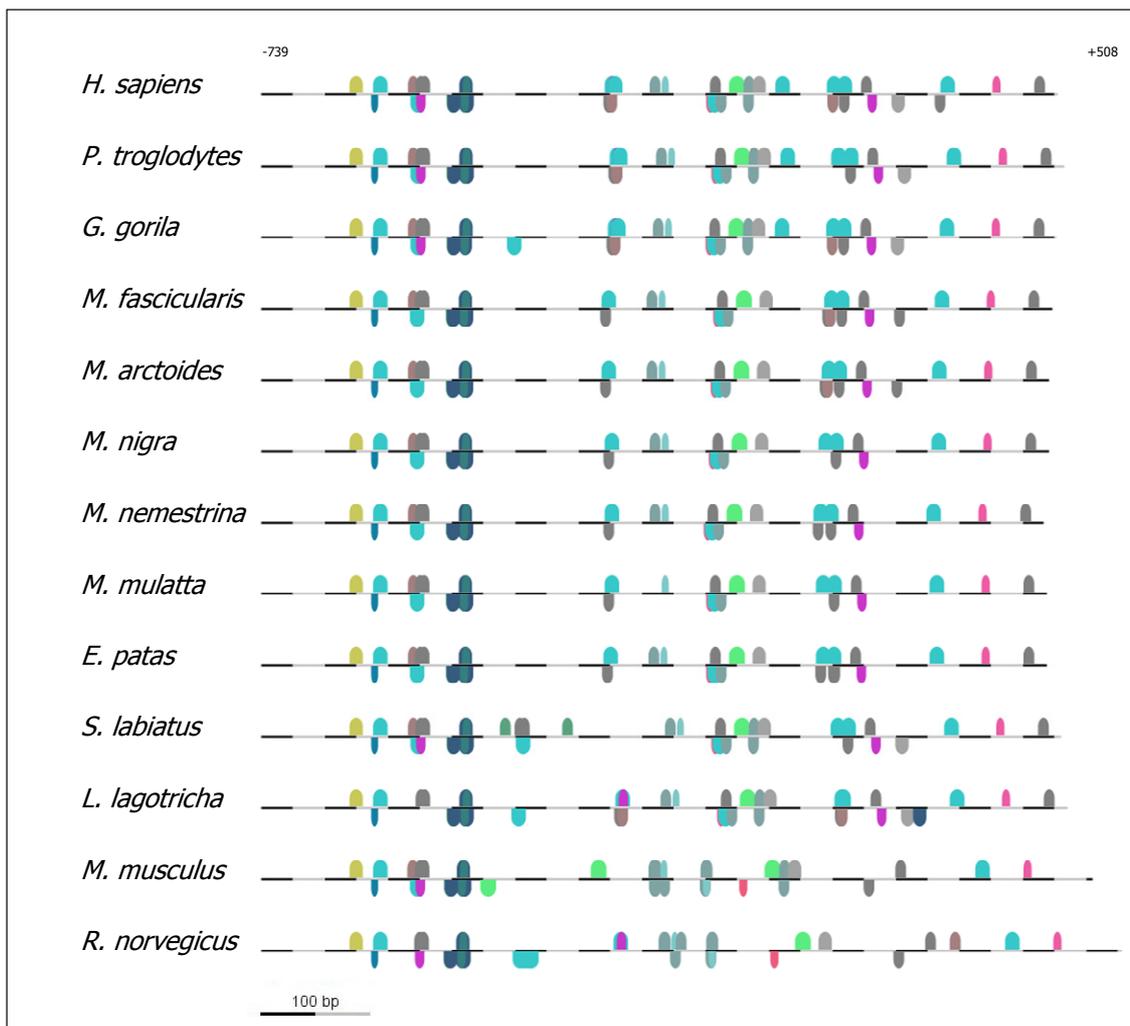


Figura R25. Resultados de la búsqueda de sitios de unión a factores de transcripción comunes para todas las especies en el propuesto núcleo del promotor. Parámetros seleccionados: presencia en todas las especies, similitud de la matriz 0.05

A continuación se llevó a cabo la búsqueda de todos los sitios de unión, aunque únicamente estuvieran presentes en una especie (Figura R26). Así, se determinó que los sitios de unión a factores de transcripción específicos de *H. sapiens* (con similitud 0.05) son los siguientes:

- Dominio POU del factor de unión a octámeros I (OCT-I), en la cadena +
- Complejo heterodimérico E2F4/DP1, en la cadena +
- Complejo heterodimérico E2F4/DP1, en la cadena -
- Producto del gen de expresión temprana *Egr-1*/*Krox-24*/*NGFI-A*

En cuanto a los sitios específicos del linaje humano-chimpancé, son:

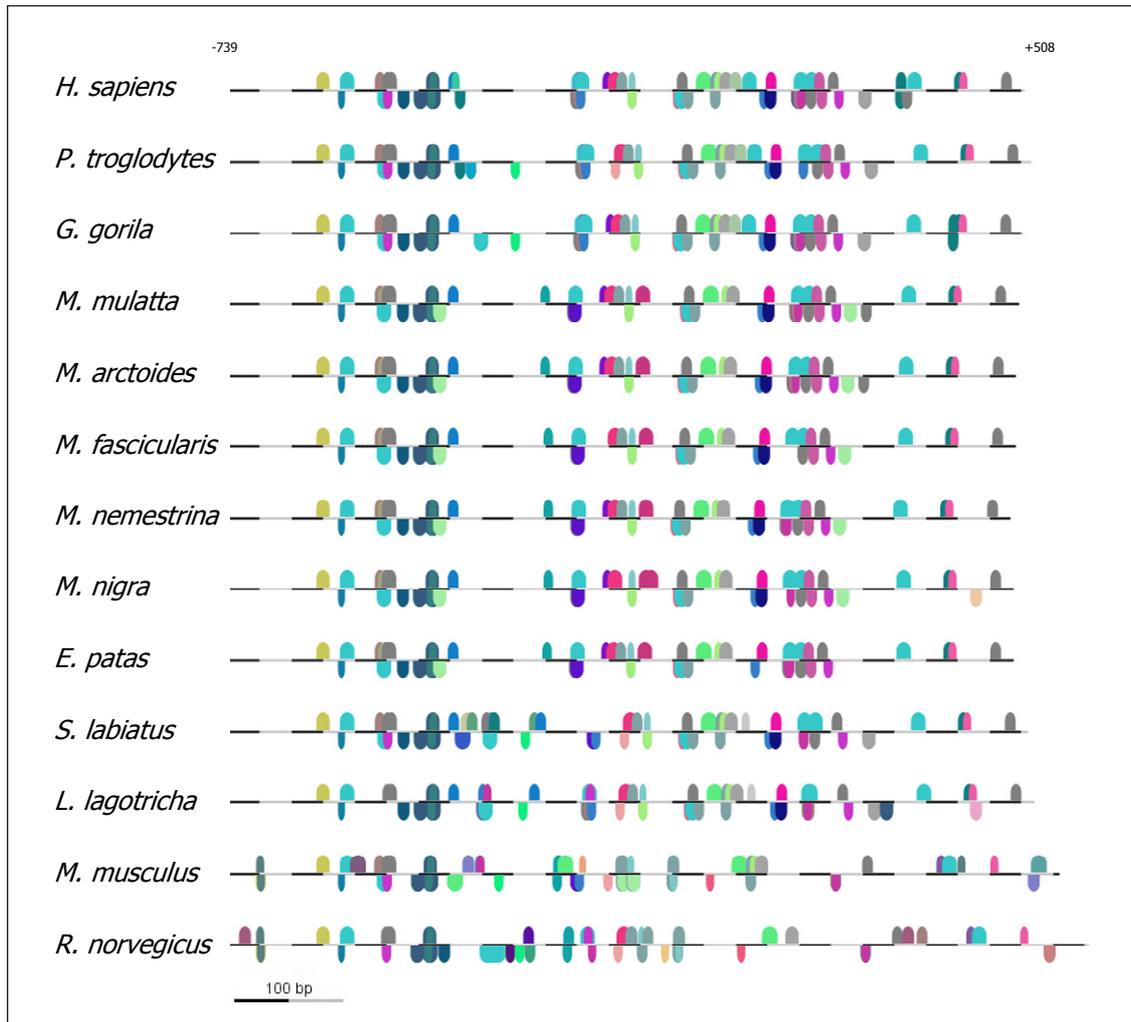
- Complejo heterodimérico E2F1/DP2

Y los sitios específicos del linaje humano-chimpancé-gorila:

- Proteína de unión a elementos de respuesta a Ras
- Factor *Kruppel-like* enriquecido en riñón KLF15

Resultados

- Represor transcripcional ZNF202.01, de unión a elementos presentes predominantemente en genes que participan en el metabolismo lipídico.



En la tabla R42 se recogen los resultados detallados de la búsqueda de sitios de unión para factores de transcripción para el propuesto núcleo del promotor. Se muestra el nombre de la matriz del factor de transcripción según la nomenclatura utilizada por *MatInspector*, así como la descripción de dicho factor, el valor optimizado para el umbral mínimo aceptado para la matriz, la posición respecto al exón s1, la cadena en la que se encuentra el sitio de unión, la similitud del núcleo o posiciones más conservadas de la matriz y la similitud de la matriz completa.

Los resultados que se muestran corresponden a la secuencia humana. En sombreado rojo se representan aquellos sitios de unión que se encuentran presentes en las 13 especies analizadas, aunque puede diferir la posición en cada una. El sombreado verde indica la localización de sitios de unión específicos en la especie humana.

Factor	Descripción de MatInspector	Opt.	Posición respecto exón s1	Cad	Core sim	Mat sim
MARE.02	Sitios de unión para homodímeros de proteínas Maf	0.88	-608, -588	+	0.875	0.930
MZF1	Proteína de dedos de zinc mieloides MZ1	0.95	-574, -564	-	1.000	1.000
ZNF219.01	Proteína de dedos de zinc 219 <i>Kruppel-like</i>	0.91	-571, -549	+	1.000	0.982
			-44, -22	-	1.000	0.992
			+149, +166	+	1.000	0.993
			+161, +183	+	1.000	0.974
KKLF.01	Factor <i>Kruppel-like</i> enriquecido en riñón, KLF15	0.91	-516, -500	+	1.000	0.960
			-203, -187	-	1.000	0.960
			+145, +161	-	1.000	0.963
ZNF202.01	Represor transcripcional de unión a elementos presentes predominantemente en genes que participan en el metabolismo lipídico.	0.73	-512, -490	-	1.000	0.819
			-206, -184	+	1.000	0.838
			+63, +85	+	0.776	0.787
			+333, +345	+	1.000	0.782
ZBP89.01	Factor de transcripción de dedos de zinc ZBP-89	0.93	-506, -484	-	1.000	0.976
WT1.01	Supresor del tumor de Wilms	0.92	-504, -488	+	1.000	0.997
			-498, -482	+	1.000	0.984
			-208, -192	-	1.000	0.987
			-40, -24	+	1.000	0.986
			+163, +179	-	1.000	0.974
RREB1.01	Proteína 1 de unión al elemento de respuesta a Ras	0.80	-503, -489	-	1.000	0.876
			-201, -187	+	1.000	0.877
			+208, +222	-	1.000	0.920
CP2.01	CP2	0.90	-469, -462	-	0.909	0.9585
FKHRL1.01	Factor FKHRL1 de dominio Fkh (FOXO)	0.83	-455, -439	-	1.000	0.989
E4BP4.01	Represor transcripcional E4BP4, dominio bZIP	0.80	-454, -434	-	1.000	0.851
CREB1.01	Proteína de unión al elemento respuesta de cAMP	0.86	-435, -415	+	1.000	1.000
CREB.02	Proteína de unión al elemento respuesta de cAMP	0.89	-433, -413	-	1.000	0.979
E4F.01	Factor regulador del promotor E4 de adenovirus, relacionado con <i>GLI-Krueppel</i>	0.82	-433, -421	-	1.000	0.887
			-430, -418	+	1.000	0.892
SOX5.01	Sox-5	0.87	-400, -384	+	1.000	0.988
OCT1P.01	Dominio POU del factor de unión a octámeros I (OCT-I)	0.86	-395, -383	+	1.000	0.910
E2F1 DP2.01	Complejo heterodimérico E2F-1/DP-2	0.78	-390, -374	-	0.795	0.837
HMX3.01	Factor de transcripción HMX3/Nkx5.1, dominio H6	0.89	-250, -236	+	1.000	0.937
PLAG1.01	Gen adenoma pleomórfico (PLAG) 1, proteína de dedos de zinc C2H2 regulada en desarrollo	0.88	-206, -186	-	1.000	0.928
ZF9.01	Proteína de unión al núcleo del promotor (CPBP) con tres dedos de zinc <i>Krueppel-type</i> .	0.87	-201, -179	+	1.000	0.926
SP1.01	Proteína estimulante 1, factor de transcripción con dedos de zinc, ubicuo.	0.88	-199, -185	-	1.000	0.987
			+90, +104	-	1.000	0.976
PLZ.01	Dedo de zinc de leucemia promielocítica (Factor de transcripción con 9 dedos de zinc <i>Krueppel-like</i>)	0.86	-157, -143	+	1.000	0.913

Tabla R42. Resultados del programa *MatInspector* para el núcleo del promotor. Similitud mínima de la matriz 0.05. Los resultados corresponden a la secuencia humana, indicándose su posición en el contig. El sombreado rojo representa aquellos sitios de unión que se encuentran presentes (aun en distinta posición) en las 13 especies estudiadas. El sombreado en verde en la posición representa las posiciones específicas de humanos.

Resultados

Factor	Descripción de MatInspector	Opt.	Posición respecto exón s1	Cad	Core sim	Mat sim
NRF1.01	Factor nuclear respiratorio 1 (NRF1), factor de transcripción bZIP que actúa sobre genes nucleares que codifican para proteínas mitocondriales.	0.78	-135, -119	+	0.750	0.846
			-31, -15	-	1.000	0.864
			+12, +28	-	1.000	0.836
			+13, +29	+	0.750	0.831
HES1.01	Homólogo a <i>Drosophila hairy/ enhancer of split 1</i> (HES-1)	0.92	-118, -104	-	0.944	0.969
			+19, +33	+	1.000	0.969
ZF5.01	Factor de transcripción: Dedo de zinc/dominio POZ	0.95	-116, -106	+	1.000	1.000
DMP1.01	Proteína <i>myb-like</i> que interacciona con CiclinaD, factor de transcripción myb-like 1 que interacciona con DMTF1-CiclinaD	0.82	-46, -34	-	1.000	0.877
HSF1.03	Factor de choque térmico 1	0.76	-10, +14	+	1.000	0.863
NRSE.01	Elemento neural restrictivo silenciador	0.67	+27, +47	+	0.782	0.78
			+246, +266	-	0.782	0.719
PAX1.01	Proteína de dominio <i>Pax1 paired</i> , que se expresa la columna vertebral del embrión de ratón en desarrollo.	0.62	+43, +61	+	0.750	0.729
BRN5.01	Clase de proteínas Brn-5, POU-VI (también conocida como emb y CNS-1)	0.74	+98, +116	-	1.000	0.838
HOXA9.01	Sitio de unión a PBX- HOXA9	0.79	+100, +116	+	0.750	0.842
CTCF.01	Factor de unión a CCCTC	0.80	+140, +164	-	1.000	0.849
GC.01 (fam SP1)	Elementos de cajas GC	0.88	+144, +158	-	0.819	0.935
CKROX.01	Proteína de colágeno krox (proteína de dedos de zinc 67-zfp67)	0.88	+147, +163	-	1.000	0.920
MAZ.01	Proteína A de dedos de zinc asociada a Myc	0.90	+149, +161	-	1.000	0.959
MYF5.01	Proteína bHLH miogénica Myf5	0.90	+180, +196	-	1.000	0.960
EGR3.01	Producto del gen de respuesta temprana al crecimiento 3	0.77	+198, +214	+	1.000	0.901
			+471, +487	+	1.000	0.850
E2F4 DP1.01	Complejo heterodimérico E2F-4/DP1	0.84	+304, +320	-	1.000	0.969
			+305, +321	+	1.000	0.963
EGR1.01	Producto del gen de expresión temprana inmediata Egr-I/Krox-24/NGFI-A	0.79	+314, +330	-	0.827	0.852
E2F.02	E2F, implicado en la regulación del ciclo celular, interacciona con proteína Rb p107	0.84	+397, +413	+	1.000	0.909
MYT1.02	Factor de transcripción de dedos de zinc MyT1, implicado en neurogénesis primaria	0.88	+405, +417	+	1.000	0.980

Tabla R42. Continuación.

II.5 Análisis filogenético

Se llevaron a cabo reconstrucciones filogenéticas para diversos conjuntos de secuencias, cómo son el alineamiento completo, que se corresponde con la secuencia comprendida entre -4640 y +1618 en *H. sapiens*, las diversas predicciones de regiones promotoras por separado (Predicción 1, Predicción 2 y Predicción 3) obtenidas con el programa *PromoterInspector* y la que hemos propuesto como región promotora nuclear en función de los resultados obtenidos.

Para la obtención de los árboles de parsimonia se utilizó el programa *PaupUp*, mientras que para los de máxima verosimilitud se utilizó el programa *PhyML Online*

(<http://atgc.lirmm.fr/phyml/>), seleccionando el modelo disponible más cercano al obtenido tras implementar la herramienta *Modeltest*.

Los árboles obtenidos por parsimonia mediante el programa *PaupUp* se muestran en las figuras R27 y R28 (A: Secuencia completa, B: Predicción 1, C: Predicción 2, D: Predicción 3, E: Promotor extraído de *EIDorado*, F: Núcleo del promotor) y los obtenidos por máxima verosimilitud mediante el programa *PhyML Online* en las figuras R29 y R30 (A: Secuencia completa, B: Predicción 1, C: Predicción 2, D: Promotor extraído de *EIDorado*, E: Núcleo del promotor).

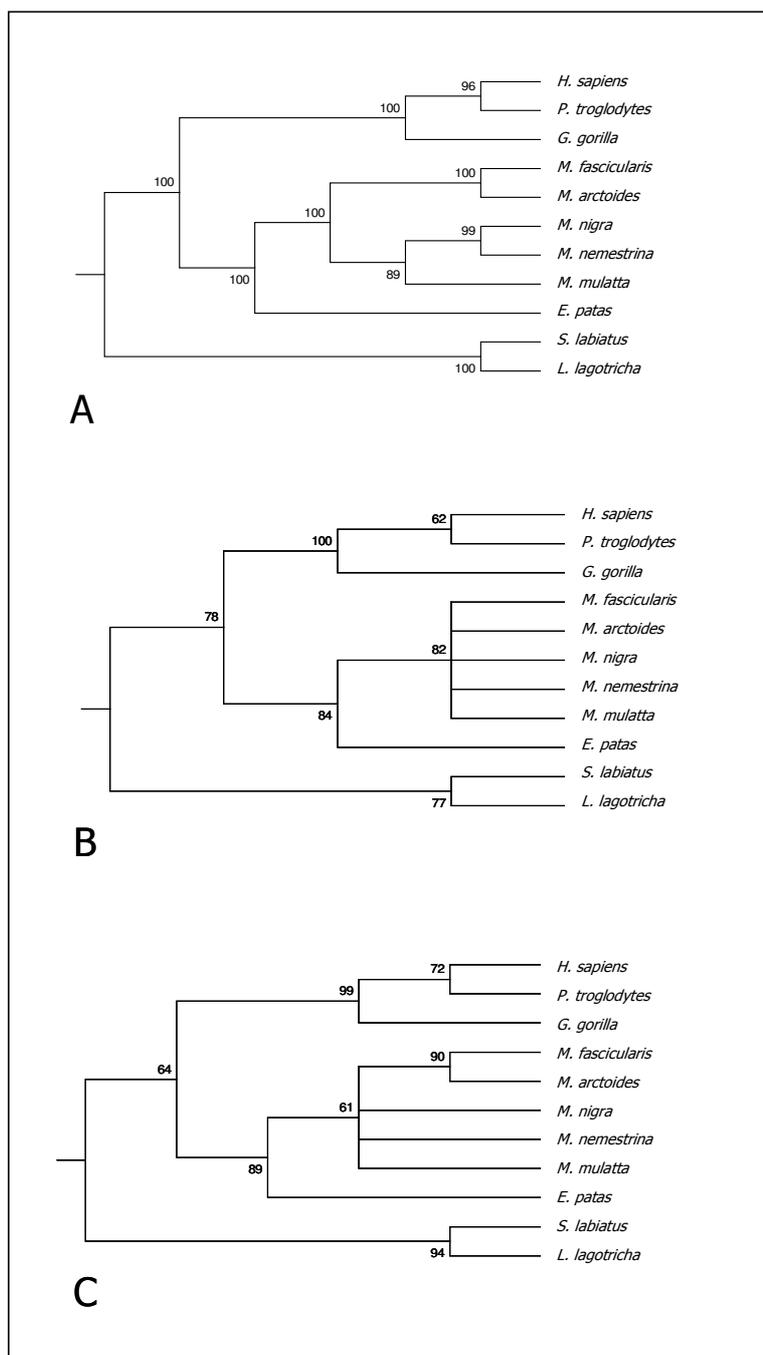


Figura R27. Árboles filogenéticos obtenidos por parsimonia. Bootstrap 1000. A. Secuencia completa (-4640, +1618). B. Predicción 1. C. Predicción 2.

Resultados

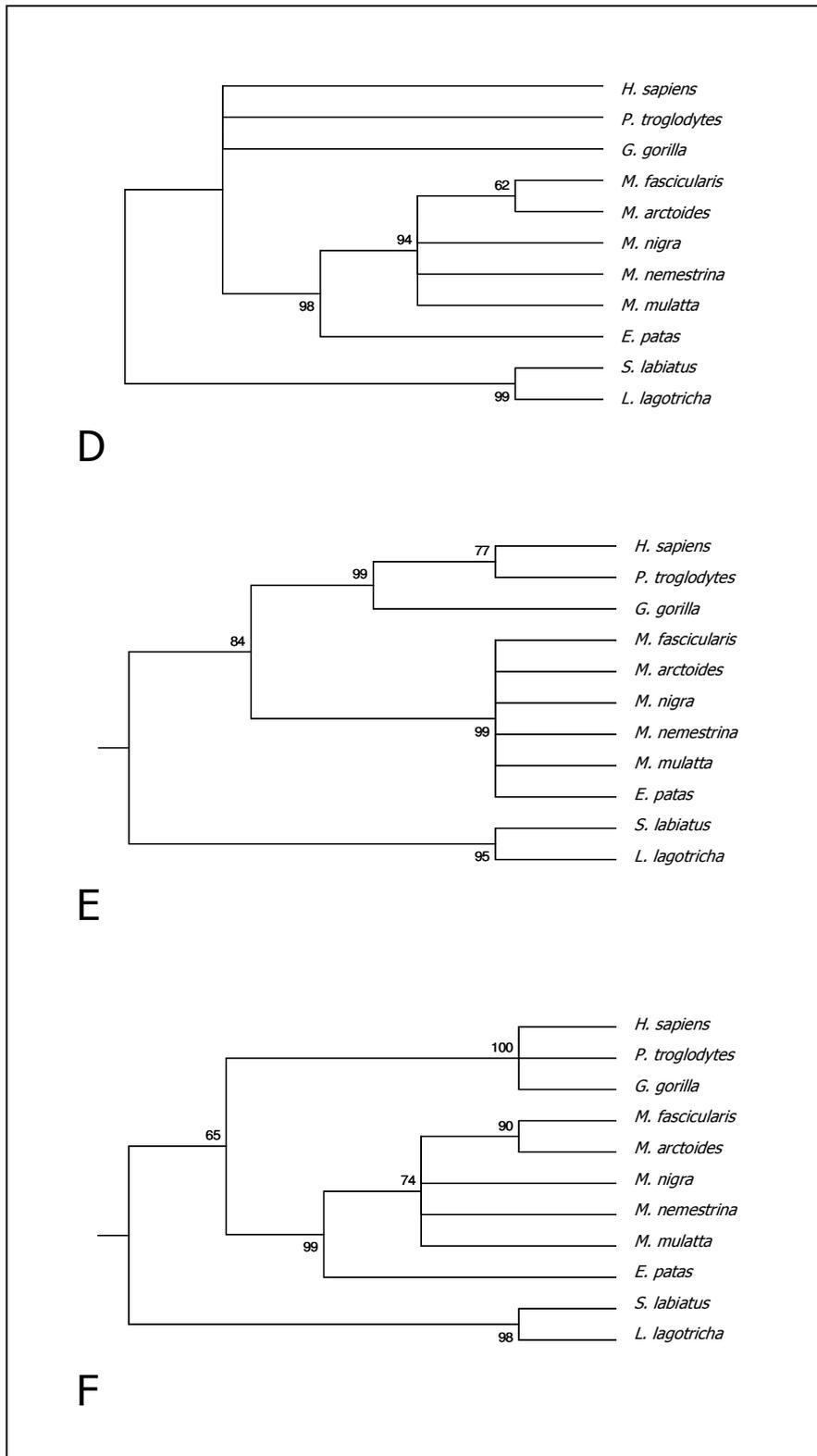


Figura R28. Árboles filogenéticos obtenidos por parsimonia. Bootstrap 1000. D. Predicción 3. E. Promotor. F. Núcleo del promotor.

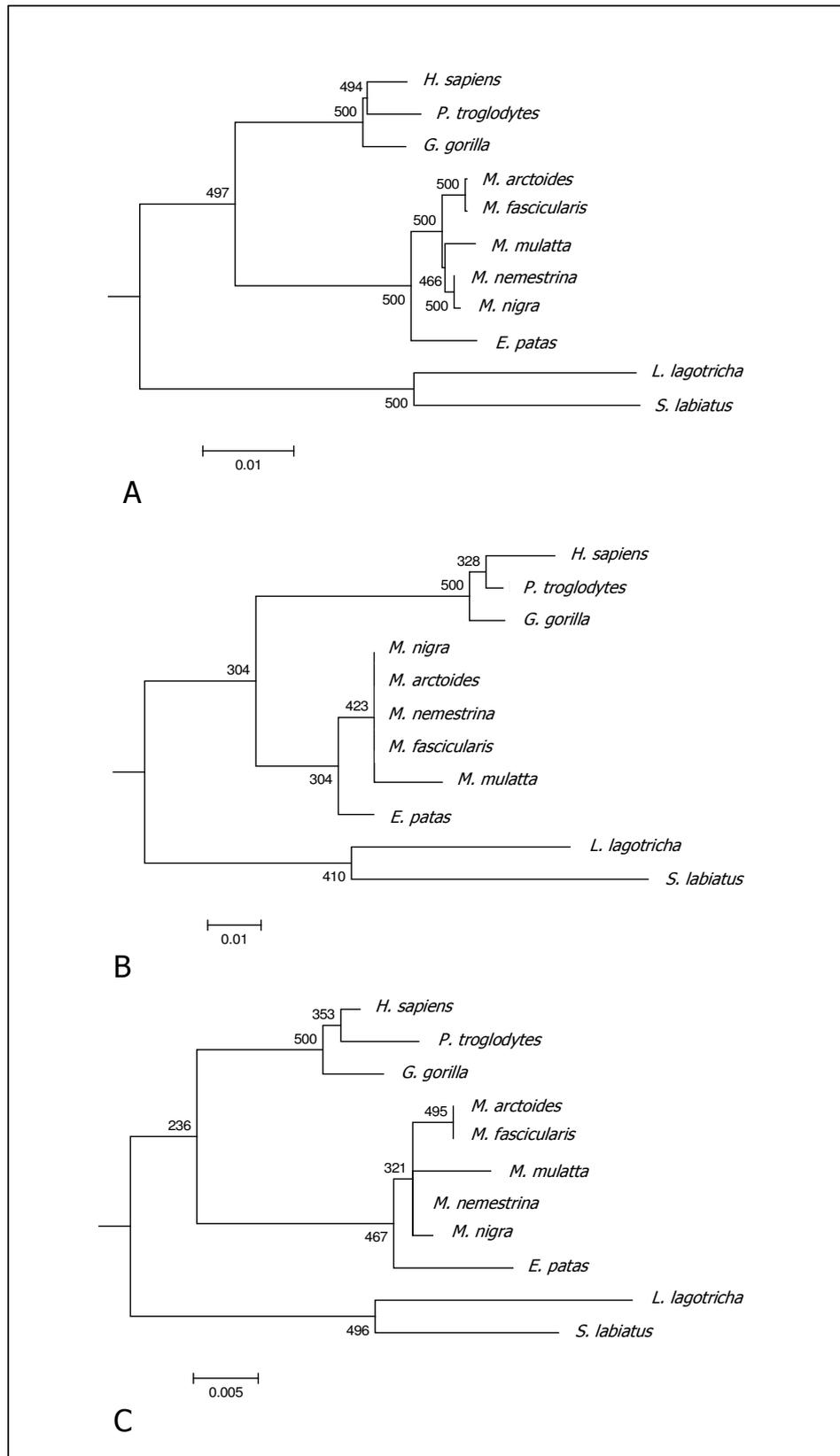


Figura R29. Árboles filogenéticos obtenidos por máxima verosimilitud. Bootstrap 500. A. Secuencia completa. B. Predicción 1. C. Predicción 2.

Resultados

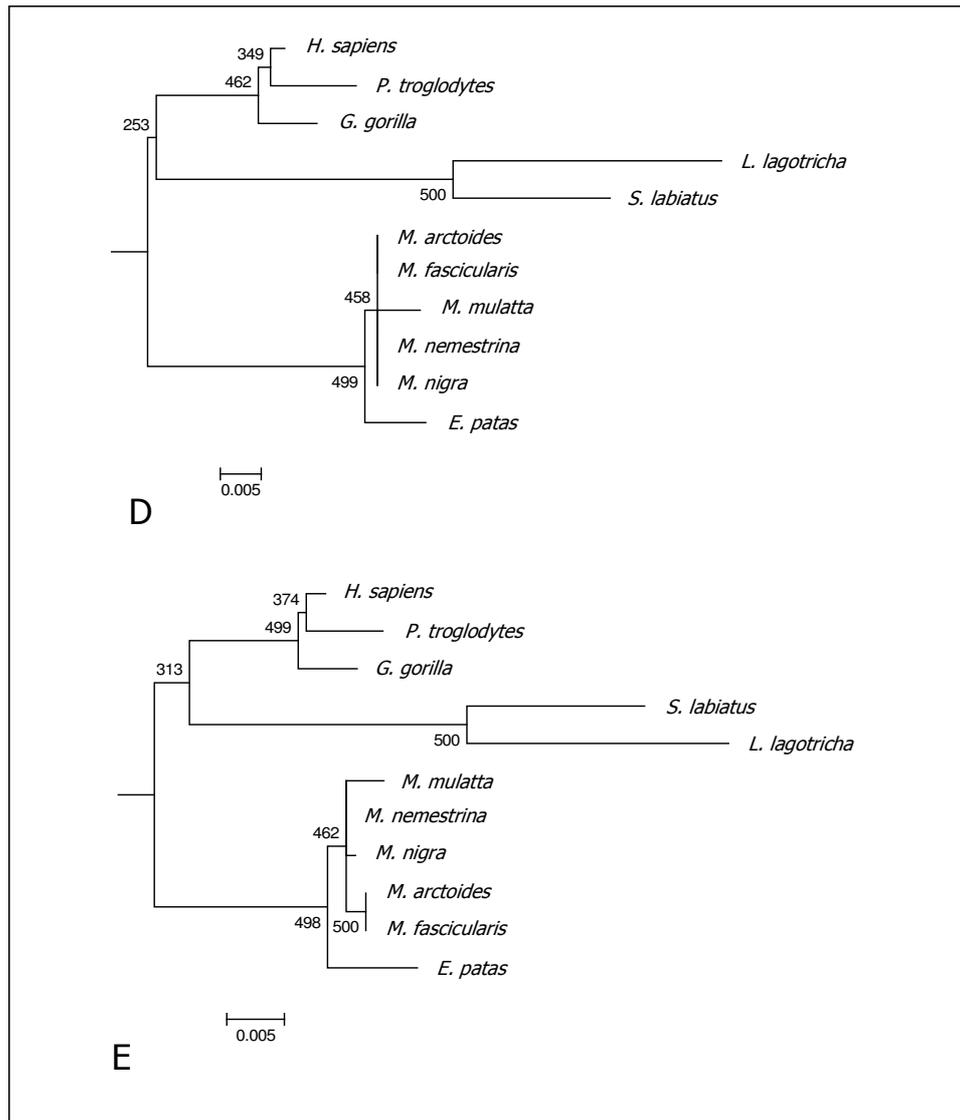


Figura R30. Árboles filogenéticos obtenidos por máxima verosimilitud. Bootstrap 500. D. Promotor. E. Núcleo del promotor.

Los árboles filogenéticos obtenidos se ajustan a los esperados según la filogenia aceptada para estas especies (Purvis, 1995; Goodman *et al.*, 1998; Glazko y Nei, 2003).

La topología de los árboles obtenidos por parsimonia y máxima verosimilitud es similar. En el caso de los árboles obtenidos por el segundo método destacan dos cosas. En primer lugar, aunque los resultados coinciden con los obtenidos por parsimonia para la secuencia completa y para las predicciones 1 y 2, en el caso de la predicción 3 no se pudo obtener un árbol robusto. Además, en los árboles obtenidos para el promotor extraído de *EIDorado* (Figura R30, D) y para el núcleo del promotor (Figura R30, E), la rama correspondiente a las especies de primates más alejadas, *S. labiatus* y *L. lagotricha* aparece junto a la correspondiente a *G. gorilla*, *P. troglodytes* y *H. sapiens*, en lugar de mostrarse como rama más alejada. Sin embargo, los valores bajos de *bootstrap* para esa diferente posición indican que la posición de esta rama no es robusta.

En segundo lugar, se observa que en función del fragmento analizado, y dentro de la grupo formado por *G. gorilla*, *P. troglodytes* y *H. sapiens*, diferentes ramas muestran aceleración respecto a las otras. Con excepción de la predicción 1 (Figura R29, B) del programa *PromoterInspector*, la rama correspondiente a *P. troglodytes* aparece acelerada respecto a *H.sapiens* en el resto de árboles. Este efecto aumenta especialmente al analizar la predicción 2 (Figura R29, C), el promotor anotado de *Genomatix* (Figura R30, D) y el que consideramos como núcleo del promotor (Figura R30, E), siendo los tres fragmentos solapantes. Este efecto, podría reflejar las consecuencias de diferentes procesos evolutivos a lo largo de la secuencia.

III. Delimitación funcional del promotor

El primer exón del gen *FOXP2*, el exón s1, fue descrito por Bruce y Margolis en 2003, quienes llevaron a cabo una primera aproximación a la determinación de la región promotora adyacente a éste. Sin embargo, en la actualidad no se dispone de un promotor caracterizado, a pesar de los esfuerzos realizados hasta la fecha (Schroeder y Myers, 2008).

En el análisis evolutivo de la región promotora del gen *FOXP2* en primates, rata y ratón, se detectó una región de elevado grado de conservación que podría tener importancia a nivel funcional y formar parte de lo que propusimos como núcleo del promotor. Para confirmar nuestra hipótesis, se decidió testar esta región mediante un ensayo funcional. Este ensayo consistió en la construcción de vectores de expresión en los que la obtención de un producto cuantificable se ponía bajo el control del promotor a testar, en nuestro caso, fragmentos de longitud decreciente de la región promotora del gen *FOXP2*. A continuación se transfectaron células de mamífero con cada una de las construcciones a probar por separado, con un vector de expresión independiente cuyo producto servía para normalizar los resultados. Los niveles de producto obtenidos, en nuestro caso Fosfatasa Alcalina Secretada, se normalizaron con los de la β -galactosidasa producidos por el vector de expresión intacto. La comparación de los resultados de las diferentes construcciones respecto a un vector control carente de promotor proporciona información sobre la actividad promotora de los fragmentos analizados.

Para el ensayo funcional, tal como se especifica en el apartado II.7 de la sección de Material y Métodos, se prepararon tres construcciones a partir de la región promotora del gen *FOXP2* situada en las cercanías del exón s1. En concreto se preparó un fragmento mayor, que incluía la región conservada, denominado Prom12, y dos fragmentos de progresivamente menor tamaño, denominados Prom15 y Prom16 (figura R31).

III.1.1 Obtención de las construcciones

Las tres construcciones con fragmentos del promotor del gen *FOXP2* se obtuvieron tras la amplificación por PCR de dichos fragmentos, clonación en vector pCR[®]2.1TOPO[®], y subclonación en el vector de expresión pSEAP2-Basic. Los clones seleccionados para la transfección fueron amplificados y secuenciados con cebadores internos presentes en los insertos y en el vector pSEAP2-basic para confirmar que la secuencia compartida era la misma.

La secuencia final del fragmento de mayor tamaño, Prom12, corresponde a la que aparece como referencia en el NCBI (NT_007933.14) con la particularidad de que 3 polimorfismos tienen diferente estado (rs10539256, rs35182478 y 35017137). En este estudio no se evalúa la posible funcionalidad de ninguno de los 10 polimorfismos presentes en el fragmento existentes en las bases de datos. No obstante, esta información podría resultar útil en el caso de comparar los resultados con resultados obtenidos en otros estudios. Los estados de los polimorfismos localizados dentro de la secuencia se muestran en la tabla R43.

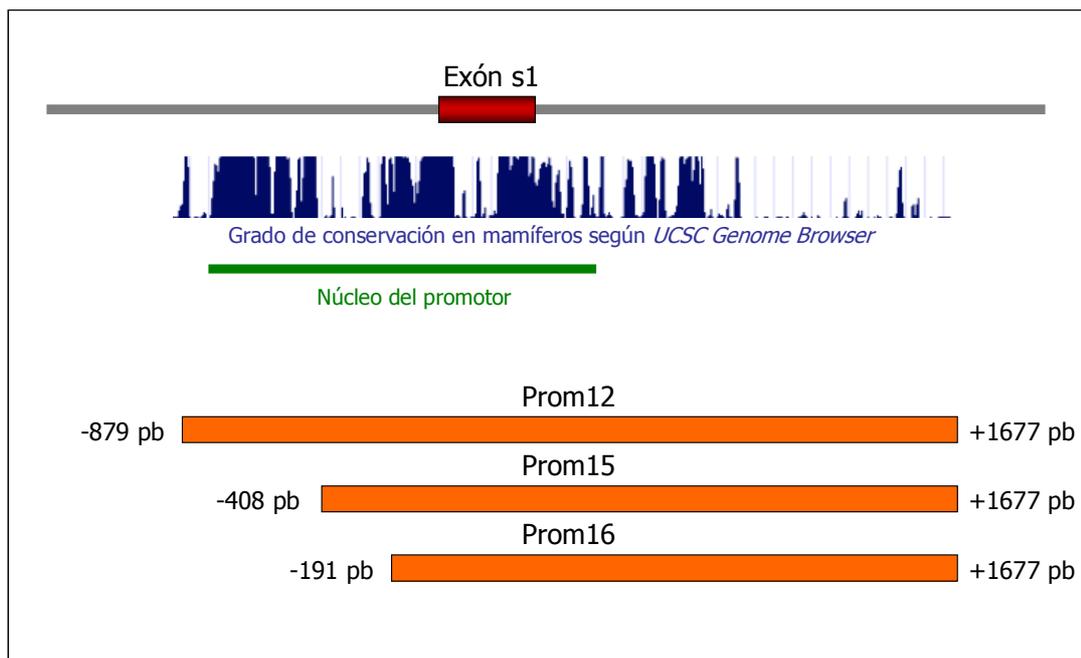


Figura R31. Esquema de las construcciones utilizadas. Se muestran las tres regiones analizadas en el contexto del exón s1, así como el diagrama de conservación en mamíferos de la región obtenido del *UCSC Genome Browser*, y el que proponemos como núcleo del promotor.

SNP	Localización respecto exón S1	<i>H. sapiens</i>	Secuencia de referencia	Prom12
rs7784307	- 456	A/C	C	C
rs10539256	-69	-/ACAC	ACAC	-
rs13309866	+ 315	A/C	C	C
rs13309912	+ 355	C/G	C	C
rs10276237	+387	G/T	G	G
rs13310148	+ 666	C/G	G	G
rs35182478	+ 1143	-/C	-	C
rs35017137	+ 1145	G/T	T	G
rs6961558	+ 1492	A/G	G	G
rs6961633	+ 1616	C/T	C	C

Tabla R43. Estado de los polimorfismos localizados en el fragmento Prom12.

Los fragmentos Prom15 y Prom16 fueron obtenidos a partir de Prom12, por lo que los estados de los polimorfismos incluidos en ambos coincidían con los del fragmento de mayor tamaño.

III.1.2 Transfección

Se hicieron dos ensayos de transfección para cada uno de los tipos celulares utilizados: células CHO-K1, procedentes de ovario de hámster y células H23, obtenidas de una cepa de células tumorales de pulmón humano. Para cada ensayo de transfección, se

Resultados

hicieron 5 experimentos diferentes, con tres réplicas cada uno. En cada uno de ellos se transfirieron los vectores y construcciones especificados en la figura R32:

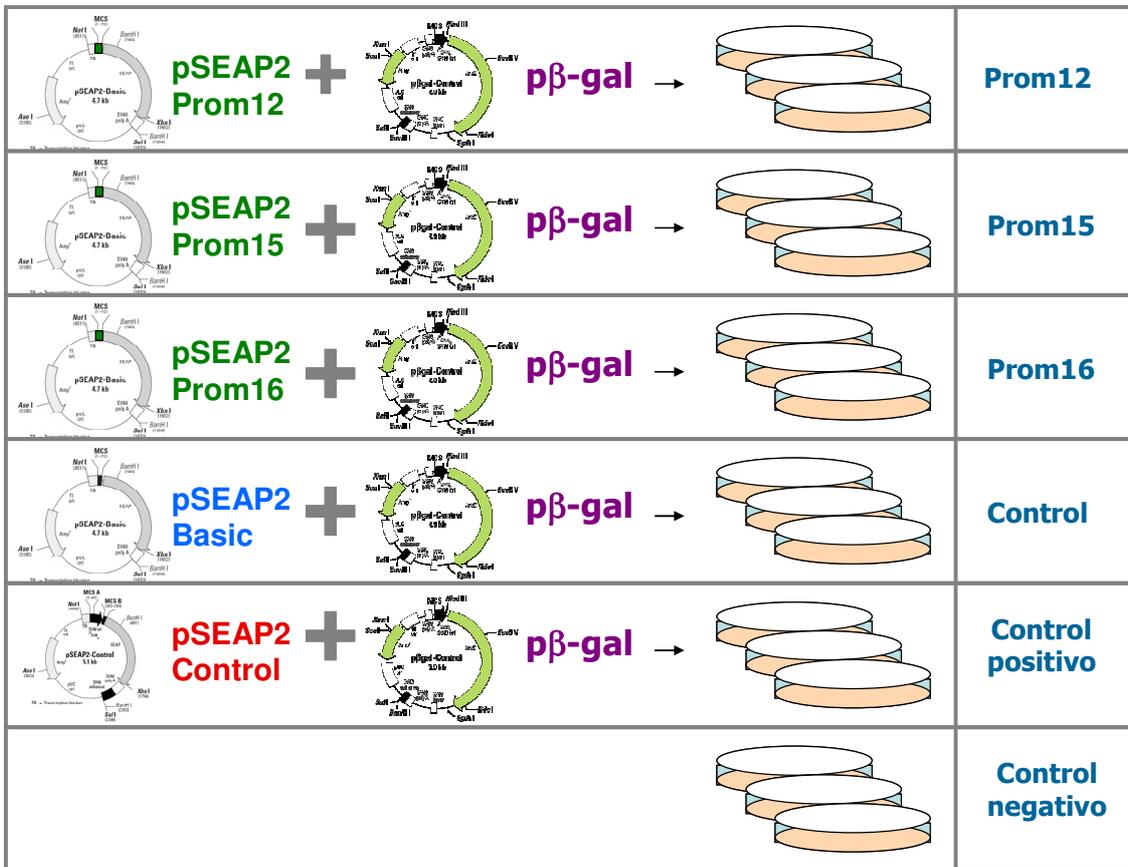


Figura R32. Esquema de los experimentos de transfección realizados.

Como se ha mencionado anteriormente, el vector pβgal fue utilizado para normalizar la actividad de la fosfatasa alcalina secretada (SEAP).

En cada ensayo se incluyó como control negativo de la transfección un experimento con células sin transfectar.

III.1.3 Medidas de la actividad de la Fosfatasa Alcalina Secretada y la β-galactosidasa

La actividad de la SEAP se midió en un luminómetro de placa (WALLAC Victor 1420 Multilabel HTS Counter) y la actividad de la β-galactosidasa en el lector de placas Fluostar OPTIMA de BMG Labtech.

Una vez transformadas las medidas de actividad a las mismas unidades, se procedió a normalizar los valores de actividad SEAP con los correspondientes a la actividad β-galactosidasa, según la fórmula:

$$\text{Actividad SEAP normalizada} = \frac{\text{Actividad SEAP normalizada}}{\text{Actividad } \beta\text{-galactosidasa}}$$

En las tablas R44 y R45 se muestran los valores obtenidos para cada uno de los experimentos de transfección en las células CHO-K1 y H23 respectivamente.

Células CHO-K1					
Ensayo 1			Ensayo 2		
	SEAP normalizada	Desviación estándar		SEAP normalizada	Desviación estándar
pSEAP2-Control	2450,82	336,67	pSEAP2-Control	4953,84	1623,68
pSEAP2-Basic	15,78	4,53	pSEAP2-Basic	8,64	4,63
Prom12	16,24	3,59	Prom12	8,89	7,09
Prom15	10,66	4,63	Prom15	4,33	3,50
Prom16	10,87	6,35	Prom16	6,73	1,42

Tabla R44. Resultados del análisis funcional para los experimentos realizados con células CHO-K1.

Se puede comprobar que el sistema de transfección utilizado funcionó a través de los resultados obtenidos del control positivo, las células trasfectadas con pSEAP2-Control, ya que se obtienen valores de actividad normalizada SEAP significativamente muy superiores a los obtenidos con el resto de construcciones. Los controles negativos para la transfección no mostraron valores de actividad para SEAP ni β -galactosidasa (datos no mostrados).

Los resultados obtenidos al transfectar células CHO-K1 muestran una variabilidad elevada tanto entre ensayos como entre las réplicas. Ésta última se ve reflejada en los elevados valores de desviación estándar (tabla R44, figura R33).

Los resultados no mostraron diferencias en la actividad SEAP normalizada de las tres construcciones. Sorprendentemente además, ninguna de las construcciones mostró mayor actividad que el control negativo para la actividad SEAP. Esto podría deberse al origen de las células CHO-K1. La secuencia de las construcciones corresponde a la secuencia promotora del gen *FOXP2* humano y las células CHO-K1, aunque constituyen un buen sistema estandarizado, son de ovario de hámster. En nuestro cultivo desconocemos si estas células expresan el gen *Foxp2*, puesto que los intentos de llevar a cabo RT-PCR a partir de RNA extraído de las mismas no tuvieron éxito. Esto fue debido no disponer de la secuencia del gen correspondiente a hámster y tener que utilizar secuencias obtenidas de ratón y rata. Respecto al tipo de tejido, en la base de datos de expresión SymAtlas (<http://symatlas.gnf.org/SymAtlas/>), se indica que el gen *FOXP2* se expresa en ovario en las especies analizadas, por lo que no es descartable la expresión del gen en nuestras células. El hecho de obtenerse menor actividad en las construcciones que en vector vacío, también podría deberse a que las secuencias reguladoras presentes son represoras de la transcripción o que en el caso de ser activadoras necesitan de otros módulos más alejados en la secuencia genómica y no presentes en nuestras construcciones.

Las diferencias en el rango de actividad SEAP normalizada observadas en los dos ensayos, probablemente se deban a diferentes factores experimentales no controlables.

Resultados

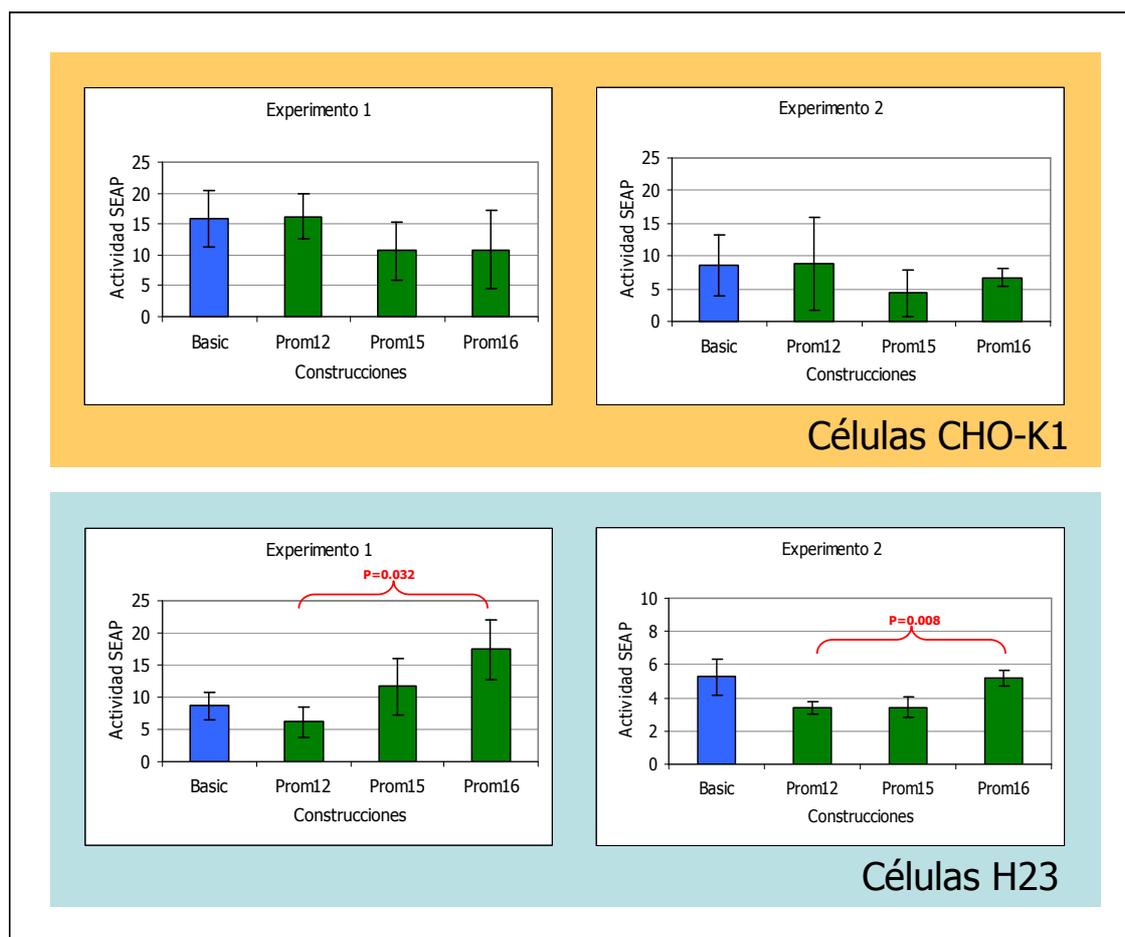


Figura R33. Representación de los resultados de análisis funcional. Se representa la actividad SEAP normalizada con actividad β -galactosidasa para el vector pSEAP2-Basic y las construcciones con insertos Prom12, Prom15 y Prom16.

Los resultados obtenidos con las células CHO-K1 nos llevaron a probar otro sistema alternativo con el que comparar los resultados o determinar si las células CHO-K1 eran las adecuadas. Las células H23 son células tumorales procedentes de pulmón, en las que se comprobó por RT-PCR la expresión del gen *FOXP2*. En este sentido, sabemos que contienen toda la maquinaria enzimática necesaria para la regulación de la expresión del gen, a diferencia de las células CHO-K1, en las que lo desconocemos.

Células H23					
Ensayo 1			Ensayo 2		
	SEAP normalizada	Desviación estándar		SEAP normalizada	Desviación estándar
pSEAP2-Control	860,89	179,18	pSEAP2-Control	868,47	132,09
pSEAP2-Basic	8,66	2,14	pSEAP2-Basic	5,24	1,09
Prom12	6,13	2,36	Prom12	3,42	0,37
Prom15	11,64	4,41	Prom15	3,43	0,63
Prom16	17,43	4,59	Prom16	5,16	0,47

Tabla R45. Resultados del análisis funcional para los experimentos realizados con células H23.

Al transfectar las células H23 también se observó variabilidad elevada tanto entre ensayos como entre las diferentes medidas de actividad para una misma construcción (tabla R45, figura R32). En el primer experimento se observa una tendencia hacia menor actividad cuanto mayor es el fragmento, lo que podría indicar que la región conservada estaría enriquecida en elementos reguladores represores de la transcripción. Sin embargo, en el segundo experimento realizado se pierde parte de la tendencia, aunque se mantiene la presencia de diferencias significativas entre la actividad normalizada de la SEAP obtenida para transfección con el fragmento Prom12 que incorpora la región conservada y el fragmento Prom16, fragmento menor de los testados, que se inicia 191 nucleótidos antes de la primera posición del exón s1 ($P=0.032$ en la Prueba T para muestras independientes en el experimento 1 y $P=0.008$ en el experimento 2).

Al comparar los dos sistemas celulares utilizados (células CHO-K1 y células H23), los resultados obtenidos en el sistema de células humanas (células H23) aunque variables, se muestran más consistentes. El resultado replicado de diferencias entre las construcciones Prom12 y Prom16 apunta a la presencia de secuencias represoras de la transcripción en la región presente en Prom12 y ausente en Prom16.

IV. Variaciones funcionales

El análisis de variaciones estructurales del gen *FOXP2*, a través del estudio de asociación y búsqueda de expansiones de trinucleótidos no proporcionó evidencias de que este tipo de variaciones tuviera un elevado peso en la vulnerabilidad a esquizofrenia, por lo que se planteó iniciar un estudio de posibles variaciones funcionales que nos permitiera obtener más información sobre su posible implicación en esta enfermedad. Se llevaron a cabo dos aproximaciones: el análisis del grado de metilación del promotor y la cuantificación de los niveles de expresión en diferentes áreas de la corteza cerebral de pacientes esquizofrénicos y controles.

IV.1 Análisis de metilación en la región promotora

Muchos genes de mamíferos contienen islas CpG en sus extremos 5', promotores o regiones 5' no traducidas cuya metilación puede asociarse a una mayor o menor expresión del gen en cuestión (Bird, 1986; Larsen *et al.*, 1992). La región promotora del gen *FOXP2* se había descrito como una región rica en GC (Bruce y Margolis, 2002), por lo que en este estudio se planteó analizar el grado de metilación de esta región para intentar determinar su posible relación con los niveles de expresión.

A través de las bases de datos *UCSC Genome Browser (Human Mar. 2006 Assembly)* y *Human Genome Segmental Duplication Database* se localizó una isla CG en la región 5' del gen *FOXP2*, solapante con el primer exón transcrito, el exón s1 (figura R34), que además coincide con la región que habíamos propuesto como núcleo del promotor. La isla CG cumple los tres requisitos necesarios establecidos para su denominación como tal: contenido en GC mayor del 50%, longitud mayor de 200 pb y razón de dinucleótidos CG mayor de 0.6 respecto a los esperados según el contenido en Cs y Gs.

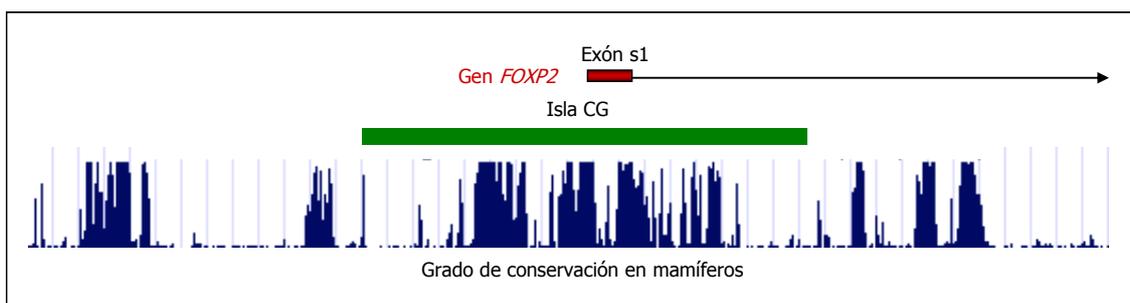


Figura R34. Contexto genómico de la Isla CG localizada en la región promotora del gen *FOXP2*.

Para analizar el estado de metilación de la isla CG se utilizaron dos estrategias:

- Patrones de digestión con enzimas de restricción sensibles a la metilación
- Modificación por bisulfito

Resultados

metilación parcial, es decir, que existían fragmentos con la diana metilada y fragmentos con diana no metilada. Puesto que la cantidad de DNA molde para la reacción de PCR era la misma, las diferencias provenían de la digestión parcial del DNA.

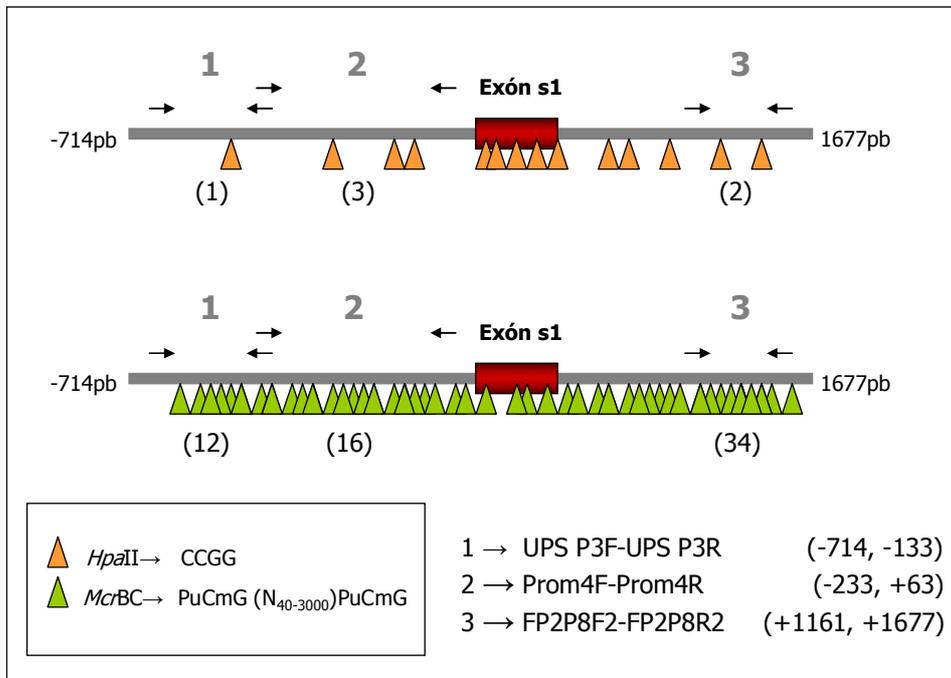


Figura R36. Esquema del análisis de metilación mediante la utilización de enzimas de restricción sensibles a metilación. 1, 2 y 3 corresponden a las regiones analizadas en el contexto del exón s1. Las dianas de restricción se representan como triángulos. Debajo de cada fragmento amplificado, y entre paréntesis se muestra el número de dianas potenciales para cada uno de los enzimas utilizados. Las distancias en pares de bases respecto al exón s1 se muestran junto a la descripción de cada fragmento.

Para el enzima de restricción *HpaII* se observó que todas las dianas analizadas se encontraban no metiladas, sin obtenerse diferencias entre pacientes con esquizofrenia y controles, ni entre áreas de la corteza cerebral. Esto quiere decir que no se obtuvo producto amplificado después de la digestión con el enzima de restricción, frente a la amplificación del DNA sin digerir.

Área de la corteza cerebral	Individuos	1		2		3	
		<i>HpaII</i>	<i>McrBC</i>	<i>HpaII</i>	<i>McrBC</i>	<i>HpaII</i>	<i>McrBC</i>
Circunvolución temporal izquierda	Controles	NM	PM	NM	M	NM	PM
	Pacientes	NM	PM	NM	M	NM	PM
Circunvolución temporal derecha	Controles	NM	PM	NM	M	NM	PM
	Pacientes	NM	PM	NM	M	NM	PM
Circunvolución del cíngulo izquierda	Controles	NM	NM	NM	NM	NM	M
	Pacientes	NM	NM	NM	NM	NM	M
Circunvolución del cíngulo derecha	Controles	NM	NM	NM	NM	NM	M
	Pacientes	NM	NM	NM	NM	NM	M

Tabla R46. Resultados del análisis de metilación mediante la utilización de los enzimas sensibles a metilación *HpaII* y *McrBC*. M indica diana de restricción metilada, NM, diana no metilada y PM diana parcialmente metilada.

En el caso del enzima *McrBC* tampoco se obtuvieron diferencias entre pacientes y controles, sin embargo sí se observaron evidencias de diferente metilación entre las dos áreas de la corteza cerebral analizadas, encontrándose evidencias de mayor metilación en la circunvolución temporal superior respecto a la del cíngulo. También se observaron diferencias entre los tres fragmentos de DNA analizados, con evidencias de metilación en el fragmento localizado en intrón s1, en la circunvolución del cíngulo, frente a ausencia de dianas metiladas en los dos fragmentos localizados aguas arriba del primer exón.

Al analizar los resultados correspondientes a este apartado es importante tener en cuenta que la utilización de enzimas de restricción sensibles a metilación proporciona una idea general del grado de metilación sin ser muy sensible. Los resultados obtenidos nos muestran que no hay diferencias en el grado de metilación para las 6 dianas analizadas del enzima *HpaII* en las dos áreas cerebrales estudiadas: circunvolución temporal superior y circunvolución del cíngulo. En cambio, sí parece haberlas para el caso de las dianas del enzima *McrBC* con un mayor grado de metilación en la circunvolución temporal superior. En este caso, sin embargo, el elevado número de dianas analizadas (12, 16 y 34 dianas en los tres fragmentos) dificulta estimar cuantitativamente las diferencias puesto que sólo con que una de las dianas se encontrara metilada, la reacción de PCR no podría tener lugar. No se puede, por tanto, diferenciar entre una o más dianas metiladas.

IV.1.2 Análisis de metilación mediante la técnica de bisulfito

Como se menciona en el apartado de Material y Métodos (epígrafe II.8.2), la técnica del bisulfito constituye un método sensible para analizar el estado de metilación de un DNA, ya que permite mostrar las 5' metilcitosinas en cadenas individuales. Tras la modificación del DNA genómico con bisulfito, los residuos de citosina no metilados se convierten en uracilo que será amplificado en la posterior reacción de PCR como timina, mientras que los residuos de citosina metilados en dinucleótidos CG, se mantendrán como citosinas (esquema en la figura M5 de Material y Métodos). Los productos de PCR obtenidos se clonan y se secuencian un mínimo de 4 clones, de forma que a partir de la secuencia se puede determinar la metilación de las citosinas en el DNA original.

Para llevar a cabo este análisis se partió de DNA genómico obtenido a partir de muestras individuales de cerebro *postmortem* humano de diferentes regiones de la corteza cerebral (circunvolución temporal superior, circunvolución del cíngulo y circunvolución del parahipocampo), procedentes de pacientes esquizofrénicos e individuos control, así como DNA genómico procedente de leucocitos humanos y DNA de una línea celular de gorila. Estas muestras habían sido donadas al centro *SAVE POWIC (SANE Prince of Wales Internacional Centre)* de Oxford. En la tabla M41 se muestran los DNAs utilizados en este análisis.

En la figura R37 se muestra un esquema en el que aparecen representados mediante líneas verticales todos los dinucleótidos CG potencialmente metilables de la región adyacente al exón s1, así como la localización de la isla CG y los fragmentos analizados mediante la técnica del bisulfito en este estudio. Se analizaron dos fragmentos. El primero de ellos, denominado *Región bisulfito CG1* o abreviadamente, *CG1*, está localizado aguas arriba del primer exón, comprende las posiciones -430 hasta -192

Resultados

respecto a la primera base del exón s1 y contiene 14 dinucleótidos CG. El segundo fragmento, denominado *Región bisulfito CG2*, o *CG2*, está localizado en el intrón s1, comprende las posiciones +1230 hasta +1455 respecto a la primera posición del exón s1 y contiene 41 dinucleótidos CG (figura R37).

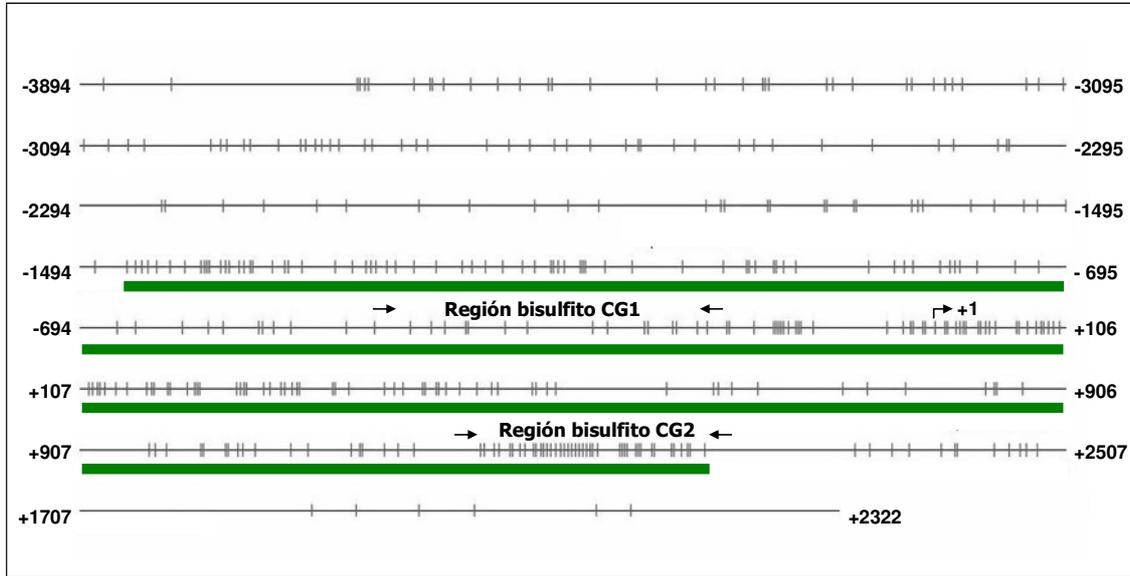


Figura R37. Representación del mapa de dinucleótidos CG potencialmente metilables en el contexto de la isla CG localizada en la región promotora del gen *FOXP2*. La localización de cada dinucleótido CG aparece representada como una línea vertical. En color verde se muestra la región que abarca la isla CG. Las dos regiones analizadas mediante la técnica del bisulfito se encuentran acotadas por flechas.

Tras la modificación con bisulfito, amplificación por PCR, clonación y secuenciación de los clones seleccionados para cada región cerebral analizada, se estimó el estado de metilación de cada uno de los dinucleótidos CG comprendidos en los fragmentos, en función de la conversión a uracilo (reflejado como timina en la secuencia) o conservación de la citosina. En las figuras R38-R41 se muestran los resultados. Las posiciones de los dinucleótidos CG se identifican con círculos, negros, si estaban metilados en la secuencia original y por tanto las Cs no habían sido modificadas por el bisulfito y blancos, si no estaban metilados y por tanto las Cs fueron transformadas.

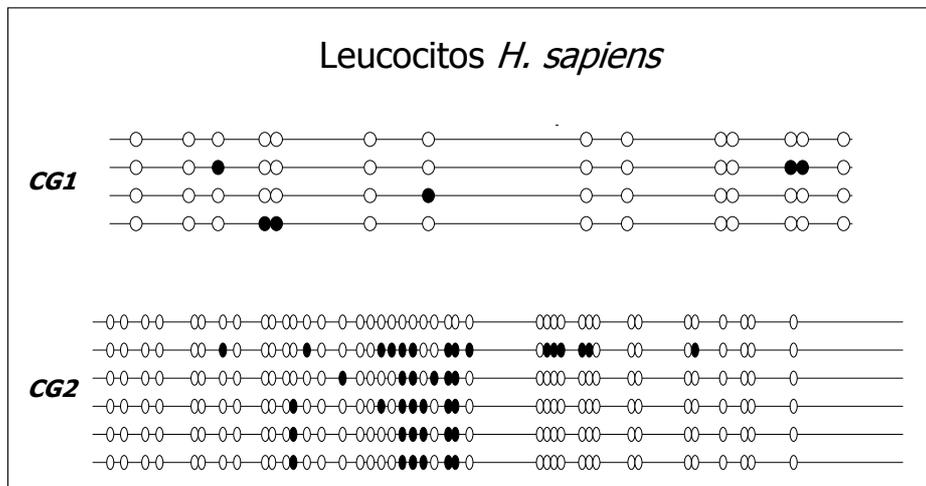


Figura R38. Resultados del análisis de metilación mediante bisulfito de los clones obtenidos a partir de DNA convertido procedente de leucocitos humanos.

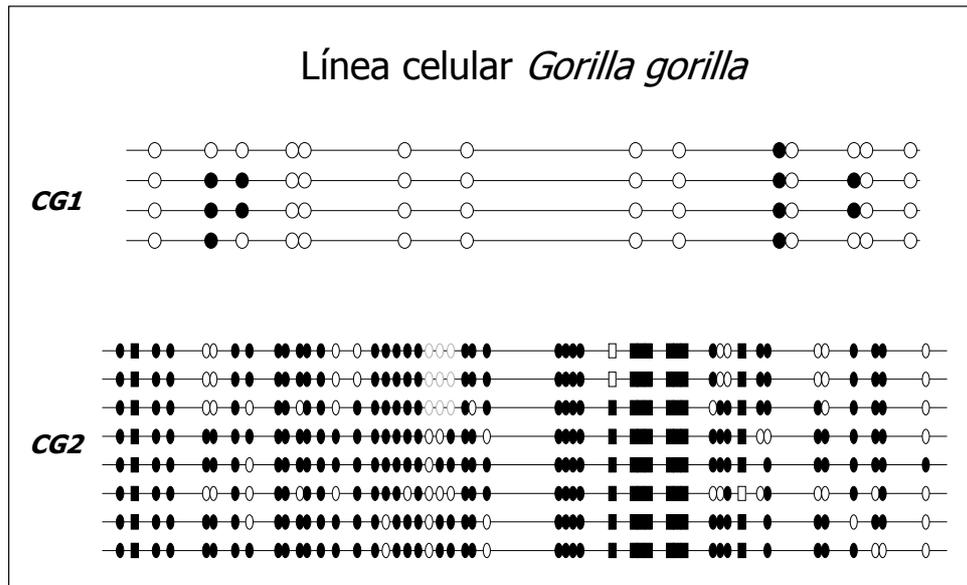


Figura R39. Resultados del análisis de metilación mediante bisulfito de los clones obtenidos a partir de DNA convertido procedente de una línea celular de gorila. Los dinucleótidos específicos de *G. gorilla* aparecen representados como rectángulos en su correspondiente posición. Las posiciones para las cuales el estado de metilación no se pudo determinar muestran el contorno de los óvalos en gris.

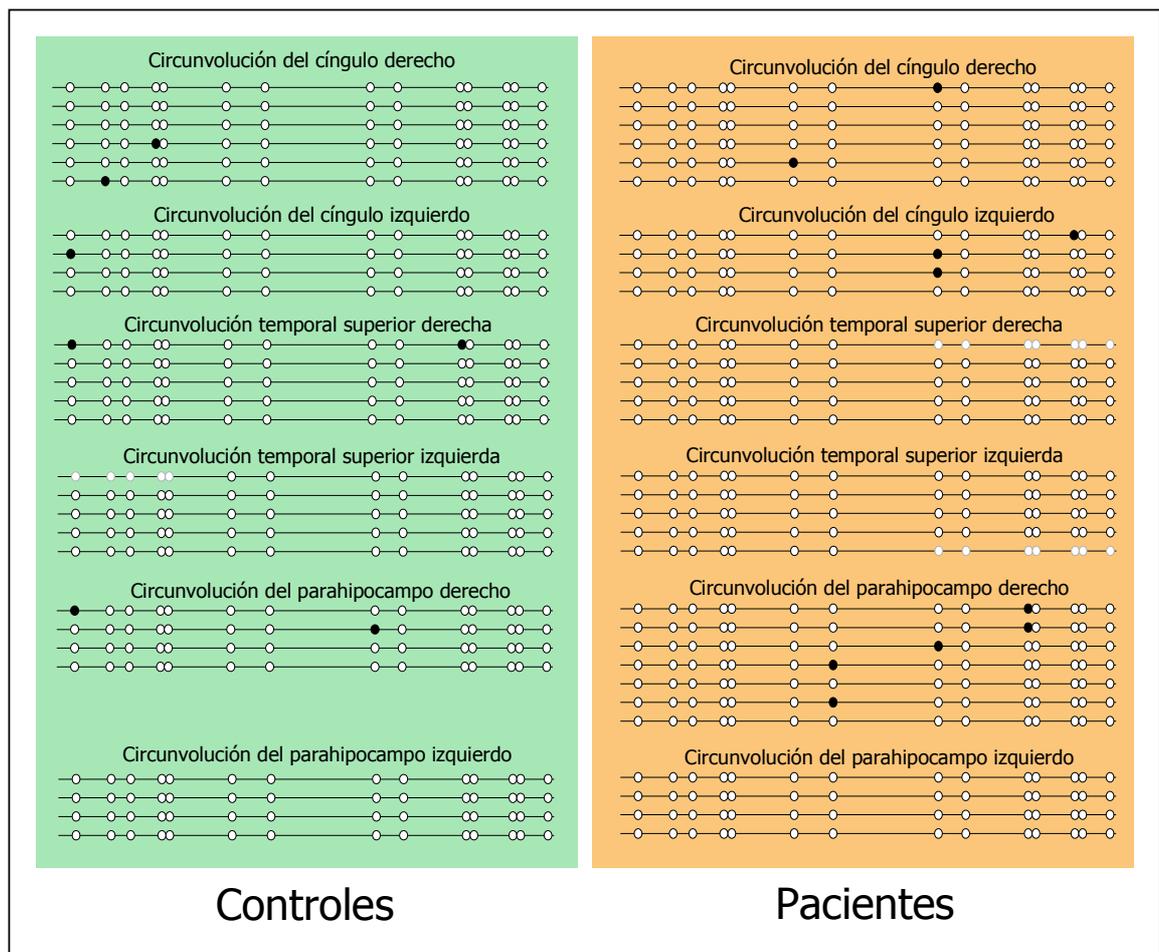


Figura R40. Resultados del análisis de metilación mediante la utilización de bisulfito de todos los clones de la región *CG1* obtenidos para cada región cerebral estudiada. Las posiciones para las cuales el estado de metilación no se pudo determinar muestran el contorno de los círculos en gris.

Resultados

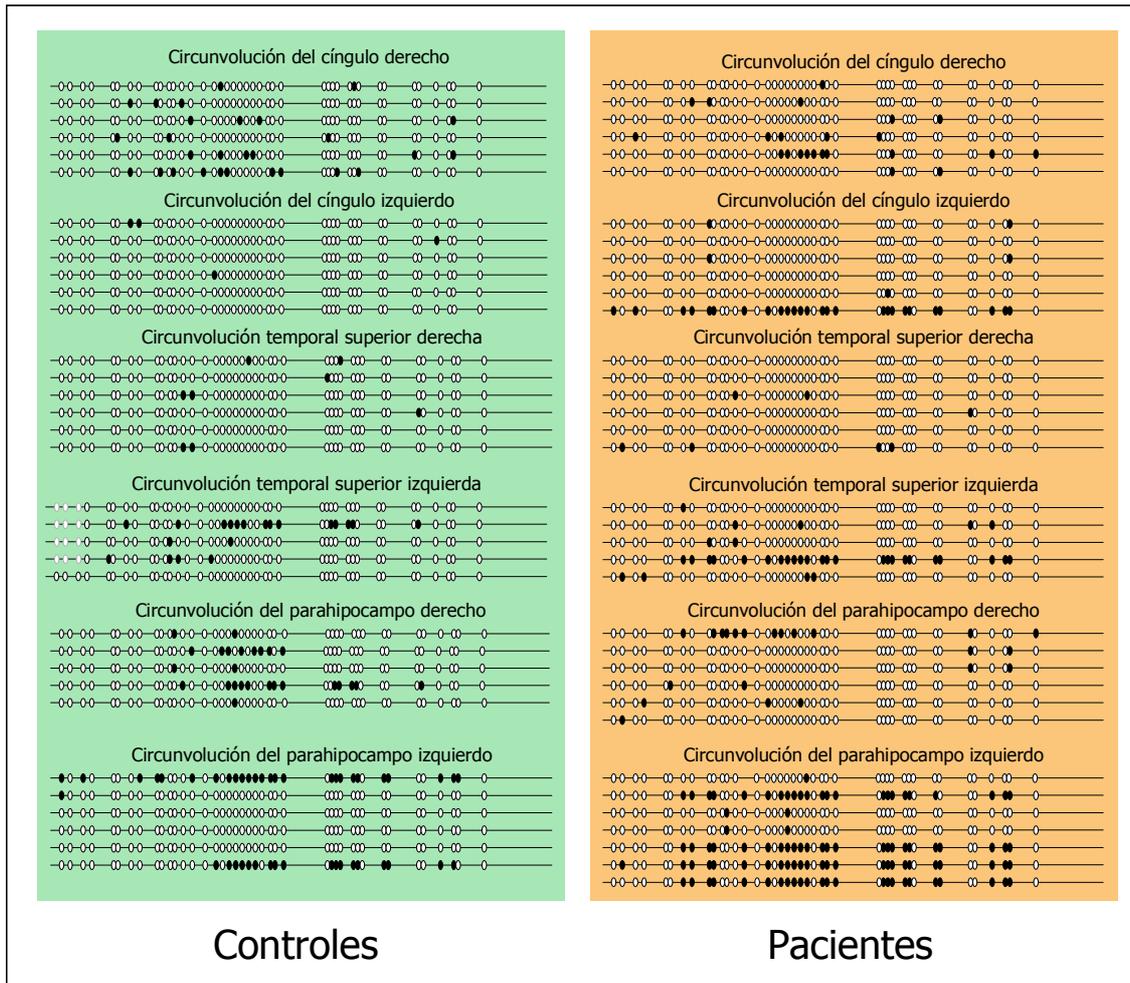


Figura R41 Resultados del análisis de metilación mediante la utilización de bisulfito de todos los clones de la región CG2 obtenidos para cada región cerebral estudiada. Las posiciones para las cuales el estado de metilación no se pudo determinar muestran el contorno de los óvalos en gris.

En las figuras R38 y R39, se muestran los resultados de los clones obtenidos para las muestras correspondientes a una línea celular de gorila y leucocitos humanos. La muestra correspondiente a la línea celular de gorila nos sirvió como control positivo de metilación.

En las figuras R40 y R41 se muestran los resultados correspondientes a controles y pacientes en cada una de las regiones cerebrales estudiadas.

Una vez obtenidos los clones de cada región analizada, se elaboró un patrón consenso de cada una de ellas, en función del tanto por cien de clones que mostraban una determinada posición CG como metilada. En la figura R42. se muestra un ejemplo de cómo se llevó a cabo este paso.

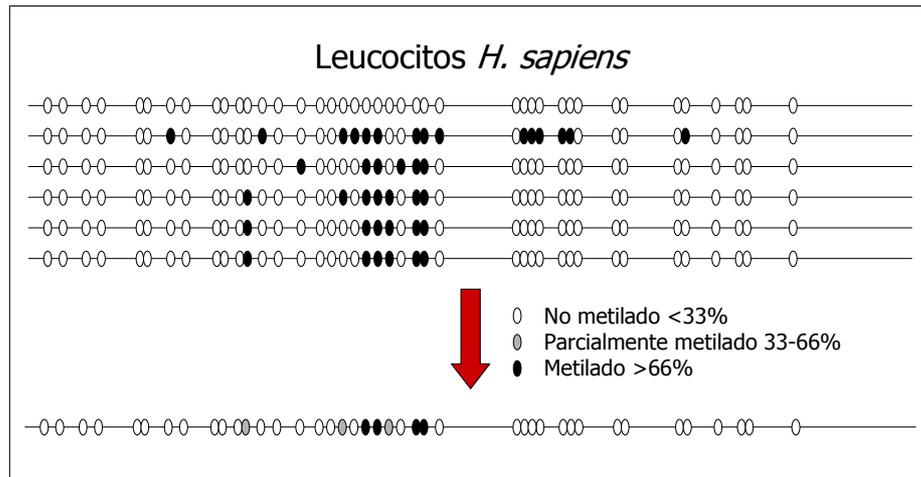


Figura R42. Creación de un patrón consenso para cada una de las regiones analizadas.

El patrón consenso permite determinar la presencia en un determinado tejido de patrones de metilación comunes, ya que estos se verían reflejados en el resultado final, frente a la metilación de dinucleótidos individuales en posiciones aleatorias. Así, después de crear el patrón para cada una de las regiones en cada área cerebral analizada, los resultados obtenidos se muestran en las figuras R43 y R44.

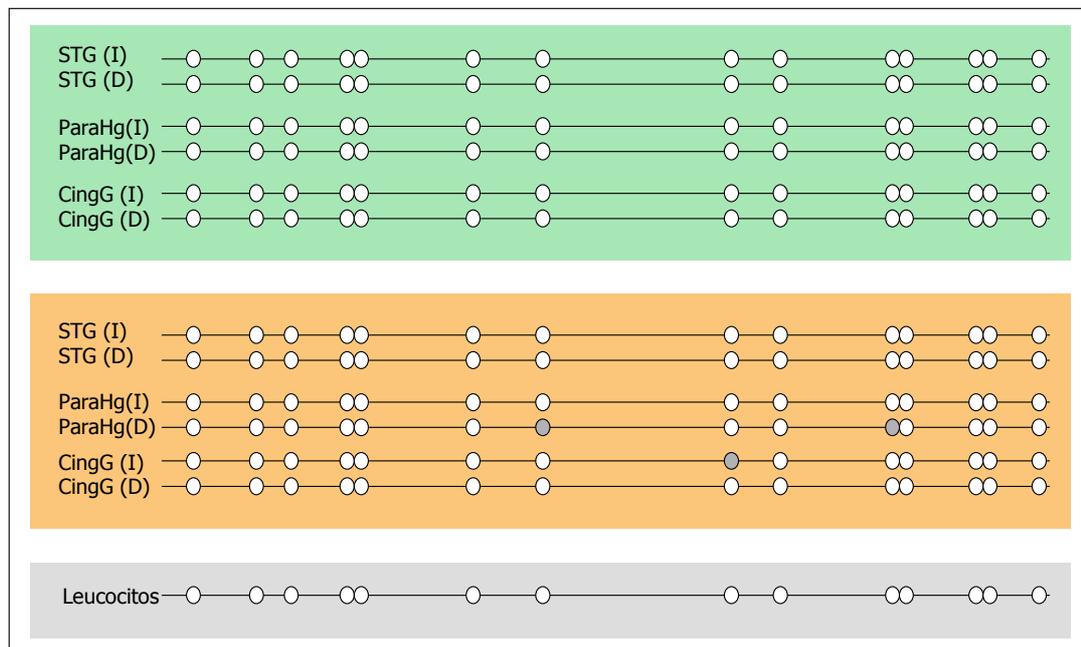


Figura R43. Patrones consenso de la región bisulfito *CG1* para cada una de las áreas cerebrales estudiadas, en pacientes (sombreado en naranja), y controles (sombreado en verde), así como para leucocitos (sombreado en gris).

Respecto a la región denominada *Región bisulfito CG1*, localizada aguas arriba del inicio del primer exón del gen, el exón s1, no se observan diferencias apreciables en cuanto al estado de metilación entre pacientes y controles. Se encontraron posiciones metiladas puntualmente, pero el patrón general es de ausencia de metilación en ambos grupos de muestras.

en cuenta que no se expresa en todos los tipos celulares presentes, por lo que la existencia de diferentes patrones es coherente.

Los resultados obtenidos mediante la técnica de modificación del DNA con bisulfito son consistentes con los obtenidos mediante enzimas de restricción sensibles a metilación, puesto que en ambos casos se observó una mayor presencia de metilación en las regiones localizadas en el intrón s1 frente a las localizadas aguas arriba del primer exón del gen.

IV.2 Estudio de los niveles de mRNA en diferentes áreas cerebrales en pacientes con esquizofrenia y controles

Dentro del análisis de variaciones funcionales se llevó a cabo un estudio de los niveles de mRNA en diferentes áreas de la corteza cerebral. Con este estudio se pretendía, en primer lugar, determinar si los niveles de expresión del gen *FOXP2* eran diferentes entre pacientes con esquizofrenia y controles, así como entre diferentes áreas de la corteza cerebral y en segundo lugar, relacionar los resultados preliminares obtenidos del análisis de metilación con datos de expresión. La presencia de metilación en islas CG localizadas en promotores suele encontrarse asociada a una disminución de la expresión (revisado en Costello y Plass, 2001; Feinberg, 2007) por lo que lo que se esperaba encontrar es que en aquellas áreas que mostraron indicios de mayor metilación la expresión fuera menor.

Las muestras utilizadas para el análisis de expresión consistieron en cDNA obtenido a partir de 1 µg de RNA extraído a partir de tejido cerebral *postmortem*. Las muestras de tejido, de diferentes regiones de la corteza cerebral procedían de pacientes esquizofrénicos e individuos control y habían sido donadas al centro *SANE POWIC* (*SANE Prince of Wales Internacional Centre*) de Oxford. En la tabla M32 se muestran los DNAs disponibles para este análisis.

El estudio de los niveles de expresión del gen *FOXP2* se llevó a cabo en sucesivos experimentos de PCR cuantitativa en tiempo real. En todos ellos se normalizó el resultado obtenido para el gen con los resultados obtenidos para el gen de expresión constitutiva *RPII*. Además, puesto que la cantidad normalizada obtenida es un número sin unidades, para comparar las cantidades relativas en diferentes muestras, se designó en cada experimento una de ellas como calibrador.

Las muestras utilizadas varían entre experimentos, por lo que en cada uno de ellos se especificará en detalle las que se utilizaron.

El **primer experimento** consistió en la comparación general de la cantidad relativa de cDNA del gen *FOXP2* en una serie de mezclas compuestas por diferentes muestras, agrupadas, bien en función de las diferentes **áreas cerebrales** (Tabla R47), bien en función de la pertenencia al grupo de **controles o pacientes** (Tabla R48), o bien en función del **sexo** (Tabla R48).

Las muestras estudiadas fueron las siguientes:

Resultados

Área de la corteza cerebral	Número de muestras en cada mezcla	
	Hemisferio Izquierdo	Hemisferio Derecho
Circunvolución del cíngulo (CingGy)	18	15
Paracingulado (Paracing)	19	17
Circunvolución Temporal Superior (STG)	22	15
Circunvolución del Parahipocampo (ParaHg)	17	11

Tabla R47. Características de las mezclas de muestras agrupadas en función del área de la corteza cerebral y hemisferio utilizadas en el experimento de PCR cuantitativa.

Paciente o control	Sexo	Hemisferio	Número de muestras en cada mezcla
Controles	Mujer	Izquierdo	4
		Izquierdo	4
		Izquierdo	4
		Derecho	4
		Derecho	3
		Derecho	4
	Hombre	Derecho	4
		Izquierdo	4
		Izquierdo	4
		Derecho	5
Derecho		4	
Pacientes	Mujer	Ambos	8
		Izquierdo	4
		Ambos	8
		Ambos	7
	Hombre	Izquierdo	4
		Izquierdo	4
		Ambos	9
		Ambos	8
		Ambos	6

Tabla R48. Características de las mezclas de muestras agrupadas en función de la pertenencia al grupo de controles o grupo de pacientes, del sexo o del hemisferio utilizadas en el experimento de PCR cuantitativa.

Los resultados obtenidos para el gen *FOXP2*, una vez normalizados con los correspondientes al gen *RPII* se muestran en las figuras R45 y R46.

En la comparación de la cantidad relativa del gen *FOXP2* en función de las diferentes áreas de la corteza cerebral analizadas, incluyendo en las mezclas tanto pacientes como controles, no se observan diferencias apreciables entre hemisferios de una misma área estudiada. En cuanto a la consideración de las áreas los resultados apuntarían a una tendencia hacia menor expresión del paracingulado respecto a otras áreas, aunque dados los valores de desviación estándar, no se observa claramente.

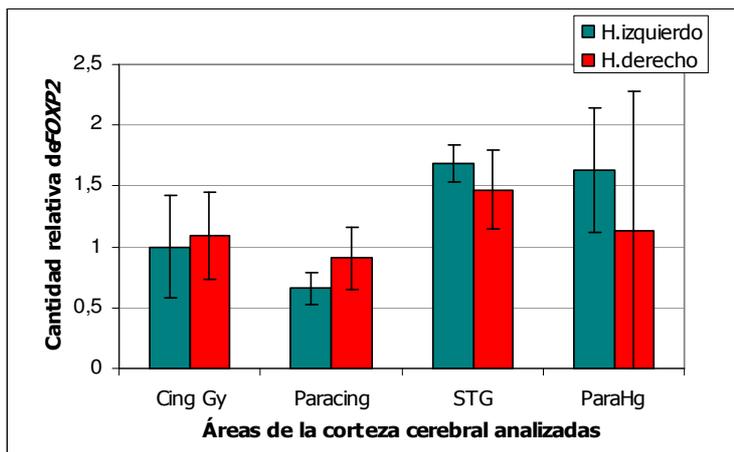


Figura R45. Resultados normalizados utilizando como calibrador la muestra correspondiente a la circunvolución del cíngulo, CingGy, del hemisferio izquierdo de la PCR cuantitativa para las muestras agrupadas correspondientes a las diferentes regiones cerebrales.

Al comparar los resultados obtenidos para pacientes frente a controles, o en función del sexo, se observa una clara ausencia de diferencias globales entre ellos, sin ningún tipo de tendencia observable en alguna dirección.

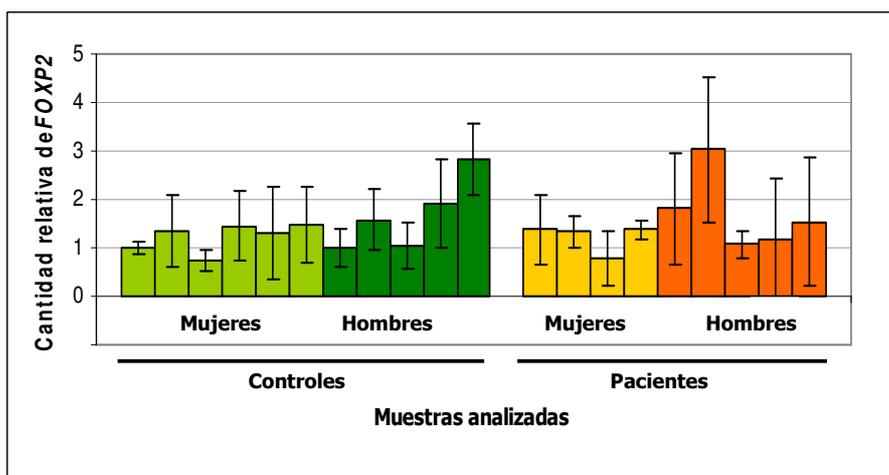


Figura R46. Resultados normalizados de la PCR cuantitativa utilizando como calibrador la primera muestra de controles del sexo femenino para las muestras agrupadas correspondientes a las diferentes regiones cerebrales.

En un **segundo experimento** tratamos de determinar la posible existencia de **diferencias de expresión entre pacientes y controles** en cada una de las **áreas de la corteza cerebral** analizadas, para lo cual se llevaron a dos aproximaciones.

En la primera se analizaron las áreas cerebrales de pacientes y controles varones indicadas en la tabla R49. El experimento se llevó a cabo en placas separadas para controles y pacientes. Las muestras utilizadas son las siguientes:

Resultados

Región	Número de muestras			
	Controles		Pacientes	
	Hemisferio Izquierdo	Hemisferio Derecho	Hemisferio Izquierdo	Hemisferio Derecho
Cing Gy	6	4	4	3
STG	6	3	4	3
ParaHg	5	3	4	-

Tabla R49. Muestras de controles y pacientes de sexo masculino utilizadas para el análisis en el experimento de comparación de la expresión entre pacientes y controles.

Los resultados obtenidos se muestran a continuación:

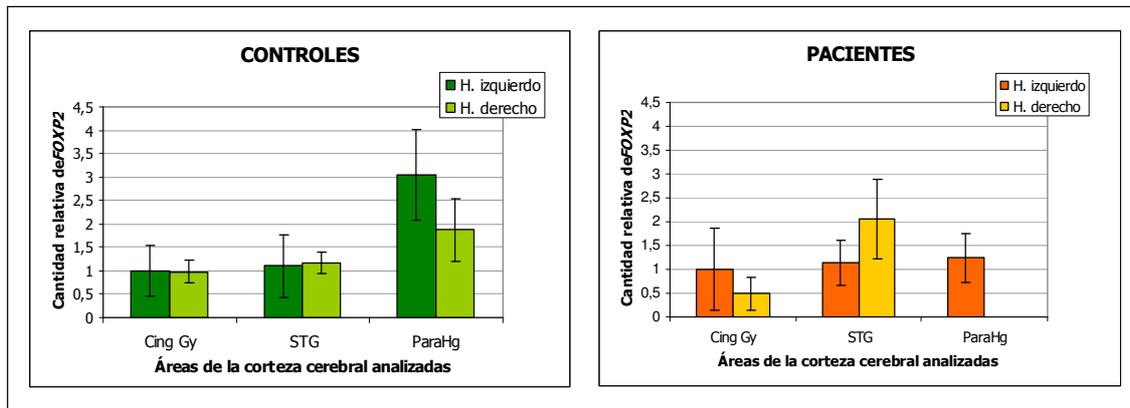


Figura R47. Resultados de la PCR cuantitativa normalizados utilizando como calibrador la muestra correspondiente al Cing Gy del hemisferio izquierdo, para las muestras agrupadas correspondientes a las diferentes regiones cerebrales. Los resultados correspondientes a individuos varones control aparecen en color verde en la gráfica de la izquierda, mientras que los individuos varones pacientes se muestran en naranja en la gráfica de la derecha.

Los resultados apuntan hacia una mayor expresión del gen *FOXP2* en la circunvolución del parahipocampo (ParaHg) izquierdo, respecto a la circunvolución temporal superior (STG) y la circunvolución del cíngulo (CingGy) en controles respecto a pacientes, aunque el tamaño de las barras de error indica que los resultados deben tomarse con precaución.

En la segunda aproximación, se volvieron a analizar las áreas que mostraban indicios de expresión diferencial del gen, esto es la circunvolución temporal superior y la circunvolución del parahipocampo. Para minimizar la posible variabilidad tanto entre experimentos como entre réplicas de una misma muestra, se analizaron todas las muestras en una misma placa de reacción, se llevaron a cabo tres réplicas de cada muestra y en el análisis de resultados se eliminó el valor obtenido más alejado para cada muestra.

Las muestras utilizadas se muestran en la tabla R50 y los resultados obtenidos en la figura R48.

En este caso se obtiene ausencia de diferencias en el caso del área de la circunvolución temporal superior entre pacientes y controles, así una tendencia hacia mayor expresión en la circunvolución del parahipocampo en pacientes respecto a controles. Las diferencias respecto al estudio anterior podrían ser debidas en parte a la utilización de muestras procedentes de individuos de ambos sexos en este caso.

Área cerebral	Hemisferio	Individuos	Número de muestras
Circunvolución temporal superior (STG)	Izquierdo	Controles	9
		Pacientes	10
	Derecho	Controles	8
		Pacientes	7
Circunvolución del Parahipocampo (ParaHg)	Izquierdo	Controles	6
		Pacientes	9
	Derecho	Controles	7
		Pacientes	3

Tabla R50. Muestras de individuos pacientes y controles de ambos sexos utilizadas en el experimento para determinar diferencias entre ambos grupos en las áreas de la corteza cerebral analizadas.

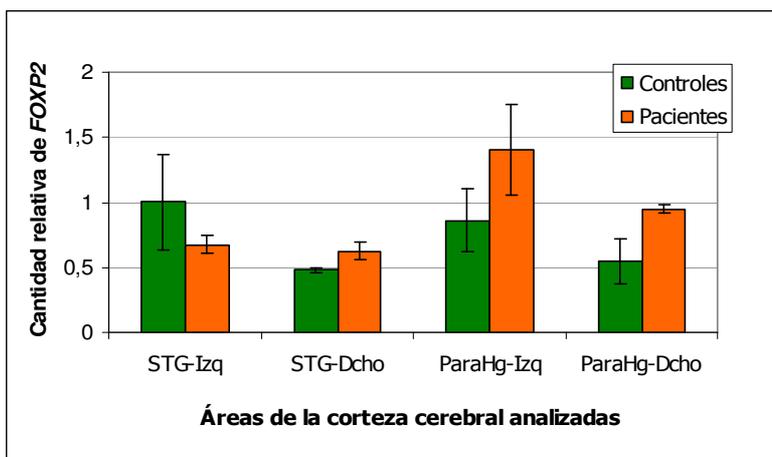


Figura R48. Resultados normalizados de la PCR cuantitativa utilizando como calibrador la muestra correspondiente al STG del hemisferio izquierdo, para la comparación de niveles de expresión de *FOXP2* en pacientes y controles.

En los experimentos anteriores se trabajó en todo momento con mezclas de individuos. En un último experimento nos planteamos analizar el grado de **variación individual** en la expresión del gen *FOXP2* en un área determinada. Concretamente, tomamos como área de estudio la circunvolución del parahipocampo izquierdo. Las muestras utilizadas en este experimento se muestran en la tabla R51.

Controles		Pacientes	
Muestra	Sexo	Muestra	Sexo
55	M	4	M
88	M	5	F
93	F	28	M
118	F	37	M
122	M	38	F
130	F	41	F
		76	F
		80	M
		111	F

Tabla R51. Muestras utilizadas en el experimento de PCR cuantitativa destinado a evaluar la variabilidad entre muestras individuales.

Resultados

Los resultados obtenidos para la cuantificación de niveles de mRNA de *FOXP2* en muestras individuales correspondientes al hemisferio izquierdo de la circunvolución del parahipocampo se muestran en la figura R49.

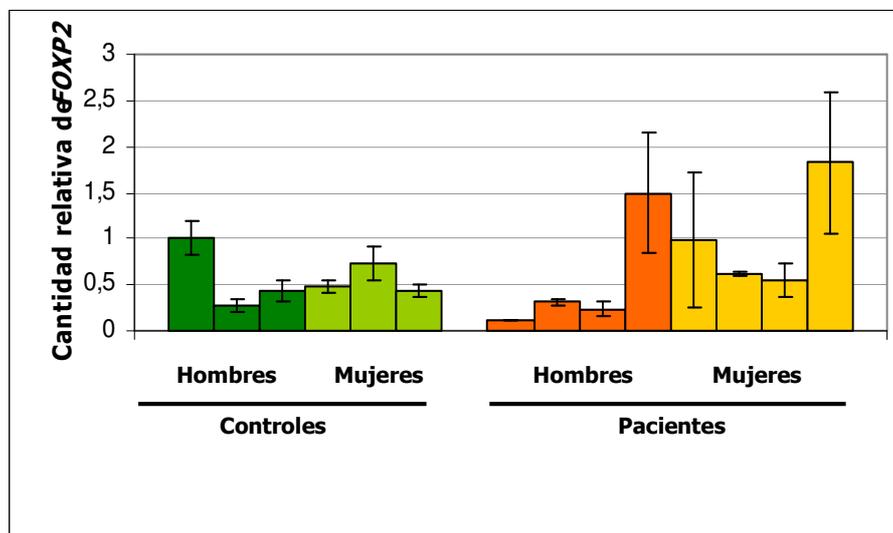


Figura R49. Resultados de la reacción de PCR cuantitativa normalizados utilizando como calibrador la primera muestra analizada, para las muestras individuales correspondientes a la región de la circunvolución del parahipocampo izquierdo.

Se observó una elevada variabilidad en la cantidad relativa de *FOXP2* en las diferentes muestras analizadas, tanto en controles como especialmente en pacientes, donde se observan algunas muestras con más del doble que otras. Si se obtienen valores medios para pacientes y controles, se obtiene una representación semejante a la observada para la región del hipocampo izquierdo en el experimento anterior (figura R49). Sin embargo, este efecto no refleja la variabilidad observada de las muestras.

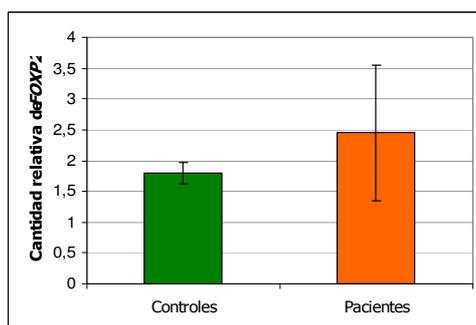


Figura R50. Resultados promedio de la reacción de PCR cuantitativa para muestras de la región de la circunvolución del parahipocampo izquierdo.

La elevada variación en los resultados obtenidos al analizar las muestras independientes indica que las posibles diferencias observadas en anteriores experimentos en los que se utilizaban mezclas podrían ser debidas al efecto de unas pocas muestras en la mezcla. Esto refleja una de las desventajas de utilizar mezclas en

este tipo de estudios, aunque esta utilización dependa de la limitación del material de partida. En este sentido, la utilización de una muestra más amplia permitiría reducir el efecto que el tamaño de la muestra puede ocasionar.

No se han obtenido por tanto unos resultados lo suficientemente robustos como para poder correlacionar datos de expresión con datos de metilación.

Resultados

Discusión



I. Variaciones estructurales de genes sometidos a selección positiva y esquizofrenia

I.1 Estudio de asociación con variantes de los genes *FOXP2* y *HAR1A*

Durante los últimos años, los estudios de asociación han protagonizado una de las etapas más importantes en el estudio de las bases genéticas de las enfermedades psiquiátricas. La imposibilidad de identificar mediante análisis de ligamiento, *loci* replicados de forma inequívoca en múltiples muestras, impulsó el desarrollo de estos estudios. Inicialmente, los estudios de asociación estaban basados en genes candidatos, bien en función de su localización genómica, bien por su posible participación en rutas funcionales relacionadas con la enfermedad. Dentro del contexto de la esquizofrenia se han analizado más de 500 genes (Allen *et al.*, 2008).

Sin embargo, los estudios de asociación de genes candidatos con la esquizofrenia tampoco han dado resultados consistentes. Esto ha propiciado la búsqueda de vías alternativas de estudio. Desde el punto de vista clínico se busca una mejor caracterización del síndrome, o la utilización de endofenotipos o fenotipos alternativos con los que llevar a cabo los estudios genéticos. Desde la perspectiva del análisis genético, en primer lugar se ha producido la evolución de los estudios de asociación hacia los conocidos como GWAs (*Genome Wide Association Studies*), basados en el análisis de variantes a lo largo de todo el genoma. En segundo lugar, han surgido nuevas aproximaciones al estudio genético de la enfermedad, bien a través de los recientes estudios sobre CNVs (*Copy Number Variations*), estudios de expresión o estudios epigenéticos.

En esta tesis doctoral se han abordado dos de estas aproximaciones a través de diferentes métodos. En primer lugar se ha utilizado un fenotipo alternativo a la esquizofrenia para llevar a cabo un estudio de asociación caso-control con genes candidatos. En este sentido, en nuestro estudio se ha decidido abordar una aproximación más simple a la de los estudios GWAs, donde no hace falta tener una hipótesis previa sobre qué genes pueden conferir susceptibilidad a una determinada enfermedad, y se consideraron unos genes candidatos específicos. Esta aproximación permite el análisis más detallado de las regiones genómicas donde se encuentran los genes estudiados aunque tiene la desventaja de que toda la variabilidad analizada se ve limitada a estos *loci*. En segundo lugar, se llevó a cabo un estudio epigenético destinado a identificar diferencias de metilación en el promotor de un gen candidato, entre pacientes con esquizofrenia y controles.

En el presente trabajo se han tomado las alucinaciones auditivas, como fenotipo alternativo al síndrome psicótico general. La utilización de las AA, que constituye uno de los principales síntomas de la esquizofrenia, se basa en la hipótesis de que las bases genéticas de las mismas deben ser menos complejas que las de la enfermedad global, o al menos es un fenotipo mejor definido que la esquizofrenia. En nuestro caso, la utilización de las AA ha demostrado resultar más robusta que la del síndrome general como se explicara más adelante.

Discusión

En el apartado de la Introducción ya se describieron diferentes modelos de origen de las AA (ver epígrafe II.2.3). Recordamos aquí que la presencia de alucinaciones también tiene lugar en otros desórdenes psiquiátricos, e incluso en la población normal (Brasic, 1998; David, 1999; Johns, 2005). Entonces, ¿tienen lugar los mismos mecanismos de presencia de AA en individuos normales e individuos psicóticos? ¿Es la forma en la que reaccionan a sus experiencias lo que ocasiona que se conviertan en pacientes psiquiátricos? (van Os *et al.*, 2000; Bentall, 2003). Según el modelo de aparición de las AA de nuestro grupo de investigación, debe existir una vulnerabilidad genética que acompañada de determinados factores ambientales del desarrollo, da lugar a una tendencia a experimentar como ajeno determinados estímulos, en definitiva, a tener alucinaciones. Dentro del contexto de la psicosis, donde el tipo de alucinación más frecuente son las AA, la vulnerabilidad genética se desglosa en: aquella que ocasiona la experiencia de voces, en la que intervendrían genes relacionados con el lenguaje, y aquella que da lugar a un estado emocional anómalo.

Así, utilizando las AA como fenotipo central, se ha llevado a cabo un estudio de asociación caso control en el que se compararon frecuencias genotípicas, alélicas y haplotípicas entre controles y pacientes psicóticos, entre los que se encontraban pacientes con y sin AA, así como con diferentes variables clínicas.

Genes candidatos en este estudio

Es un hecho ampliamente admitido desde los primeros estudios de esta enfermedad, que la esquizofrenia es una patología específicamente humana (Crichton-Browne, 1879; Conrad, 1957, en la versión española, 1997). Esta importante característica de la esquizofrenia se basa en la constatación clínica de que los aspectos nucleares de la enfermedad tienen que ver con rasgos característicos de nuestra especie: la conciencia y el lenguaje (Sanjuán y González, 2005). De este modo, dos cuestiones científicas tan importantes como la búsqueda de los hechos diferenciales humanos y la búsqueda del origen de la esquizofrenia, se proponen como caminos compartidos. Así, la sintomatología de esta enfermedad, junto con las paradojas de la universalidad y de la persistencia en las poblaciones humanas, han permitido la elaboración de planteamientos evolutivos sobre el origen de la esquizofrenia. Un ejemplo de ello son las aproximaciones llevadas a cabo por Crow o Crespi, sobre la implicación de genes que han sufrido selección positiva en el linaje humano en el desarrollo de la esquizofrenia (Crow, 1997; Crespi *et al.*, 2007). Esta aproximación parte de que los genes y regiones reguladoras que han evolucionado de forma adaptativa pueden haber contribuido tanto al desarrollo de los rasgos fenotípicos característicamente humanos, como por ejemplo, el lenguaje, como de alguna forma, también a enfermedades típicamente humanas, como la esquizofrenia. Uno de los indicios que sugieren un fuerte impacto de la selección darwiniana en el linaje humano es la extraordinaria rapidez con la que han sido adquiridos algunos rasgos claves, como el tamaño del cerebro, las avanzadas habilidades cognitivas, los órganos vocales o la bipedestación (Sanjuán y González, 2005).

Para este estudio de asociación se utilizaron polimorfismos de dos genes: *FOXP2* y *HAR1A*. La selección de estos genes vino propiciada principalmente por sus importantes características evolutivas, como se muestra a continuación.

El gen *FOXP2*, constituye un paradigma de la acción de la selección positiva en el linaje humano, como ya se mencionó en el apartado de la Introducción (epígrafe V.6). Diferentes análisis muestran evidencias de la actuación de la selección positiva en este

locus en el linaje humano, siendo los candidatos más probables los dos cambios aminoacídicos ocurridos tras la separación del ancestro común con chimpancé (Enard *et al.*, 2002; Zhang *et al.*, 2002) y sitúan el tiempo de fijación de estos cambios en un momento compatible con la aparición de un lenguaje humano eficiente. Además, este gen es el primer y único gen descrito hasta la fecha relacionado directamente con una alteración del lenguaje (Lai *et al.*, 2001).

Las alteraciones del lenguaje son uno de los principales síntomas del síndrome psicótico (McKenna y Oh, 2005) y tienen un claro componente en las AA. Como avanzábamos antes, según nuestro modelo etiopatogénico de las AA en psicosis, debe existir en los pacientes con AA una vulnerabilidad a sufrir alteraciones del lenguaje que ocasionen la experiencia de voces. Este hecho, junto con la hipótesis de que la vulnerabilidad genética a la esquizofrenia haya evolucionado como consecuencia de la selección en rasgos cognitivos humanos (Crow, 1997), hace que la selección de un gen relacionado con el lenguaje y sometido a selección positiva en el linaje humano, como gen candidato a la esquizofrenia y en especial a las AA, sea una buena aproximación.

Por otra parte, el gen *HAR1A* es un gen de RNA no codificante que contiene el elemento HAR1, identificado como el elemento con tasa de sustitución más acelerada del linaje humano (Pollard *et al.*, 2006). Existen dos transcritos alternativos en sentido contrario a *HAR1A*, *HAR1R α* y *HAR1R β* que también contienen el elemento HAR1, y que por su patrón temporal y espacial de expresión podrían regular a *HAR1A* mediante inhibición antisentido. El gen *HAR1A* coexpresa con la reelina, proteína importante en la migración de células del neuroblasto y para la especificación de la estructura en capas de la corteza cerebral humana. Los niveles de expresión de la reelina se ven incrementados en la escala evolutiva del linaje de los primates y además, se ha encontrado que la cantidad de RNA mensajero en cerebros postmortem de pacientes esquizofrénicos se encuentra reducida en un 50% en todos los tejidos analizados (revisión en Costa *et al.*, 2001). Así, las características evolutivas del gen *HAR1A*, juntamente con su posible relación con el gen de la reelina, permiten considerar a este gen como un buen candidato para la esquizofrenia.

SNPs analizados

Para el análisis del gen ***FOXP2*** se analizaron 25 SNPs, distribuidos a lo largo de las 600 Kb que cubre el gen.

El descubrimiento de los bloques de haplotipos en el genoma humano, que pueden variar en tamaño desde pocas kilobases hasta más de 100 kb (Wall *et al.*, 2003), mostró que el LD se extiende usualmente a lo largo de distancias cromosómicas más largas de lo esperado hasta el momento y permitió estimar que el análisis de un SNP dentro de cada bloque sería suficiente en los estudios de asociación caso-control con enfermedades. Esto ha llevado a intentar aprovechar estas propiedades para llevar a cabo la selección de SNPs (*tagSNPs*) que capturen la variabilidad de una determinada región cromosómica. De este modo, se dispone de diferentes programas bioinformáticos, que utilizan información de bases de datos para obtener valores de LD entre SNPs ya analizados en población general y permiten la selección de los *tagSNPs*, como *Tagger*, o *BNTagger* (de Bakker *et al.*, 2005; Lee y Shatkay, 2006). Así, la información del LD permite seleccionar los SNPs que se van a utilizar en un estudio de asociación.

En este trabajo se han utilizado dos formas de selección de SNPs: una en la que los SNPs se seleccionaron por su localización, utilización en otros estudios o capacidad de

Discusión

ser genotipados por determinada técnica, y otra en la que se seleccionaron mediante el programa *Tagger*. Las ventajas de la utilización de un programa de selección de SNPs en función del LD existente entre ellos, así como la utilidad de bases de datos como la del *HapMap*, de la que toma la información el programa *Tagger*, se ven reflejadas al considerar varios factores. En primer lugar, como se menciona en el apartado de Resultados (epígrafe I.1.1), no se observan grandes bloques de LD entre los 25 SNPs del gen *FOXP2* analizados cuando se representan gráficamente los valores de r^2 , constituyendo una buena representación de la variabilidad de la región. Únicamente se pueden considerar como posibles bloques los formados por los polimorfismos rs7803667/ rs10447760/ rs6961558/ rs923875, que abarcan 12 Kb, rs1916977/rs2396722/ rs2253478/rs2694941, de 50 Kb, y rs1456031/rs2396765/ rs1456021 de 31 Kb (Figuras R3 y R4).

En segundo lugar, se observa en el hecho de que a pesar de que los polimorfismos rs7799652, rs1456029, rs12670585 y rs1456031 se encuentran en una región de poco más de 5 Kb, los valores de LD entre ellos son bajos. Con excepción de rs1456031, estos polimorfismos fueron seleccionados con el programa *Tagger*. Así, el bajo LD encontrado entre estos polimorfismos refleja la causa por la que se seleccionaron tantos SNPs para una región tan pequeña, que es la existencia de mayores valores de la tasa de recombinación que las regiones adyacentes.

La región de 5 Kb en la que se localizan los polimorfismos mencionados se encuentra localizada desde el intrón 8 del gen *FOXP2* hasta el intrón 10, región que incluye el dominio de cremallera de leucina, así como la presencia de un exón 10 alternativo, más largo que da lugar a una proteína truncada con posibles implicaciones funcionales.

Los patrones de LD obtenidos son similares entre pacientes y controles, indicando ausencia de grandes diferencias en las frecuencias de recombinación entre los diferentes polimorfismos en ambas muestras.

En el análisis de desequilibrio de ligamiento para los SNPs analizados del gen ***HAR1A*** se observó que el rango de valores de r^2 no resultó ser muy elevado. De los SNPs seleccionados, el rs750696, es un *tagSNP*, o SNP capaz de capturar la mayor parte de variabilidad de una región. Por esta razón, se esperaría que el resto de los SNPs seleccionados estuviera en relativo LD con este. Al no obtenerse elevados valores de LD podemos concluir que el *tagSNP* rs750696 probablemente no sea suficiente como para capturar toda la variabilidad de la región y que la utilización de los otros SNPs analizados enriquece nuestro poder para estimarla.

Así, a pesar de la gran utilidad de obtener información sobre la variabilidad de una región cromosómica mediante la genotipación de unos pocos SNPs (*tagSNPs*), para determinados fragmentos de genoma todavía no hay caracterizados los suficientes *tagSNPs* como para proporcionar un conocimiento exhaustivo de la región. En todo caso, siempre conviene llegar a un compromiso entre el número de SNPs a genotipar y el objetivo a llevar a cabo. Para el gen *HAR1A* se seleccionaron 6 SNPs que cubren 5 Kb, mientras que para el caso del gen *FOXP2*, que cubre más de 600 Kb se seleccionaron 25. Obviamente, se habrían podido seleccionar más SNPs para caracterizar mejor la región del *FOXP2*, pero se decidió que con 25 se tendría una idea aproximada de la variabilidad y que si se obtenían resultados interesantes para algún polimorfismo, se podía genotipar nuevos SNPs.

Métodos de genotipación

Para llevar a cabo la genotipación de los SNPs analizados se utilizaron dos métodos, uno tradicional, la genotipación por RFLPs y otro utilizado por las plataformas de genotipación a gran escala, un ensayo iPLEX™ utilizando tecnología MassArray de Sequenom®.

Cuando se iniciaron los estudios de asociación, la genotipación por RFLPs surgió como un método fácil, rápido, reproducible y relativamente económico de determinar los alelos presentes de un determinado SNP. El método sencillo de amplificar por PCR un fragmento de DNA que incluyera el polimorfismo en cuestión y posteriormente digerir con un enzima de restricción que cortara o no el fragmento en función del nucleótido presente resultaba accesible a la mayoría de laboratorios interesados en llevar a cabo este tipo de estudios, puesto que no era necesario un equipamiento especial.

Sin embargo este tipo de metodología tiene una serie de limitaciones a tener en cuenta:

- En primer lugar, la posibilidad de genotipar un SNP por RFLPs viene determinada por la presencia o no de un sitio de reconocimiento de una enzima de restricción así como de la capacidad para poder discriminar entre los diferentes alelos. En ocasiones se puede solventar el problema creando cebadores con una modificación que cree una diana de enzima de restricción al amplificar por PCR, pero en estos casos la determinación del genotipo se ve complicada por el tamaño de los fragmentos generados, como fue el caso del polimorfismo rs6961558. Para este SNP los fragmentos que permitían diferenciar los diferentes alelos diferían únicamente en 21 pb.
- En segundo lugar, el creciente tamaño de las muestras utilizadas en los estudios de asociación, así como el incremento en el número de SNPs analizados ha convertido lo que era un método rápido en un método largo y tedioso en comparación con los que se utilizan en plataforma.
- Por último, el coste económico puede ser alto, dependiendo del tipo de enzima de restricción utilizado.

De este modo, aunque la utilización de los RFLPs puede ser conveniente e incluso ventajosa en algunos casos, en la actualidad, con la aparición de plataformas y servicios de genotipación a gran escala, ya no es un método tan extendido.

El desarrollo de las técnicas de genotipación automatizadas y de programas informáticos necesarios para la interpretación de resultados ha marcado el inicio de los estudios de asociación del genoma completo o GWAs. Sin estas tecnologías, el análisis de cientos de miles de SNPs en miles de muestras habría conllevado una inversión en tiempo y coste económico impensable con otro método de genotipación. Aún así, es una tecnología en continuo perfeccionamiento que todavía tiene algunos factores por mejorar:

- Determinados SNPs son difíciles o imposibles de genotipar. Esto sucede especialmente cuando el polimorfismo se localiza en regiones ricas en GC. Puesto que una elevada proporción de promotores de genes contiene islas GC, que potencialmente pueden contener elementos reguladores afectados por SNPs, la dificultad de genotipar estos SNPs que pudieran ser funcionales supone una limitación.
- Fallos en la genotipación. En algunos casos un elevado % de las muestras fallan en la genotipación. Este hecho puede observarse en nuestro estudio, donde en

Discusión

muchos casos para algunos SNPs no se obtuvo información para el 10% de las muestras analizadas. Esto puede redundar en considerar el SNP analizado como un polimorfismo de calidad cuestionable. En algunos estudios como el de The Wellcome Trust Case Control Consortium, los SNPs con frecuencia de pérdida >3% se eliminan (WTCCC, 2007)

En este trabajo se han utilizado dos metodologías muy diferentes con un mismo objetivo: la genotipación de un total de 37 SNPs, 31 de los cuales resultaron polimórficos en nuestra muestra. La obtención de los resultados de la genotipación de 10 SNPs por RFLPs conllevó un trabajo de más de dos años, mientras que la genotipación mediante un método de gran escala supuso tres meses. En este sentido se puede observar que al comparar ambos métodos la genotipación en plataforma es mejor para la obtención rápida de resultados. En cualquier caso, al iniciarse este trabajo, no se disponía de las herramientas necesarias para llevar a cabo el análisis a gran escala, y de las opciones disponibles (secuenciación, hibridación, etc.), la genotipación por RFLPs era la más ventajosa.

▪ *Gen FOXP2*

Estudio de asociación

Desde su aislamiento y caracterización el gen *FOXP2* ha sido ampliamente utilizado en diferentes estudios de asociación caso-control con patologías relacionadas con el lenguaje, principalmente autismo (ver tabla I5 de la Introducción). Sin embargo, hasta la fecha existen únicamente tres estudios de asociación de este gen en relación con la esquizofrenia, dos de ellos llevados a cabo en nuestro grupo de investigación, con una submuestra del presente estudio (Sanjuán *et al.*, 2005; Sanjuán *et al.*, 2006b) y un tercero realizado recientemente en población coreana (Jung *et al.*, 2008).

En el primero de los estudios, en el que se analizaron 149 pacientes y 137 controles, se encontró ausencia de asociación con los dos polimorfismos analizados: rs923875 y rs17137124. En el segundo estudio, en el que se comparaban las frecuencias genotípicas, alélicas y haplotípicas de 10 SNPs (rs7803667, rs10447760, rs923875, rs1597548, rs2396722, rs1358278, rs1852469, rs2396753, rs17137124 y rs1456031) en 186 pacientes y 160 controles se encontraron diferencias significativas para las frecuencias alélicas y genotípicas del polimorfismo rs2396753 al comparar entre pacientes con AA y controles ($P=0.0027$ y $P=0.007$, respectivamente), aunque las diferencias no se mantuvieron al corregir por Bonferroni secuencial. El análisis haplotípico en este estudio proporcionó evidencias de un haplotipo protector (rs7803667T/rs10447760C/rs923875A/rs1358278A/rs2396753A $P=0.009$).

Todos los polimorfismos de este estudio están incluidos en el análisis de la presente tesis doctoral, con excepción del polimorfismo rs1358278, para el que se había encontrado un desequilibrio de ligamiento completo con el polimorfismo rs2396722, localizado 166 pb aguas arriba (Sanjuán *et al.*, 2006b).

El estudio de asociación realizado con los 25 SNPs en el presente trabajo no mostró evidencias de diferencias significativas en las frecuencias alélicas, genotípicas entre pacientes y controles, pacientes con AA y controles o pacientes con esquizofrenia y controles. Únicamente el polimorfismo rs2396753, el mismo para el que se había obtenido significación en el estudio anterior (Sanjuán *et al.*, 2006b), muestra una tendencia hacia la significación. Al comparar pacientes con controles se obtiene diferencias significativas para las frecuencias genotípicas, aunque la significación se

pierde al llevar a cabo la corrección múltiple. Cuando se comparan pacientes con AA con controles, las diferencias son sustancialmente mayores, encontrándose un aumento en las frecuencias del alelo C en pacientes (posible alelo de riesgo en nuestra muestra) aunque nuevamente no resisten la corrección por Bonferroni.

Este polimorfismo, rs2396753, ha sido analizado previamente en relación con autismo en población japonesa (Gong *et al.*, 2004) aunque no se encontró asociación con él. Posteriormente, también se analizó su relación con AA dentro del síndrome psicótico, sin obtenerse tampoco diferencias significativas (Jung *et al.*, 2008).

Se encuentra localizado en el intrón 2b del gen *FOXP2*, aproximadamente 9 kilobases aguas abajo del exón 2b y 26 kilobases aguas arriba del exón 3, por lo que no parece estar implicado directamente en la vulnerabilidad a las AA, si no más bien a través del desequilibrio de ligamiento con otros polimorfismos potencialmente funcionales. Nos parece importante hacer notar que las frecuencias alélicas correspondientes a este polimorfismo varían sustancialmente entre las poblaciones para las que se dispone de esa información (tabla D1), por lo que cualquier susceptibilidad a las alucinaciones conferida por uno de los alelos tendría que tenerse en cuenta únicamente dentro de dicha población. Esto explicaría porqué en el estudio de Jung y colaboradores no replicaron la asociación encontrada por nuestro grupo.

Población del <i>HapMap</i>	Frecuencias alélicas del SNP rs2396753	
	C	A
CEU (residentes en UTAH con ascendencia europea)	0.44	0.56
CHB (población china de Beijing)	0.39	0.61
JPT (población japonesa de Tokio)	0.54	0.46
YRI (población de Yoruba, Nigeria)	0.16	0.84

Tabla D1. Frecuencias alélicas del SNP rs2396753 en poblaciones incluidas en el *HapMap*.

En el estudio de Jung y colaboradores, en el que se analizaron los polimorfismos rs923875, rs2396753 y rs1456031 (todos incluidos en el presente trabajo) en 198 pacientes con AA y 196 controles, se encontraron diferencias significativas entre ambos grupos para las frecuencias genotípicas del polimorfismo rs17137124, así como diferencias para este mismo SNP con el ítem correspondiente a Delirios de la escala PANSS (Jung *et al.*, 2008). Sin embargo, el polimorfismo rs17137124 no muestra diferencias significativas en nuestra muestra.

En el artículo de Gong y colaboradores se encontró transmisión preferente del alelo C del polimorfismo rs1456031 en niños con autismo, así como asociación con el haplotipo rs1456031C/rs2396753C/rs1852469. En nuestro estudio no se ha encontrado asociación con dicho SNP ni con el haplotipo en cuestión.

En este sentido nos gustaría resaltar dos cosas en relación al hecho de que la utilización de pacientes con AA proporciona resultados más significativos que la utilización del grupo con todos los pacientes al compararse con los controles. Por una parte, estos resultados reflejan la validez de la utilización de las AA como fenotipo alternativo al síndrome psicótico en un estudio genético, constituyendo un fenotipo más acotado y homogéneo. Por otra parte, es necesario comparar los resultados obtenidos con los correspondientes a la comparación entre pacientes con AA y pacientes sin AA, para poder diferenciar entre vulnerabilidad genética a las AA y vulnerabilidad a las AA dentro del síndrome psicótico. Además, puesto que las AA también tienen lugar entre población normal (Johns y van Os, 2001), sería conveniente comparar dentro de individuos sanos, dos grupos, uno con AA y otro sin AA.

Discusión

Los valores de poder estadístico obtenidos para las comparaciones pacientes-contróles y pacientes con AA-contróles son muy bajos (en su mayoría menores del 0.7, que se considera por consenso un valor aceptable), lo que indica que la capacidad para detectar falsos negativos en nuestra muestra para los valores de OR estimados es baja. No obstante, podemos considerar que este cálculo es atípico, en el sentido de que se obtuvo *a posteriori*, una vez obtenidos los valores de OR y no *a priori*, y seleccionando valores de OR típicos en enfermedades mentales, como se utiliza normalmente. Además, los valores de OR no dejan de ser una estima, hecho agravado por los amplios intervalos de confianza obtenidos. La determinación del poder estadístico se llevó a cabo de la forma indicada debido a la ausencia de información, antes del inicio del estudio, de las frecuencias alélicas de algunos de los polimorfismos utilizados, datos necesarios para el análisis.

Respecto al análisis de haplotipos, al comparar tanto pacientes frente a controles, pacientes con AA frente a controles como pacientes con esquizofrenia frente a controles, no se obtuvieron diferencias significativas para ninguno de los test globales aplicados a las combinaciones de 4 SNPs consecutivos. Sin embargo, al analizar la combinación empleada en Sanjuán *et al.*, 2006 sí se obtuvieron diferencias globales entre controles y pacientes con AA y se reprodujo la tendencia hacia un posible haplotipo de protección: rs7803667T/ rs10447760C/ rs923875A/ rs2396722C/ rs2396753A. Este haplotipo incluye el polimorfismo rs2396753, el cual como se ha comentado, presentaba tendencia a mostrar diferencias significativas entre controles e individuos psicóticos con AA. En concreto, este potencial haplotipo protector, con mayor frecuencia en controles, contiene el alelo A, que es el de menor frecuencia en pacientes. Esto refuerza el resultado obtenido en las comparaciones de las frecuencias genotípicas y alélicas, donde se encontró un aumento en las frecuencias del alelo C en pacientes con AA.

La obtención de resultados negativos para el análisis por ventanas consecutivas y positivos para una combinación determinada de SNPs no consecutivos, es un indicador de las limitaciones de los análisis de haplotipos por ventanas consecutivas.

Un paso crucial en el análisis de haplotipos es la selección de marcadores. Cuando se trabaja con muchos SNPs, los haplotipos posibles pueden ser demasiados, bien para poder estimar las frecuencias, especialmente si hay poco LD entre ellos, o bien para poder ser analizados mediante los algoritmos disponibles (revisión en Zhao *et al.*, 2003). Además, el poder estadístico de los análisis de haplotipos se reduce por el elevado número de haplotipos potenciales que necesitan ser estudiados. Ante la imposibilidad de llevar a cabo el análisis de haplotipos con todos los SNPs genotipados, debido a que ante cada haplotipo diferente la muestra se fraccionaría cada vez más, la utilización de ventanas surge como una de las alternativas para reducir el número de haplotipos considerados en el estudio de asociación. Sin embargo en el proceso se pierde información. Esto sucede debido a que se descarta la posible interacción entre variantes lejanas, bajo la base de que será cerca de la variante causal o que confiera vulnerabilidad a la enfermedad donde mayor probabilidad haya de encontrar un bloque de SNPs o haplotipo de riesgo.

Análisis de variables clínicas

Como se indicó en la sección de Resultados (epígrafe I.1.5) nuestro análisis se ha centrado principalmente en variables relacionadas con las AA, variables relacionadas con el lenguaje y variables relacionadas con el inicio de la enfermedad.

En relación con las AA se analizaron diferentes variables: presencia o ausencia de AA, cronicidad y las variables incluidas en la escala PSYRATs de evaluación de las AA.

No se han obtenido evidencias de diferencias significativas entre pacientes con AA y pacientes sin AA. El polimorfismo rs2396753 anteriormente mencionado mostró una tendencia, sin embargo tras la corrección múltiple no se mantuvo la significación. En cuanto al análisis de haplotipos, se obtuvieron diferencias significativas después de la corrección para algunos de los test individuales de haplotipos. No obstante, los tests globales de estas combinaciones no resultaron significativos.

Del mismo modo, no se encontró evidencia de que variaciones en los polimorfismos analizados influyan en la vulnerabilidad a ninguno de los ítems de las AA incluidos en la escala **PSYRATs**. Esta escala, incluye tanto ítems relacionados con las características físicas de las alucinaciones (frecuencia, duración, localización, intensidad) como relacionados con las características emocionales (grado de convicción, contenido negativo, ansiedad, repercusión). En este sentido, no parece que el gen *FOXP2* esté relacionado con estas características, sino más bien con la aparición en sí de las AA, como se presume a partir de los resultados de la comparación de pacientes con AA y controles.

Dada la relación del gen *FOXP2* con el lenguaje, se analizaron dos variables relacionadas con este rasgo, ambas incluidas en la escala Manchester: **Pobreza del Lenguaje e Incoherencia del Lenguaje**. Únicamente el polimorfismo rs2253478 mostró diferencias significativas frente al ítem de Pobreza del Lenguaje tras la corrección múltiple. Este polimorfismo está localizado en el intrón s3. Se desconoce si podría resultar ser funcional por sí mismo o estar en desequilibrio de ligamiento con algún otro.

El gen *FOXP2* se ha relacionado con la dispraxia verbal del desarrollo, que consiste en problemas en el aprendizaje y producción de las secuencias de movimientos vocales necesarias para el lenguaje, y que a menudo va acompañado de problemas en gramática y lenguaje expresivo o receptivo (Watkins *et al.*, 2002). En aves con aprendizaje vocal en las que se ha provocado una disminución de la expresión de FoxP2 se observa una alteración en la estructura del canto de forma que los individuos jóvenes no son capaces de imitar correctamente a los adultos (Haesler *et al.*, 2007). El resultado de asociación obtenido con el ítem de Pobreza del lenguaje resulta difícil de encuadrar dentro de las alteraciones descritas, de no ser que los problemas gramaticales y de lenguaje expresivo o receptivo desencadenaran una pobreza en el lenguaje.

Pese a constituir un gen del neurodesarrollo y estar relacionado con fenómenos de plasticidad (Haesler *et al.*, 2007; Teramitsu *et al.*, 2007), característicos entre otros de la transición a la edad adulta, los análisis de asociación relativos a la variable de **inicio temprano o tardío de la enfermedad** indican que es poco probable que el gen *FOXP2* esté implicado en la aparición temprana o tardía del primer episodio psicótico.

En cuanto a la **cronicidad** de las alucinaciones se encontraron evidencias de diferencias significativas para los polimorfismos rs1456031 y rs1456021, apoyadas por el hecho de que en el análisis de haplotipos el haplotipo que contiene ambos SNPs muestra diferencias globales entre ambos grupos. Este resultado sin embargo, debe ser tomado con precaución y sería recomendable la repetición del análisis con una muestra mayor, para descartar que este resultado sea un falso positivo.

Discusión

- Gen *HAR1A*

Estudio de asociación caso-control

Hasta el momento, únicamente se ha llevado a cabo un estudio de asociación caso-control con polimorfismos del gen *HAR1A* (Tolosa *et al.*, 2008). En dicho estudio no se encontraron diferencias en las frecuencias genotípicas, alélicas o haplotípicas al comparar controles con pacientes con AA, aunque sí se encontró un haplotipo de riesgo, rs6011613C /rs2427496C /rs6122371C /rs750696C /rs750697G /rs2427498C en relación con la aparición de AA dentro del síndrome psicótico. Las muestras empleadas en ambos casos son prácticamente idénticas, con excepción de unos pocos pacientes y controles, no incluidos en el presente trabajo.

No se encontraron diferencias significativas en las frecuencias genotípicas, alélicas, o haplotípicas al comparar controles, pacientes, pacientes con AA y pacientes con esquizofrenia, por lo que no parece que el gen *HAR1A* esté relacionado con la vulnerabilidad al síndrome psicótico. Como era de esperar, estos resultados coinciden con los obtenidos en Tolosa *et al.*, 2008.

En cualquier caso, al igual que en el caso del análisis con el gen *FOXP2*, los bajos valores de poder estadístico que se obtuvieron son bajos, lo que indica que la capacidad para detectar asociación de nuestra muestra es limitada.

Análisis de variables clínicas

Al tener en cuenta las variables clínicas relacionadas con las AA los resultados más relevantes se obtuvieron al comparar las frecuencias haplotípicas entre pacientes con AA y pacientes sin AA. Estos resultados coinciden parcialmente con los obtenidos en Tolosa *et al.*, 2008, donde se encontraron diferencias significativas para el haplotipo rs6011613C /rs2427496C /rs6122371C /rs750696C /rs750697G /rs2427498C entre ambos grupos. En la presente tesis, aunque se encontró una tendencia para este mismo haplotipo, las diferencias no se mantuvieron significativas tras la corrección por el número de haplotipos. Esto, como se menciona en el apartado de Resultados (epígrafe I.2.5.2), puede ser debido, bien a la pequeña variación en el tamaño de la muestras utilizadas en cada caso (las frecuencias obtenidas son prácticamente las mismas), o bien a la modificación del parámetro de valor umbral para haplotipos raros (0.05 en el artículo y 0.03 en la presente tesis).

En cualquier caso, puesto que las frecuencias en pacientes con AA y controles no difieren significativamente entre sí, la implicación funcional del haplotipo en cuestión se daría dentro del contexto psicótico.

I.2 Búsqueda de expansiones de trinucleótidos en el gen *FOXP2*

Una aproximación al estudio genético de la esquizofrenia que está cobrando actualmente gran interés es el estudio de las variaciones en el número de copias (CNVs): duplicaciones o deleciones de segmentos genómicos entre 1 Kb y millones de bases (Cantor y Geschwind, 2008; Feuk *et al.*, 2006). Un ejemplo lo constituyen los dos primeros estudios a gran escala realizados, en los que se observa que el número

de deleciones o duplicaciones de baja frecuencia aumenta en pacientes con esquizofrenia (*The International Schizophrenia Consortium*, 2008; Stefansson *et al.*, 2008).

Aunque no entran dentro de la definición más restrictiva de los CNVs (Feuk *et al.*, 2006), las expansiones de repeticiones de DNA constituyen otra fuente de variabilidad, referida también al número de copias de determinados segmentos. Existen cerca de 20 trastornos neurológicos causados por expansiones inestables de repeticiones de DNA (Gatchel y Zoghbi, 2005). Dentro de estas expansiones inestables destacan las correspondientes a las repeticiones de trinucleótidos: CTGn en distrofia miotónica y ataxia espinocerebelar tipo 8 (SCA8); la cadena reversa complementaria, CAGn, en SCA1, SCA2, SCA3, SCA6, SCA7 y SCA12, en la enfermedad de Machado-Joseph y la atrofia dentatorubral-palidoluisiana (DRPLA); AAG en ataxia de Friedreich y CGGn y su cadena complementaria CCGn en el síndrome de X Frágil y retraso mental específico del cromosoma X (FRAXE) respectivamente (Wilmont y Warren, 1998; Koob *et al.*, 1999; Holmes *et al.*, 1999). Aunque las expansiones de repeticiones de trinucleótidos son las más frecuentes, también se conocen expansiones inestables de tetranucleótidos y pentanucleótidos (Liquori *et al.*, 2001; Matsuura *et al.*, 2000)

En la expansión de repeticiones en estas enfermedades intervienen diferentes trinucleótidos y diferentes rangos en la longitud patológica de las repeticiones. En algunas, las repeticiones codifican para glutamina (Enfermedad de Parkinson, DRPLA, SCA1, SCA2, SCA3, SCA6, SCA7) mientras que en otras ocurren en la región 3' no traducida (distrofia miotónica y SCA8), en la región 5' no traducida (Síndrome del X Frágil, FRAXE, SCA12) o en un intrón (Ataxia de Friedreich). Los mecanismos patogénicos en los trastornos producidos por repeticiones inestables incluyen pérdida de función, ganancia de función (debido al aumento de la función o bien a adquisición de nueva función de la proteína), o función alterada como consecuencia de interacciones RNA-proteína aberrantes (revisado en Gatchel y Zoghbi, 2005). Además, mientras que muchas de estas enfermedades comparten rasgos clínicos similares, los genes causales de la enfermedad codifican para proteínas de diversa función.

Las repeticiones de DNA inestable están relacionadas con el fenómeno de anticipación genética (el corea de Huntington constituye el ejemplo más claro), en la que en sucesivas generaciones el inicio de una enfermedad se produce de forma más temprana al tiempo que aumenta la severidad de los síntomas.

Desde el descubrimiento del DNA inestable, y ante la posibilidad de una relación con la anticipación genética para explicar algunos factores de los trastornos psicóticos, se han realizado diferentes estudios en los que se planteaba que la anticipación genética tuviera lugar en psicosis. Los estudios llevados a cabo al respecto proporcionaron resultados sorprendentemente consistentes, aunque se ha cuestionado si la utilización de los métodos estadísticos aplicados es la apropiada para una enfermedad compleja como es la psicosis (revisado en Vincent *et al.*, 2000). En cualquier caso, la idea de anticipación genética se mantiene en la actualidad, a través de los resultados obtenidos de la observación de casos familiares (Blicinc-Pesek y Agius, 2007).

Se han llevado a cabo diferentes estudios buscando evidencias de expansiones en repeticiones de trinucleótidos a nivel genómico y en determinados genes, tanto en esquizofrenia como en trastorno bipolar (revisado en Vincent *et al.*, 2000; Glatt *et al.*, 2003; Tsutsumi *et al.*, 2004). Sin embargo, los métodos utilizados inicialmente no tenían sensibilidad suficiente para detectar expansiones de menor longitud y aunque se

Discusión

obtuvieron algunos resultados positivos, no parece que grandes variaciones en la longitud de fragmentos con repeticiones de trinucleótidos contribuyan de forma importante a la vulnerabilidad en la esquizofrenia.

El gen *FOXP2* contiene dos tramos de poliglutaminas de 40 y 10 repeticiones, localizados en los exones 5 y 6 respectivamente. La repetición de 40 glutaminas constituye el tramo de glutaminas no patológico de mayor longitud observado en una proteína. De hecho, mediante estudios *in vitro* se ha determinado que tramos de poliglutaminas con más de 37 repeticiones forman agregados insolubles, de forma que la longitud normal del tramo de *FOXP2* se encontraría en el límite establecido para expansiones patológicas (Scherzinger *et al.*, 1999).

Diferentes estudios han llevado a cabo el análisis de estos tramos de poliglutaminas en relación con diversos trastornos motores o del lenguaje, principalmente autismo, encontrándose poca o ninguna variación en el número de glutaminas (Newbury *et al.*, 2002; Wassink *et al.*, 2002; Bruce y Margolis, 2002; Gauthier *et al.*, 2003; Li *et al.*, 2005; MacDermot *et al.*, 2005). Esto es coherente con las observaciones de Butland y colaboradores, quienes, estimaron que debido a no tratarse de un tramo de repeticiones CAG puras (y estar interrumpido con codones CAA), el nivel de polimorfismo del tramo de 40 glutaminas debía ser bajo (Butland *et al.*, 2007).

Recientemente, durante la redacción de esta tesis, se publicó la primera búsqueda de expansiones de glutaminas en pacientes esquizofrénicos (Laroche *et al.*, 2008). Estos autores no encontraron ninguna variante de riesgo en los tramos de poliglutaminas asociada a la enfermedad, aunque sí encontraron en pacientes con autismo dos cambios, una delección de 4 aminoácidos y una inserción de 4 aminoácidos.

Los resultados obtenidos en nuestro trabajo, es decir, la ausencia de expansiones en los tramos de poliglutaminas del gen *FOXP2* en pacientes esquizofrénicos, con y sin AA, coinciden con los de Laroche y colaboradores. Este hecho, unido a la poca variabilidad mostrada en estas repeticiones en otras patologías, indica, en primer lugar una elevada estabilidad en la secuencia repetida, también mantenida a lo largo de la evolución (el tramo de 40 poliglutaminas se mantiene con pocas variaciones entre especies, desde *Xenopus* hasta *Homo sapiens*). En segundo lugar, podemos concluir que no es probable que expansiones en este gen tengan implicación en enfermedades neurológicas. De ambas observaciones se deduce la importancia que debe tener la función de los tramos de poliglutaminas en la proteína *FOXP2*, para que la región codificante correspondiente se mantenga estable.

En nuestro trabajo se analizó también la posible expansión de trinucleótidos en una región rica en trinucleótidos CGG, localizada en el intrón s1, unas 1280 pares de bases aguas arriba del inicio de transcripción. Éste análisis se llevó a cabo debido a dos razones. En primer lugar debido a la detección de una delección de 9 pares de bases en un paciente con esquizofrenia durante la genotipación de un polimorfismo cercano. En segundo lugar, la región se localiza en el extremo de una isla CG que incluye el primer exón del gen, y la inestabilidad en las repeticiones podrían dar lugar a alteraciones en la expresión del gen, como se ha visto en el caso de otros trastornos neurológicos. Por ejemplo, la expansión de trinucleótidos CGG en el extremo 5' no traducido del gen *FMR1* causa el retraso mental del Síndrome del X Frágil (Verkerk *et al.*, 1991) y la expansión de CCG en extremo 5' no traducido del gen *FMR2* da lugar al Síndrome del X Frágil E a través de la hipermetilación de una isla CG y el silenciamiento transcripcional del gen (Knight *et al.*, 1993; Gecz *et al.*, 1996; Gu *et al.*, 1996).

Nuestro análisis, sin embargo, no reveló ninguna variación en la longitud de la región en pacientes o controles, con excepción de la ya mencionada, por lo que no parece que se produzcan expansiones en este tramo que puedan afectar a la expresión del gen.

I.3 Búsqueda de mutaciones en la región promotora

En el momento de iniciarse el proyecto de esta tesis, con excepción de la mutación identificada en la familia KE (Lai *et al.*, 2001) y dos cambios sinónimos encontrados en el exón 5, G247G y Q190Q, el primero en un paciente con autismo y el segundo en 2 pacientes con autismo y 4 controles (Wassink *et al.*, 2002; Gauthier *et al.*, 2003), no se había identificado ningún cambio o mutación en la región codificante del gen *FOXP2*.

La mutación encontrada en la familia KE fue analizada en nuestra muestra de controles y pacientes con psicosis, sin encontrarse en ninguno de los casos. Esto concuerda con el hecho de que se trate de una mutación puntual familiar.

Durante el desarrollo de nuestro trabajo, se describieron dos nuevos cambios en secuencia codificante: un cambio no sinónimo localizado en el exón 2 (Q17L), cerca del extremo N-terminal y un codón de parada en el exón 7 (R328X) que da lugar a una proteína truncada. Ambos cambios se identificaron en una búsqueda de mutaciones en región codificante llevada a cabo en individuos con dispraxia verbal (McDermot *et al.*, 2005). Posteriores ensayos funcionales con las proteínas resultantes para ambos cambios indican que el primero de ellos probablemente no sea una mutación patológica; el segundo, sin embargo sí parece afectar a las funciones normales de la proteína nuclear *FOXP2* (Vernes *et al.*, 2006).

Esta ausencia de variabilidad en la secuencia de DNA y el elevado grado de conservación a lo largo la evolución de la proteína para la cual codifica, nos hizo centrar nuestro análisis en la región promotora, concretamente en la región 5' al exón s1, bajo la hipótesis de que cambios en esta región podrían afectar la expresión del gen. De este modo, y aunque no constituía uno de los objetivos principales del trabajo, se llevó a cabo una búsqueda de variabilidad en la región promotora en una muestra de niños con Trastorno Específico del Lenguaje. La utilización de esta muestra y no de los pacientes con psicosis de los que disponíamos se basó en el hecho de que en estos individuos, la patología asociada al lenguaje es más específica que en el caso de los psicóticos.

Se identificaron 3 variantes no presentes en bases de datos, localizadas en las posiciones -3954, -1048 y +147 respecto a la primera posición del exón s1. La primera de ellas, un cambio G/A en posición -3954, encontrada en heterocigosis en 5 individuos, fue identificada también en controles, por lo que probablemente se trate de un polimorfismo todavía no descrito.

En cuanto a las otras dos, ambas se encontraron en el mismo individuo, en heterocigosis y no se detectó su presencia en controles. Esto plantea diferentes marcos de actuación de los cambios. En primer lugar, ambos podrían tratarse de polimorfismos de baja frecuencia que no han sido identificados en las muestras control utilizadas, debido al bajo número de controles analizados para ello. En segundo lugar, podrían constituir cambios que afectaran a la expresión del gen. El cambio localizado aguas arriba del inicio del gen se encuentra apenas una kilobase antes del inicio de la transcripción, por lo que no es descartable su aspecto funcional. Un cambio de base en

Discusión

esa posición podría afectar dos potenciales dianas de factores de transcripción correspondientes a Elk-1 y la proteína de unión a X-box. En cuanto al cambio localizado en posición +147, este se encuentra en el primer exón del gen, aunque se trata de un exón no codificante. Esta región además, constituye una región de elevado contenido en GC, en la que el cambio identificado podría afectar también a la unión de potenciales dianas de factores de transcripción como: WT1, el factor de unión a CCCTC, el factor tipo *kruppel* enriquecido en riñón, o elementos de la caja GC, entre otros. En cualquier caso, la presencia de ambos cambios en un mismo individuo podría hacer que el pequeño efecto sobre la expresión del gen que pudieran tener individualmente, se viera amplificado. Aunque este hecho no deja de ser una especulación.

II. Análisis del promotor del gen *FOXP2*

En el momento de iniciar el trabajo de esta tesis, la única aproximación a la estructura del promotor del gen *FOXP2* había sido llevada a cabo por Bruce y Margolis, quienes al analizar las regiones 5' de los exones s1, s2, s3 y 1, únicamente identificaron una potencial región promotora aguas arriba del exón s1. Estos autores detectaron por métodos bioinformáticos dos regiones prometedoras cerca de este exón, una entre las posiciones -235 y +220 y la otra entre -1531 y -1208, ambas coincidentes con la isla CG (Bruce y Margolis, 2002).

Es por esta razón que en nuestro estudio fue esa la región considerada como promotora.

II.1 Análisis evolutivo de la región promotora

El elevado grado de conservación observado en las comparaciones genómicas hombre-chimpancé, especialmente pronunciado en las regiones codificantes, es insuficiente para explicar las diferencias anatómicas y de comportamiento entre ambas especies. Por esta razón se ha propuesto que cambios en regiones reguladoras que den lugar a diferencias en expresión podrían haber jugado un papel importante en la evolución (King y Wilson 1975; Carroll, 2005). Aunque las regiones promotoras poseen una plasticidad ausente en regiones codificantes (Tirosh *et al.*, 2008), más sensibles a las variaciones en la secuencia, cambios en las regiones promotoras pueden influir en los niveles de expresión, probablemente debido a la unión diferencial de factores reguladores. Son por tanto sensibles también a la actuación de selección positiva y negativa (Kohn *et al.*, 2004). Así, son de especial interés observaciones como la de Prabhakar y colaboradores, que encontraron un exceso de sustituciones específicas en el hombre en secuencias no codificantes conservadas cerca de genes que se expresan en cerebro (Prabhakar *et al.*, 2006).

El ensombrecimiento filogenético (*phylogenetic shadowing*), ha sido utilizado con éxito para identificar elementos reguladores funcionales, bajo el supuesto de que como resultado de una selección purificadora, estas regiones se encuentran más conservadas que las secuencias que evolucionan de forma neutra. Con esta perspectiva, en este estudio se analizó la región del gen *FOXP2* localizada aguas arriba del primer exón (exón s1) en diferentes especies. Del estudio comparativo de las secuencias de primates, rata y ratón se han obtenido 3 regiones de elevada conservación en las que

podrían existir elementos reguladores importantes: (-3050, -2962), (-743, -401) y (-21, +35).

Entre las especies de primates, como era de esperar dada la cercanía filogenética entre las mismas, los bloques de mayor conservación cubrían una mayor parte de la secuencia analizada como puede verse en el apartado de Resultados (epígrafe II.3).

Una vez determinadas qué regiones podrían contener elementos funcionales, teniendo en cuenta su conservación evolutiva, se procedió a intentar caracterizar por otros métodos bioinformáticos la región promotora. Así, mediante programas de predicción de promotores se obtuvieron diferentes predicciones informáticas, alrededor de las regiones (-1480, -1157), (-184,508) y (1173, 1444).

Sorprendentemente, las regiones asignadas como predicciones de secuencias promotoras del gen no coinciden con las obtenidas como candidatas en función del grado de conservación. Se observa una discrepancia entre la asignación basada en la presencia de motivos reguladores, llevada a cabo especie por especie y el ensombrecimiento filogenético, basado en evolución a diferentes tasas. Sin embargo, el hecho de que en la mayoría de los primates las posiciones de las predicciones sean tan similares indica que deben estar presentes motivos similares o al menos la distribución de los mismos tiene que ser similar.

Las predicciones para las regiones promotoras de nuestro estudio no coinciden exactamente con las obtenidas por Bruce y Margolis (2002), como puede verse en la figura D1.

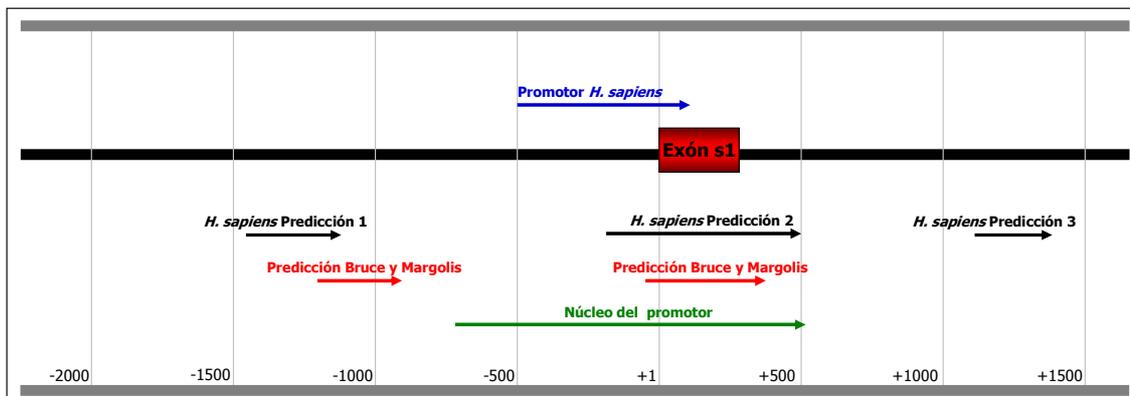


Figura D1. Localización del promotor extraído de *El Dorado* (marcado en azul), de las predicciones de promotores obtenidas en nuestro estudio con el *PromoterInspector* de *Genomatix* (marcado en negro) y las obtenidas por Bruce y Margolis (2002) por el mismo método (marcado en rojo). En verde se muestra la región que se propuso como núcleo del promotor. Las distancias respecto a la primera base del exón s1 se muestran en pares de bases.

Dado que en ambos casos el programa utilizado es el mismo, suponemos que las diferencias podrían ser ocasionadas a la actualización de las bases de datos con las que trabaja el programa *PromoterInspector*. Es de destacar que en nuestro caso utilizamos como sitio +1 el correspondiente al extremo 5' del gen *FOXP2* obtenido de la base de datos del UCSC browser y Bruce y Margolis utilizaron un extremo 5' diferente, el identificado en su estudio, aunque por comparación con ratón indicaron que el exón s1 podría extenderse aguas arriba. Curiosamente, las predicciones localizadas alrededor de -1500 tienen exactamente la misma longitud y la distancia respecto a la predicción que solapa con el primer exón es la misma. Bruce y Margolis no analizaron la región donde se encuentra la Predicción 3.

Discusión

En función de los datos obtenidos en la búsqueda de regiones conservadas y de predicción bioinformática de promotores nos planteamos la existencia de una región promotora muy conservada en mamíferos que podría constituir el núcleo del promotor entre las posiciones -749 y 507. Este fragmento reúne dos características importantes que nos hacen pensar que contiene importantes secuencias reguladoras, además de su cercana localización al sitio de inicio de la transcripción. En primer lugar incluye una región extremadamente conservada en las especies analizadas, con un único cambio en los alineamientos de secuencias de primates (figura R22, epígrafe II.3 de Resultados). Este grado de conservación nos hizo pensar que debe haber una fuerza selectiva que impida la aparición de cambios en el fragmento. En segundo lugar, la que proponemos como región núcleo del promotor incluye los promotores extraídos de bases de datos para *H. sapiens*, *P. troglodytes*, *M. mulatta* y *M. musculus*, así como las predicciones 2 de las especies de primates y las únicas predicciones obtenidas de *M. musculus* y *R. norvegicus*.

El análisis de los sitios de unión de factores de transcripción en la región analizada, reveló nuevamente patrones similares para todas las especies, especialmente en primates. Se observan áreas donde la concentración de sitios de unión es mayor que en otras, lo que es compatible con una estructura modular como la utilizada en algunos programas bioinformáticos de predicción (Frech *et al.*, 1997). En el caso de la región considerada como núcleo del promotor se encontraron 4 sitios de unión específicos de humano: un sitio de unión al dominio POU del factor de unión a octámeros I (OCT-1), dos sitios de unión al complejo heterodimérico E2F4/DP1 y un sitio de unión al producto del gen de expresión temprana *Egr-I/Krox-24/NGFI-A*. OCT-1 se expresa ampliamente en gran número de tejidos adultos y está relacionado con la regulación de múltiples procesos, incluido el desarrollo del sistema nervioso. Este factor, además, tiene un papel crítico en el reclutamiento de complejos para el inicio de la transcripción en algunos promotores que carecen de elemento TATA, como parece ser el caso del gen *FOXP2* (revisión en Phillips y Luisi, 2000). El complejo heterodimérico E2F4/DP1 participa en la regulación del ciclo celular (revisión en Kusek *et al.*, 2000). De especial interés es el gen de expresión temprana *Egr-I/Krox-24/NGFI-A*, cuya expresión en sistema nervioso parece estar relacionada con plasticidad neuronal y procesos de memoria a largo plazo (Jones *et al.*, 2001; Frankland *et al.*, 2004).

En cuanto a los sitios de unión presentes en todas las especies, destaca la presencia de 4 sitios de unión a la Proteína con dedos de zinc 219 Kruppel-like, 5 sitios de unión al Supresor del Tumor de Wilms, 4 sitios para el Factor nuclear respiratorio 1 y 3 sitios para la proteína 1 de unión al elemento de respuesta Ras. Es relevante el número de sitios de unión a diferentes factores de transcripción que contienen dedos de zinc (MZ1, ZNF219, ZBP89, OCT-1, ZF9 (en tabla R42).

Nuestros resultados son compatibles con los obtenidos por Bruce y Margolis, quienes analizaron la región correspondiente a un fragmento de 2Kb, comprendiendo 1500 pb aguas arriba de su estima del sitio +1 del inicio de la transcripción y 500 pb aguas abajo del mismo (Bruce y Margolis, 2002). La diferencia en el número de dianas, probablemente sea debida a que en nuestro trabajo únicamente analizamos en detalle la región correspondiente a lo que proponemos como núcleo de promotor.

Es importante tener en cuenta que los criterios utilizados para el análisis de los sitios de unión a factores de transcripción en nuestro estudio han sido los más restrictivos posibles, de forma que únicamente se tuvieran en cuenta los resultados más fiables.

Esto ha hecho que sin duda, se haya perdido mucha información, así como sitios de unión a factores de transcripción funcionalmente activos y no considerados en nuestros resultados. Curiosamente, en la región altamente conservada entre las posiciones -749 y -537 la concentración de sitios de unión identificados con los criterios más restrictivos es muy pequeña. Estos sitios de unión son los siguientes: Elemento Inducible por Barbitúricos, Factor con homeodominio Nkx, Represor Transcripcional Proteína 1 de maduración inducida por linfocitos B, sitios de unión no palindrómicos del Factor Nuclear I, sitios de unión a homodímeros de proteínas Maf, y Proteína con dedos de zinc asociada a Myc.

Los árboles filogenéticos obtenidos por parsimonia y por máxima verosimilitud son compatibles con la filogenia aceptada para las especies analizadas (Purvis, 1995; Goodman *et al.*, 1998; Glazko y Nei, 2003). *Lagothrix lagotricha* y *Saguinus labiatus*, cuyo linaje se separó del resto hace aproximadamente 35 millones de años, constituyen las especies más alejadas. El linaje de las *Macaca sp.* se muestra separado del que daría lugar a la especie humana hace 25 ma. Por último, la especie más cercana al hombre corresponde a *Pan troglodytes* seguida de *Gorilla gorilla*, como corresponde según sus tiempos de divergencia estimados de 6 y 9 millones de años aproximadamente.

El resultado más relevante obtenido de los árboles filogenéticos es la diferente tasa evolutiva de los fragmentos analizados. Estos resultados están de acuerdo con los resultados obtenidos en el análisis de conservación. Mediante ensombrecimiento filogenético se encontró una serie de regiones en las que la tasa de cambio era sensiblemente menor que la de las regiones contiguas. No es de extrañar así, que se haya obtenido diferente tasa evolutiva al considerar el fragmento completo, que contiene esas regiones conservadas, respecto a considerar fragmentos como las predicciones, que no las contienen.

Al obtener el árbol correspondiente al análisis de la secuencia completa de 6.4 kilobases, en primer lugar, la rama *Homo|Pan|Gorilla* era algo más corta que la correspondiente a la familia de *Macaca|Erythrocebus* y *Lagothrix|Saguinus*, indicando una tasa de evolución menor. Además, las ramas correspondientes a *H. sapiens*, *P. troglodytes* y *G. gorilla* tenían más o menos la misma longitud, indicando también una tasa de evolución similar.

Cuando se utilizó como partida la predicción 1, comprendida entre las posiciones -1480 y -1157 respecto al inicio de la transcripción, se observó un efecto diferente. La rama correspondiente a *Homo|Pan|Gorilla* es sensiblemente más larga, indicando una evolución más rápida en estos primates. Dentro de esta rama se observa también una aceleración en el linaje correspondiente a *H. sapiens*.

Sin embargo, cuando se utiliza la predicción 2, correspondiente al fragmento comprendido entre las posiciones -184 hasta +508, la predicción 3 (+1173, +1444), la región promotora obtenida de *El Dorado* (-500, +102) o la que proponemos como núcleo del promotor (-749, +507) se observa el efecto contrario. Para estos fragmentos, la rama *Homo|Pan|Gorilla* es más corta y dentro de esta, se observa que el linaje correspondiente a *H.sapiens* evolucionó más lentamente que el de *P.troglodytes*.

Así, en resumen, la región más cercana al inicio de la transcripción parece haber sido sometida a mayores restricciones evolutivas, especialmente en el linaje humano. Estos resultados son compatibles con los observados por Taylor y colaboradores, quienes encontraron que la tasa de cambio aumenta al alejarse del inicio de transcripción y que

Discusión

las tasas evolutivas pueden cambiar entre linajes, encontrándose mayores tasas en el linaje de primates, en comparación a otros mamíferos (Taylor *et al.*, 2006).

II.2 Caracterización funcional del promotor del gen *FOXP2*

El estudio comparativo nos permitió identificar una región como posible núcleo del promotor del gen *FOXP2*. Para contrastar nuestra hipótesis se llevó a cabo un estudio funcional para el que se prepararon tres construcciones de tamaño decreciente, que contenían el propuesto núcleo del promotor, controlando la expresión del gen que codifica para la Fosfatasa Alcalina Secretada (SEAP) y se compararon los niveles de expresión obtenidos con el correspondiente a una construcción sin secuencias promotoras. Puesto que los resultados obtenidos en las células H23 procedentes de una línea celular humana fueron más robustos que los obtenidos en las células CHO-K1, de ovario de hámster, nos referiremos únicamente a ellos.

No se obtuvo un incremento significativo de la expresión del gen que codifica para la SEAP, para los fragmentos analizados respecto de la construcción sin promotor. Esto apunta a la presencia de elementos represores en algún punto de las secuencias integradas en el vector o bien a la existencia de otros elementos distales necesarios para el inicio de la transcripción.

De la comparación entre las construcciones analizadas, en los experimentos realizados se observa que la construcción de mayor tamaño (-879, +1677), que incluye la región muy conservada con un único cambio entre las secuencias de primates analizadas, es la que da lugar a menores niveles de expresión en todos los casos al compararse con la construcción de menor tamaño (-191, +1677). Esto nos lleva a pensar que en la región presente en el fragmento mayor existen elementos represores.

Durante el análisis de nuestros resultados, se publicó un estudio en el que se realizó un experimento similar (Schroeder y Myers, 2008). Estos autores detectaron la presencia de 4 sitios de inicio de la transcripción alternativos, localizados en los exones s1, 1, 1b y 2 en diferentes tipos celulares y llevaron a cabo ensayos funcionales para las posibles regiones promotoras adyacentes a los cuatro sitios de inicio de la transcripción.

En el caso del exón s1, llevaron a cabo un análisis funcional de expresión en células tumorales procedentes de cerebelo humano, con 11 construcciones diferentes, que incluían fragmentos contenidos en el intervalo -780, +309. Para la mayoría de estas construcciones (figura D2) sí se detectó un aumento en la expresión respecto la construcción vacía. En los ensayos funcionales del trabajo de Schroeder y Myers, correspondientes a los posibles promotores adyacentes a los sitios de inicio de transcripción localizados en los exones 1 y 1b no se obtuvo diferencias entre las construcciones utilizadas.

Aunque las construcciones utilizadas en el estudio de Schroeder y Myers son diferentes a las utilizadas en nuestros experimentos, se puede hacer una comparación aproximada en la que equiparar nuestras construcciones de (-879, +1677), (-408, +1677) y (-191, +1677), con las obtenidas por dicho grupo de investigación: (-780, +309), (-425, +309) y (-242, +309). Así, en estas tres construcciones, aunque estos autores no encuentran diferencias significativas entre ellas, sí parece encontrarse la misma tendencia observada en nuestros resultados, es decir un aumento de la expresión en el fragmento más pequeño. No obstante, a diferencia de lo obtenido en

nuestro estudio, y como ya se ha mencionado anteriormente, se observó un aumento de la expresión en todas las construcciones respecto del vector vacío. Existe una diferencia notable en el extremo 3' de las construcciones utilizadas en nuestro estudio y las del estudio de Schroeder y Myers, a la cual pensamos se debe la menor expresión de nuestras construcciones. En este fragmento destaca la presencia de 2 cajas CG para factores SP1/GC, 6 sitios de unión para la proteína C inducida por el factor de crecimiento nervioso o EGR y factores relacionados, 5 sitios de unión al factor NRF1, 5 sitios de unión a factores con dedos de zinc mieloides y 3 sitios para la familia de proteínas del factor inducible por hipoxia bHLH/PAS.

En cuanto al extremo 5', aunque no es tan notable como en el anterior caso, también hay una diferencia en el tamaño del fragmento de algo menos de 100 nucleótidos. En esta región se localizan sitios de unión a los factores: Proteína de unión a X-box, Sox 5, factor de transcripción con dominio POZ y dedos de zinc, EF2, Factor nuclear Y y v-Myb.

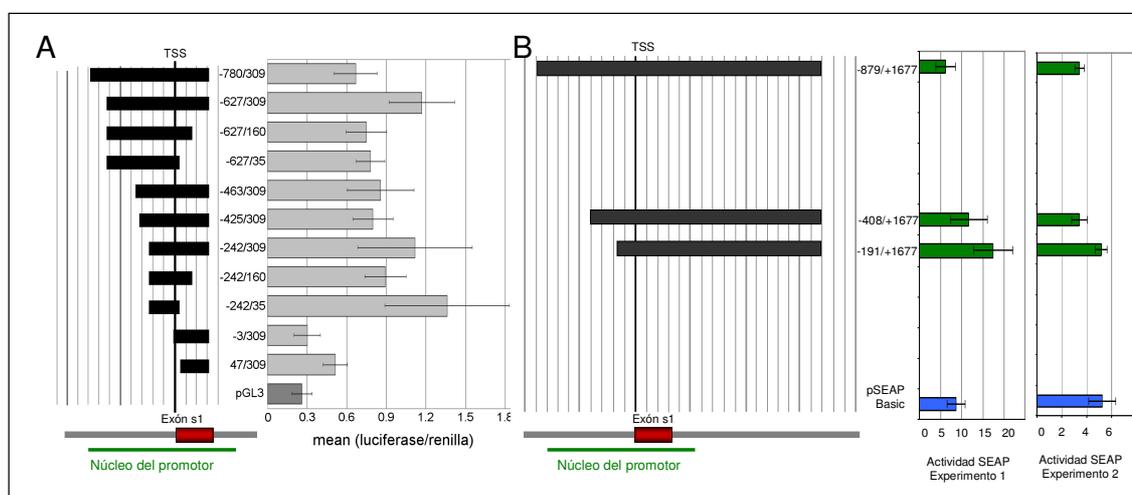


Figura D2. Comparación de los ensayos de expresión obtenidos por Schroeder y Myers (A) con los correspondientes a este trabajo (B). Las posiciones de los fragmentos analizados respecto a la posición +1 del exón s1 se muestran a la izquierda. Los resultados de los ensayos de transfección se muestran a la derecha. La localización del que proponemos como núcleo del promotor se muestra en color verde en el esquema de los fragmentos analizados. TSS se refiere al sitio de inicio de la transcripción.

Las diferencias ocasionadas por el extremo 3' y o el extremo 5', pueden ser debidas a dos razones: o bien por la presencia de sitios de unión a factores de transcripción que puedan actuar como represores, o bien, los elementos activadores presentes en el extremo 5' respecto al primer exón del gen *FOXP2*, quedan demasiado distales del inicio de la expresión del gen de la Fosfatasa Alcalina Secretada.

Otra diferencia importante es el tipo de células utilizadas. En nuestro caso utilizamos células tumorales de pulmón, mientras que Shroeder y Myers utilizaron células tumorales de cerebelo. La expresión del gen *FOXP2* es elevada en cerebelo (Lai *et al.*, 2003) y a nivel general la expresión en cerebro es mayor que la obtenida en pulmón (Lai *et al.*, 2001). En este sentido una menor expresión generalizada en las células utilizadas en nuestros experimentos respecto a las de Shroeder y Myers podría explicarse por la presencia de elementos reguladores de cada tejido. Ensayos sobre qué genes son regulados por *FOXP2* indican la presencia de diferentes dianas en pulmón y cerebro por lo que no es descartable que el mismo gen *FOXP2* tenga distinta regulación en ambos tejidos (Spiteri *et al.*, 2007; Vernes *et al.*, 2007).

Discusión

Nuestros resultados, junto con los obtenidos por Bruce y Margolis, Schroeder y Myers indican que la regulación de la expresión del gen *FOXP2* es compleja y probablemente intervengan más factores de los inicialmente planteados. Schroeder y Myers, proponen que los sitios de inicio de la transcripción localizados en los exones s1 y 2 contribuyen a la obtención de niveles basales de la proteína en una amplia variedad de tejidos adultos humanos, y que en determinados estadios del desarrollo o condiciones celulares estos niveles basales se aumentarían a través de la transcripción iniciada en los exones 1 y 1b. La ausencia de resultados positivos para los ensayos funcionales para las regiones analizadas de los exones 1 y 1b pueden ser debidas a la existencia de elementos reguladores distales.

El único de los sitios de inicio de transcripción que contiene una isla CG es el que se localiza en el exón s1, donde se observa mayor incremento de la expresión en estudios funcionales (Schroeder y Myers, 2008) por lo que probablemente apoya nuestra consideración inicial de que constituya el promotor principal.

III. Variaciones funcionales

Tal y como mencionamos en apartados anteriores, los estudios epigenéticos constituyen una de las aproximaciones que está cobrando más interés en el estudio de las enfermedades psiquiátricas, especialmente en el campo de la esquizofrenia. La participación de modificaciones más allá de la secuencia ofrece explicación a algunas características de esta enfermedad que no pueden ser explicadas bajo el modelo genético, como la discordancia entre gemelos monocigóticos, la variabilidad fenotípica o la presencia de remisiones y recaídas (Petronis, 2004). En la tabla D2 se muestran algunas de estas características, así como la explicación correspondiente al modelo genético y al modelo epigenético. Algunas de las explicaciones genéticas y epigenéticas se han visto reforzadas por los primeros estudios donde se estima la variabilidad de la metilación del DNA en diferentes tipos de individuos. Por ejemplo, Fraga y colaboradores observaron que las diferencias epigenéticas en contenido y distribución de 5-metil-citosinas y acetilación de histonas aumentan con la edad en gemelos monocigóticos (Fraga *et al.*, 2005). En un estudio posterior, a nivel genómico, Eckhardt y colaboradores no fueron capaces de reproducir esos resultados en muestras agrupadas, debido probablemente al método de análisis utilizado, aunque sí observaron regiones genómicas con diferente metilación en los tipos celulares analizados, tanto en promotores como en regiones intergénicas conservadas (Eckhardt *et al.*, 2006).

La metilación constituye la modificación epigenética más estable en la modulación de la plasticidad transcripcional de los genomas de mamíferos. En estos, la metilación se produce en dinucleótidos CG, cuya frecuencia se encuentra reducida en el genoma, con excepción de las regiones donde se localizan las islas CG que son regiones ricas en CG presentes en el 50-70% de los promotores de genes humanos (Bird, 1986). La metilación dentro de este tipo de regiones se asocia al silenciamiento génico y está involucrada en la impronta genómica e inactivación del cromosoma X, así como en patologías asociadas a enfermedades del desarrollo y cáncer (Razin y Cedar 1994; Jones y Takai, 2001; Li, 2002; Jaenisch y Bird, 2003; Shen *et al.*, 2007). Inicialmente, se consideró que la metilación del DNA estaba asociada al control de la expresión de genes específica de tejidos en células mamíferos, pero se observó que la mayoría de islas CG localizadas en promotores de genes se mantienen no metiladas

independientemente de la expresión específica de tejido (Bird *et al.*, 1987). Esta idea se ha mantenido durante mucho tiempo, aunque posteriores estudios indican que existe aproximadamente un 4% de promotores de genes que se encuentran completamente metilados en tejido normal, en los que la metilación está asociada al silenciamiento génico, de forma que una inhibición de la misma reactiva la expresión del gen (Shen *et al.*, 2007).

Característica	Modelo Genético	Modelo epigenético
Mecanismo molecular primario de la enfermedad	Predisposición a través de la variación en la secuencia de DNA	Desregulación epigenética
Discordancia entre gemelos monocigóticos	Efectos de un ambiente no compartido	Modificaciones epigenéticas diferenciales en gemelos monocigóticos, principalmente debido a factores estocásticos
Efecto del origen parental	Sin explicación	Modificación específica de origen parental del DNA e histonas (impronta genómica)
Efectos medioambientales	Interacción con el genoma que puede introducir cambios en la expresión génica	El estado epigenético de un gen o genes puede ser modificado por factores medioambientales para controlar la expresión del gen.
Efecto del sexo	Ligamiento a cromosomas sexuales (si no se encuentra ligamiento en cromosomas sexuales no hay explicación).	Diferentes efectos epigenéticos de andrógenos y estrógenos.
Variabilidad fenotípica	Resultado de diferentes combinaciones de genes que confieren vulnerabilidad, conjuntamente con factores ambientales.	Resultado de cambios epigenéticos inducidos por eventos estocásticos, ambientales y del desarrollo.
Remisiones y recaídas	Sin explicación	Fluctuaciones en la regulación metaestable epigenética; cambios epigenéticos relacionados con la edad; efectos epigenéticos del tratamiento.
Presencia de casos familiares y esporádicos	Genes con diferente efecto (para casos familiares y esporádicos respectivamente). Sin embargo los estudios de ligamiento no han detectado todavía genes con gran efecto.	Metaestabilidad epigenética durante la meiosis.
Elevada frecuencia a pesar de la presión evolutiva	Ventajas evolutivas de genes que predisponen a la esquizofrenia (en portadores no afectados)	Tasa de epimutación <i>de novo</i> elevada durante la meiosis
Inconsistencia en los datos de ligamiento y asociación	Heterogeneidad genética de la enfermedad. Falta de poder estadístico para detectar efectos aditivos pequeños. Complejidad evolutiva en la formación de los haplotipos.	La variación en la secuencia de DNA no tiene porqué jugar un papel etiológico primario en la enfermedad.

Tabla D2. Comparación del modelo genético (basado en secuencia del DNA) y epigenético de la esquizofrenia. Tomado de Petronis, 2004

Hay pocos datos en los que se haga referencia a la presencia de diferencias en metilación en psicosis. El caso más claro lo constituye el reciente estudio de Mills y colaboradores, quienes encontraron evidencias de diferencias en metilación asociadas a psicosis en *loci* relacionados con procesos de neurotransmisión y desarrollo neuronal (Mill *et al.*, 2008). En cuanto a casos concretos, el más conocido es el del promotor del gen que codifica para la reelina, para el cual se ha observado hipermetilación asociada a niveles más bajos de mRNA en cerebros *postmortem* (Chen *et al.*, 2002; Grayson *et al.*, 2005).

En la región 5' respecto al inicio de la transcripción del gen *FOXP2* se localiza una isla CG de gran longitud y que además muestra un elevado grado de conservación en algunos tramos, como ya se ha comentado en apartados anteriores (Figura R34 del

Discusión

apartado de Resultados). Esto nos llevó a pensar si mecanismos epigenéticos podían participar en la regulación del gen y si estos mecanismos podían estar alterados entre controles y pacientes con esquizofrenia. Se planteó por tanto un análisis exploratorio para determinar los patrones de metilación de la región en muestras obtenidas de diferentes áreas de la corteza cerebral procedentes de cerebros *postmortem*.

Este análisis de metilación se realizó mediante dos métodos: el análisis mediante enzimas de restricción sensibles a metilación y el análisis mediante modificación con bisulfito.

El análisis de metilación con **enzimas de restricción sensibles a metilación** permitió obtener una idea aproximada del grado de metilación de las regiones analizadas. En este sentido el enzima *McrBC* resultó más informativo a nivel general, indicando posibles diferencias entre la circunvolución temporal y la circunvolución del cíngulo. La utilización de este tipo de enzimas es un método rápido para determinar el estado de metilación, sin embargo, presenta algunas desventajas como la dependencia a las dianas de enzimas de restricción sensibles a metilación existentes en una región determinada o la posibilidad de digestiones incompletas.

El análisis de metilación mediante la **modificación con bisulfito**, demostró ser un método mucho más sensible, puesto que al poder determinar la presencia de metilación en cada dinucleótido CG se pueden obtener patrones específicos de metilación en las muestras analizadas.

En este caso se han obtenido diferencias apreciables entre las dos regiones analizadas, CG1, localizada en 5' respecto al inicio de la transcripción localizado en el exón s1 y CG2, localizada en el intrón s1. La primera de ellas mostró escasa metilación, mientras que en el caso de la segunda, el grado de metilación, reflejado en un mismo patrón en las diferentes muestras analizadas, fue mayor.

Los resultados más relevantes al comparar pacientes con controles se encuentran también al analizar la región CG2. Se ha encontrado evidencias de mayor metilación en el área cerebral de la circunvolución del parahipocampo, tanto para pacientes como para controles, aunque la relación de metilación es inversa para ambos grupos al considerar los diferentes hemisferios. En el hemisferio izquierdo se observa mayor grado de metilación en pacientes que en controles, sin embargo, en el hemisferio derecho, se observa mayor grado de metilación en controles.

La presencia de diferentes patrones de metilación en las diferentes muestras puede ser debida a la existencia de distintos tipos celulares con diferente grado de metilación. Estas diferencias en la metilación según el tipo de célula se han observado en otros estudios (Eckhardt *et al.*, 2006). En nuestro caso esto se ve apoyado por el hecho de que cuando se analiza DNA extraído de sangre periférica con un mismo tipo celular se obtiene un patrón más homogéneo.

En la región CG2 se encuentra el tramo con repeticiones CGG estudiado por análisis de fragmentos, lo que la convierte en una zona con muy elevado contenido en CG. El fragmento CG2 incluye dos cajas CG de unión a factores CG, 5 sitios de unión al factor nuclear de respiración (NRF1), dos sitios de unión a la familia de proteínas del factor inducible por hipoxia bHLH/PAS y 3 sitios de unión para la proteína C inducida por el factor de crecimiento nervioso o EGR y factores relacionados. La metilación de la región podría impedir la unión de estos factores.

La principal limitación de nuestro estudio es el reducido número de muestras utilizadas. No obstante, los resultados obtenidos apuntan de forma consistente hacia la existencia

de diferentes patrones de metilación. Otra limitación es, sin duda, el hecho de que no se ha podido determinar el estado de metilación de la isla CG al completo, sino únicamente dos fragmentos. El DNA transformado con bisulfito es difícil de amplificar y aunque se diseñaron cebadores para cubrir la región, la mayoría de ellos no funcionaron. Aún así se estudiaron dos regiones representativas, una, del contexto de metilación aguas arriba del inicio de la transcripción y la otra en el primer intrón del gen, en el extremo aguas abajo de la isla CG.

Como se ha mencionado anteriormente, la metilación de islas CG asociadas a promotores de genes se encuentra relacionada con la represión de la expresión para determinados genes. Con el objetivo de intentar correlacionar datos de metilación con datos de expresión, así como para determinar diferencias de expresión del gen en pacientes y controles o en distintas áreas cerebrales, se llevó a cabo una serie de experimentos de PCR cuantitativa. En estos experimentos se analizaron los niveles de RNAm del gen en diferentes áreas de la corteza cerebral para las cuales se ha observado variaciones estructurales o funcionales en esquizofrenia.

Estudios postmortem y de neuroimagen han encontrado una disminución en la cantidad de materia gris de la circunvolución del parahipocampo en el hemisferio izquierdo en pacientes esquizofrénicos (Razi *et al.*, 1999; McDonald *et al.*, 2000; Paillere-Martinot *et al.*, 2001; Sigmundsson *et al.*, 2001; Hulshoff Pol *et al.*, 2004; metaanálisis en Honea *et al.*, 2005 y Glahn *et al.*, 2008). ¿Podría estar relacionado el aumento del grado de metilación del promotor de *FOXP2* observado en el análisis de los patrones de metilación con esta disminución a través de un descenso en la expresión del gen?. Para otra de las áreas más replicadas por su implicación en esquizofrenia, la circunvolución temporal superior (Honea *et al.*, 2005; Wilke *et al.*, 2001; Sigmundsson *et al.*, 2001; Hulshoff Pol *et al.*, 2004; Takahashi *et al.*, 2006) no se habían encontrado diferencias pronunciadas en los patrones de metilación obtenidos, aunque no era descartable la presencia de diferencias en expresión.

Los análisis de expresión a partir de mezclas obtenidas de diferentes individuos mostraron tendencia a mayor expresión en pacientes que en controles en las áreas correspondientes a las circunvoluciones temporal y del parahipocampo derechas y circunvolución del parahipocampo izquierda y menor expresión en la circunvolución temporal superior izquierda. Dado el número de muestras utilizadas, estos resultados deben ser considerados como preliminares.

Lamentablemente, los datos de los que disponemos no permiten establecer una relación directa entre los patrones de metilación observados y los resultados de los análisis de expresión. El reducido número de muestras de los análisis de metilación junto con la elevada variabilidad interindividual observada a nivel de expresión y que podría también tener lugar a efectos de patrones de metilación, impide poder llevar a cabo cualquier tipo de correlación. En este sentido, sería conveniente repetir los análisis con un mayor número de muestras. Sin embargo, se plantea otra limitación puesto que las muestras deben ser obtenidas de un material tan limitado y en ocasiones poco disponible como es el caso de los cerebros *postmortem*.

Otra consideración que no se ha tenido en cuenta es el origen de las muestras con las que se ha trabajado, tanto en los análisis de metilación como especialmente en los de expresión. A diferencia de otras enfermedades en las que se puede obtener muestras de tejido fresco para su estudio, en el caso de las enfermedades cerebrales, debido a la inaccesibilidad del cerebro humano y la no disponibilidad de biopsias, se suele recurrir a la utilización de muestras de tejido procedentes de cerebros *postmortem*.

Discusión

Existen por tanto, una serie de factores no controlables, basados en las condiciones pre y *postmortem*, que podrían afectar a la calidad de la muestra. Así, la medida de la expresión génica en tejido *postmortem* puede verse dificultada por diferentes factores como la edad, género, pH, intervalo *postmortem*, calidad del tejido, medicación, tabaco, causa de la muerte, procesos agonales y otras variables. De estos factores, los que mayor influencia parecen tener, como demuestran diferentes estudios, son el pH y la calidad del RNA (Li *et al.*, 2004; Lipska *et al.*, 2006; Mexal *et al.*, 2006; Stan *et al.*, 2006).

La presente tesis se ha centrado en la evaluación a través de diferentes aproximaciones la implicación del gen *FOXP2*, gen sometido a selección positiva en el linaje humano y relacionado directamente con una alteración de uno de los rasgos más característicos de la especie humana, el lenguaje, en la vulnerabilidad a la esquizofrenia.

El estudio de asociación caso-control no ha permitido establecer una implicación consistente entre las variantes estructurales analizadas (SNPs y posibles expansiones de trinucleótidos) con las alucinaciones auditivas como fenotipo alternativo a la esquizofrenia. No obstante, la participación del gen *FOXP2* en la vulnerabilidad a las AA, como componente referido al lenguaje no puede descartarse por completo, ya que el SNP rs2396753 mostró una tendencia hacia la significación y el SNP2253478 mostró diferencias significativas para el ítem de Pobreza del Lenguaje de la Escala Manchester.

En cualquier caso, este estudio permite mostrar las ventajas de la utilización de las AA como fenotipo alternativo a la esquizofrenia, ya que los resultados obtenidos en las comparaciones utilizando este síntoma refuerzan tendencias observadas en la comparación respecto a los pacientes en general.

El análisis de la región promotora nos ha permitido valorar que la regulación de la expresión del gen debe ser más compleja de lo que inicialmente se esperaba.

El análisis evolutivo muestra que diferentes tramos de la región promotora analizada han evolucionado diferencialmente, encontrándose una región altamente conservada, que curiosamente no parece contener una elevada concentración de sitios de unión a factores de transcripción, como sería lo esperado dada su posible importancia funcional. Esta secuencia se localiza inmediatamente aguas arriba del potencial promotor extraído de las bases de datos mediante herramientas bioinformáticas. Bajo la hipótesis de su pertenencia al núcleo central del promotor, la evaluación funcional de esta región conservada indica que posiblemente contenga elementos represores de la transcripción, al menos en el tipo celular testado, esto es, células procedentes del pulmón ya que se obtienen mayores niveles de expresión en su ausencia.

Dentro de la hipótesis epigenética de la esquizofrenia, se llevó a cabo un análisis de los patrones de metilación en dos fragmentos incluidos en una isla CG adyacente al primer exón, no traducido, del gen en tejidos *postmortem* de pacientes esquizofrénicos y controles. En primer lugar se obtuvo una clara diferencia en el grado de metilación entre el fragmento analizado localizado aguas arriba del exón respecto al localizado aguas abajo, ambos en extremos de la isla CG. Comparando el grado de metilación en diferentes áreas cerebrales en pacientes con esquizofrenia y controles, se observó que en la circunvolución del parahipocampo el grado de metilación es mayor que en las otras áreas analizadas y además hay indicios de diferencias entre ambos grupos analizados en ambos hemisferios. Desafortunadamente, los análisis de expresión encaminados a determinar si el grado de metilación estaba correlacionado con los niveles de expresión no permitieron llegar a resultados concluyentes al respecto. Con todo esto, sería recomendable ampliar el número de muestras utilizadas para confirmar las diferencias que se observaron en nuestra reducida muestra e intentar correlacionar datos de metilación con datos de expresión, así como ampliar las áreas cerebrales analizadas, poniendo especial énfasis en áreas del lenguaje.

Con todos los resultados obtenidos, no podemos descartar que el gen *FOXP2* participe en el desarrollo de las alucinaciones auditivas, aunque no se han encontrado evidencias directas de ello.

Discusión

Conclusiones



1. El estudio de asociación caso-control muestra la ausencia de diferencias significativas en las frecuencias alélicas y genotípicas de los polimorfismos del gen *FOXP2* analizados, en relación con el desarrollo de alucinaciones auditivas en pacientes con esquizofrenia, cuando se aplican las correcciones estadísticas más conservativas. No obstante, se observa una tendencia hacia mayor frecuencia del alelo C para el polimorfismo rs2396753 en pacientes con alucinaciones auditivas. Esta tendencia, dada la severa corrección estadística aplicada, podría constituir un resultado positivo a tener en cuenta.
2. La utilización de las alucinaciones auditivas como fenotipo alternativo de la esquizofrenia permite obtener resultados más consistentes que la utilización de un fenotipo más amplio, ya que al comparar el grupo control frente a pacientes con este síntoma se refuerzan tendencias observadas en la comparación del grupo de controles frente a los pacientes con esquizofrenia.
3. La relación del gen *FOXP2* con la esquizofrenia se ve reforzada por la asociación entre un polimorfismo del gen *FOXP2* con el ítem de Pobreza del Lenguaje de la escala Manchester.
4. Las regiones con repeticiones de trinucleótidos del gen *FOXP2* analizadas son altamente estables por lo que se descarta que su expansión inestable constituya un mecanismo por el cual se alteren los niveles de expresión o la estructura del producto obtenido del gen *FOXP2*.
5. El análisis evolutivo de la región reguladora localizada en el extremo 5' del gen *FOXP2* permitió detectar una región muy conservada aguas arriba del primer exón del gen (posiciones -749, -537) que forma parte de lo que proponemos como núcleo del promotor del gen.
6. El análisis funcional de la región conservada indicó que no constituye un elemento *enhancer* de la región promotora, si no más bien un elemento represor.
7. No se han encontrado diferencias consistentes en los niveles de expresión ni en los patrones de metilación de diferentes áreas cerebrales entre pacientes y controles.
8. Se ha encontrado que dentro de la isla CG adyacente al primer exón del gen, hay una variabilidad en el grado de metilación del extremo 3' respecto al extremo 5', encontrándose mayor grado de metilación en el primero.
9. El hecho de no encontrar diferencias significativas en las frecuencias genotípicas, alélicas o haplotípicas en el estudio de asociación caso-control llevado a cabo con el gen *HAR1A* indica que no parece estar relacionado con la vulnerabilidad a la esquizofrenia, aunque podría estar relacionado con la susceptibilidad a padecer alucinaciones auditivas dentro del contexto psicótico, al obtenerse diferencias globales en la comparación de haplotipos entre pacientes alucinadores y pacientes sin alucinaciones.

Bibliografía



Aguilar,E.J., Sanjuan,J., Garcia-Marti,G., Lull,J.J., & Robles,M. MR and genetics in schizophrenia: focus on auditory hallucinations. *Eur. J. Radiol.* **67**, 434-439 (2008).

Alarcon,M., Cantor,R.M., Liu,J., Gilliam,T.C., & Geschwind,D.H. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am. J. Hum. Genet.* **70**, 60-71 (2002).

Alarcon,M., Abrahams,B.S., Stone,J.L., Duvall,J.A., Perederiy,J.V., Bomar,J.M., Sebat,J., Wigler,M., Martin,C.L., Ledbetter,D.H., Nelson,S.F., Cantor,R.M., & Geschwind,D.H. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82**, 150-159 (2008).

Alcock,K.J., Passingham,R.E., Watkins,K.E., & Vargha-Khadem,F. Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain Lang* **75**, 17-33 (2000).

Alcock,K.J., Passingham,R.E., Watkins,K., & Vargha-Khadem,F. Pitch and timing abilities in inherited speech and language impairment. *Brain Lang* **75**, 34-46 (2000).

Aleman, A. and Laroi, F. Hallucinations: The Science of Idiosyncratic Perception. 2008. American Psychological Association. Washington, DC.

Ref Type: Generic

Allen,N.C., Bagade,S., McQueen,M.B., Ioannidis,J.P., Kavvoura,F.K., Khoury,M.J., Tanzi,R.E., & Bertram,L. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat. Genet.* **40**, 827-834 (2008).

Allen,P., Laroi,F., McGuire,P.K., & Aleman,A. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neurosci. Biobehav. Rev.* **32**, 175-191 (2008).

American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders.* (2002).

Andreasen,N.C. & Flaum,M. Schizophrenia: the characteristic symptoms. *Schizophr. Bull.* **17**, 27-49 (1991).

Arbib,M.A. Other faces in the mirror: a perspective on schizophrenia. *World Psychiatry* **6**, 75-78 (2007).

Arinami,T., Ohtsuki,T., Ishiguro,H., Ujike,H., Tanaka,Y., Morita,Y., Mineta,M., Takeichi,M., Yamada,S., Imamura,A., Ohara,K., Shibuya,H., Ohara,K., Suzuki,Y., Muratake,T., Kaneko,N., Someya,T., Inada,T., Yoshikawa,T., Toyota,T., Yamada,K., Kojima,T., Takahashi,S., Osamu,O., Shinkai,T., Nakamura,M., Fukuzako,H., Hashiguchi,T., Niwa,S.I., Ueno,T., Tachikawa,H., Hori,T., Asada,T., Nanko,S., Kunugi,H., Hashimoto,R., Ozaki,N., Iwata,N., Harano,M., Arai,H., Ohnuma,T., Kusumi,I., Koyama,T., Yoneda,H., Fukumaki,Y., Shibata,H., Kaneko,S., Higuchi,H., Yasui-Furukori,N., Numachi,Y., Itokawa,M., & Okazaki,Y. Genomewide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility loci on chromosomes 1p, 14q, and 20p. *Am. J. Hum. Genet.* **77**, 937-944 (2005).

Arking,D.E., Cutler,D.J., Brune,C.W., Teslovich,T.M., West,K., Ikeda,M., Rea,A., Guy,M., Lin,S., Cook,E.H., & Chakravarti,A. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am. J. Hum. Genet.* **82**, 160-164 (2008).

Ashley-Koch,A., Wolpert,C.M., Menold,M.M., Zaeem,L., Basu,S., Donnelly,S.L., Ravan,S.A., Powell,C.M., Qumsiyeh,M.B., Aylsworth,A.S., Vance,J.M., Gilbert,J.R., Wright,H.H.,

Bibliografía

- Abramson,R.K., DeLong,G.R., Cuccaro,M.L., & Pericak-Vance,M.A. Genetic studies of autistic disorder and chromosome 7. *Genomics* **61**, 227-236 (1999).
- Badner,J.A. & Gershon,E.S. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol. Psychiatry* **7**, 405-411 (2002).
- Bakewell,M.A., Shi,P., & Zhang,J. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl. Acad. Sci. U. S. A* **104**, 7489-7494 (2007).
- Bakwin,H. Reading disability in twins. *Dev. Med. Child Neurol.* **15**, 184-187 (1973).
- Barrett,J.C., Fry,B., Maller,J., & Daly,M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* **21**, 263-265 (2005).
- Barta,P.E., Pearlson,G.D., Powers,R.E., Richards,S.S., & Tune,L.E. Auditory hallucinations and smaller superior temporal gyral volume in schizophrenia. *Am. J. Psychiatry* **147**, 1457-1462 (1990).
- Bartlett,C.W., Flax,J.F., Logue,M.W., Smith,B.J., Vieland,V.J., Tallal,P., & Brzustowicz,L.M. Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. *Hum. Hered.* **57**, 10-20 (2004).
- Bearden,C.E. & Freimer,N.B. Endophenotypes for psychiatric disorders: ready for primetime? *Trends Genet.* **22**, 306-313 (2006).
- Becker,K.G. The common variants/multiple disease hypothesis of common complex genetic disorders. *Med. Hypotheses* **62**, 309-317 (2004).
- Belton,E., Salmond,C.H., Watkins,K.E., Vargha-Khadem,F., & Gadian,D.G. Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Hum. Brain Mapp.* **18**, 194-200 (2003).
- Bentall R.P. Madness explained: Psychosis and human nature. (Penguin Books. London,2003).
- Bird,A.P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213 (1986).
- Bird,A.P., Taggart,M.H., Nicholls,R.D., & Higgs,D.R. Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *EMBO J.* **6**, 999-1004 (1987).
- Bleuler, E. Dementia praecox oder die Gruppe der Schizophrenien. 1911. **Leipzig: Deutike.**
- Blinc-Pesek,M. & Agius,M. Anticipation and the genetics of psychosis. *Br. J. Psychiatry* **191**, 181 (2007).
- Bly,M. Mutation in the vesicular monoamine gene, SLC18A1, associated with schizophrenia. *Schizophr. Res.* **78**, 337-338 (2005).
- Boffelli,D., McAuliffe,J., Ovcharenko,D., Lewis,K.D., Ovcharenko,I., Pachter,L., & Rubin,E.M. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**, 1391-1394 (2003).
- Bonkowsky,J.L. & Chien,C.B. Molecular cloning and developmental expression of foxP2 in zebrafish. *Dev. Dyn.* **234**, 740-746 (2005).

- Bonkowsky, J.L., Wang, X., Fujimoto, E., Lee, J.E., Chien, C.B., & Dorsky, R.I. Domain-specific regulation of foxP2 CNS expression by lef1. *BMC. Dev. Biol.* **8**, 103 (2008).
- Bookheimer, S. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* **25**, 151-188 (2002).
- Brasic, J.R. Hallucinations. *Percept. Mot. Skills* **86**, 851-877 (1998).
- Bruce, H.A. & Margolis, R.L. FOXP2: novel exons, splice variants, and CAG repeat length stability. *Hum. Genet.* **111**, 136-144 (2002).
- Bruder, G., Rabinowicz, E., Towey, J., Brown, A., Kaufmann, C.A., Amador, X., Malaspina, D., & Gorman, J.M. Smaller right ear (left hemisphere) advantage for dichotic fused words in patients with schizophrenia. *Am. J. Psychiatry* **152**, 932-935 (1995).
- Brune, M. Schizophrenia-an evolutionary enigma? *Neurosci. Biobehav. Rev.* **28**, 41-53 (2004).
- Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J., & Greenberg, M.E. Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857-868 (1999).
- Brzustowicz, L.M., Hodgkinson, K.A., Chow, E.W., Honer, W.G., & Bassett, A.S. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. *Science* **288**, 678-682 (2000).
- Burns, J. The social brain hypothesis of schizophrenia. *World Psychiatry* **5**, 77-81 (2006).
- Bustamante, C.D., Fedel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M., & Clark, A.G. Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153-1157 (2005).
- Butland, S.L., Devon, R.S., Huang, Y., Mead, C.L., Meynert, A.M., Neal, S.J., Lee, S.S., Wilkinson, A., Yang, G.S., Yuen, M.M., Hayden, M.R., Holt, R.A., Leavitt, B.R., & Ouellette, B.F. CAG-encoded polyglutamine length polymorphism in the human genome. *BMC. Genomics* **8**, 126 (2007).
- Caceres, M., Lachuer, J., Zapala, M.A., Redmond, J.C., Kudo, L., Geschwind, D.H., Lockhart, D.J., Preuss, T.M., & Barlow, C. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. U. S. A* **100**, 13030-13035 (2003).
- Cannon, T.D. & Keller, M.C. Endophenotypes in the genetic analyses of mental disorders. *Annu. Rev. Clin. Psychol.* **2**, 267-290 (2006).
- Cantor, R.M. & Geschwind, D.H. Schizophrenia: genome, interrupted. *Neuron* **58**, 165-167 (2008).
- Cardno, A.G., Rijdsdijk, F.V., Sham, P.C., Murray, R.M., & McGuffin, P. A twin study of genetic relationships between psychotic symptoms. *Am. J. Psychiatry* **159**, 539-545 (2002).
- Carlsson, P. & Mahlapuu, M. Forkhead transcription factors: key players in development and metabolism. *Dev. Biol.* **250**, 1-23 (2002).
- Carroll, S.B. Evolution at two levels: on genes and form. *PLoS. Biol.* **3**, e245 (2005).

Bibliografía

- Chang,J.B., Wang,P.N., Chen,W.T., Liu,C.Y., Hong,C.J., Lin,K.N., Liu,T.Y., Chi,C.W., & Liu,H.C. ApoE epsilon4 allele is associated with incidental hallucinations and delusions in patients with AD. *Neurology* **63**, 1105-1107 (2004).
- Chen,S.F., Chen,C.H., Chen,J.Y., Wang,Y.C., Lai,I.C., Liou,Y.J., & Liao,D.L. Support for association of the A277C single nucleotide polymorphism in human vesicular monoamine transporter 1 gene with schizophrenia. *Schizophr. Res.* **90**, 363-365 (2007).
- Chen,Y., Sharma,R.P., Costa,R.H., Costa,E., & Grayson,D.R. On the epigenetic regulation of the human reelin promoter. *Nucleic Acids Res.* **30**, 2930-2939 (2002).
- Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A., Todd,M.J., Tanenbaum,D.M., Civello,D., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X., White,T.J., Sninsky,J.J., Adams,M.D., & Cargill,M. Positive selection in the human genome inferred from human-chimp-mouse orthologous gene alignments. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 471-477 (2003).
- Close,H. & Garety,P. Cognitive assessment of voices: further developments in understanding the emotional impact of voices. *Br. J. Clin. Psychol.* **37 (Pt 2)**, 173-188 (1998).
- Conrad,K. La esquizofrenia incipiente. (Fundación Archivos de Neurobiología, Madrid; 1997).
- Coop,G., Bullaughey,K., Luca,F., & Przeworski,M. The timing of selection at the human FOXP2 gene. *Mol. Biol. Evol.* **25**, 1257-1259 (2008).
- Costa,E., Davis,J., Grayson,D.R., Guidotti,A., Pappas,G.D., & Pesold,C. Dendritic spine hypoplasticity and downregulation of reelin and GABAergic tone in schizophrenia vulnerability. *Neurobiol. Dis.* **8**, 723-742 (2001).
- Costa,E., Davis,J.M., Dong,E., Grayson,D.R., Guidotti,A., Tremolizzo,L., & Veldic,M. A GABAergic cortical deficit dominates schizophrenia pathophysiology. *Crit Rev. Neurobiol.* **16**, 1-23 (2004).
- Costello,J.F. & Plass,C. Methylation matters. *J. Med. Genet.* **38**, 285-303 (2001).
- Covington,M.A., He,C., Brown,C., Naci,L., McClain,J.T., Fjordbak,B.S., Semple,J., & Brown,J. Schizophrenia and the structure of language: the linguist's view. *Schizophr. Res.* **77**, 85-98 (2005).
- Crespi,B., Summers,K., & Dorus,S. Adaptive evolution of genes underlying schizophrenia. *Proc. Biol. Sci.* **274**, 2801-2810 (2007).
- Crichton-Browne, J. On the weight of the brain and its components parts in the insane. *Brain* **2**, 42-67. 1879.
- Crow,T.J. Is schizophrenia the price that Homo sapiens pays for language? *Schizophr. Res.* **28**, 127-141 (1997).
- Crow,T.J. Schizophrenia as the price that homo sapiens pays for language: a resolution of the central paradox in the origin of the species. *Brain Res. Brain Res. Rev.* **31**, 118-129 (2000).
- Crow,T.J. How and why genetic linkage has not solved the problem of psychosis: review and hypothesis. *Am. J. Psychiatry* **164**, 13-21 (2007).
- Crow,T.J. The 'big bang' theory of the origin of psychosis and the faculty of language. *Schizophr. Res.* **102**, 31-52 (2008).

- Damasio,A.R. Aphasia. *N. Engl. J. Med.* **326**, 531-539 (1992).
- David,A.S. Auditory hallucinations: phenomenology, neuropsychology and neuroimaging update. *Acta Psychiatr. Scand. Suppl* **395**, 95-104 (1999).
- David,A.S. The cognitive neuropsychiatry of auditory verbal hallucinations: an overview. *Cognit. Neuropsychiatry* **9**, 107-123 (2004).
- Davies,M.F., Griffin,M., & Vice,S. Affective reactions to auditory hallucinations in psychotic, evangelical and control groups. *Br. J. Clin. Psychol.* **40**, 361-370 (2001).
- de Bakker,P.I., Yelensky,R., Pe'er,I., Gabriel,S.B., Daly,M.J., & Altshuler,D. Efficiency and power in genetic association studies. *Nat. Genet.* **37**, 1217-1223 (2005).
- DeFries,J.C., Fulker,D.W., & LaBuda,M.C. Evidence for a genetic aetiology in reading disability of twins. *Nature* **329**, 537-539 (1987).
- Delespaul,P., deVries,M., & van Os,J. Determinants of occurrence and recovery from hallucinations in daily life. *Soc. Psychiatry Psychiatr. Epidemiol.* **37**, 97-104 (2002).
- Delisi,L.E. Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. *Schizophr. Bull.* **27**, 481-496 (2001).
- Delisi,L.E., Shaw,S.H., Crow,T.J., Shields,G., Smith,A.B., Larach,V.W., Wellman,N., Loftus,J., Nanthakumar,B., Razi,K., Stewart,J., Comazzi,M., Vita,A., Heffner,T., & Sherrington,R. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am. J. Psychiatry* **159**, 803-812 (2002).
- Delisi,L.E. Reviewing the "facts about schizophrenia:: a possible or impossible task? *Schizophr. Res.* **102**, 19-20 (2008).
- Demonet,J.F., Thierry,G., & Cardebat,D. Renewal of the neurophysiology of language: functional neuroimaging. *Physiol Rev.* **85**, 49-95 (2005).
- Dorus,S., Vallender,E.J., Evans,P.D., Anderson,J.R., Gilbert,S.L., Mahowald,M., Wyckoff,G.J., Malcom,C.M., & Lahn,B.T. Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell* **119**, 1027-1040 (2004).
- Dudbridge,F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum. Hered.* **66**, 87-98 (2008).
- Eastwood,S.L. & Harrison,P.J. Cellular basis of reduced cortical reelin expression in schizophrenia. *Am. J. Psychiatry* **163**, 540-542 (2006).
- Eckhardt,F., Lewin,J., Cortese,R., Rakyán,V.K., Attwood,J., Burger,M., Burton,J., Cox,T.V., Davies,R., Down,T.A., Haefliger,C., Horton,R., Howe,K., Jackson,D.K., Kunde,J., Koenig,C., Liddle,J., Niblett,D., Otto,T., Pettett,R., Seemann,S., Thompson,C., West,T., Rogers,J., Olek,A., Berlin,K., & Beck,S. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**, 1378-1385 (2006).
- Enard,W., Przeworski,M., Fisher,S.E., Lai,C.S., Wiebe,V., Kitano,T., Monaco,A.P., & Paabo,S. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869-872 (2002).
- Enard,W., Khaitovich,P., Klose,J., Zollner,S., Heissig,F., Giavalisco,P., Nieselt-Struwe,K., Muchmore,E., Varki,A., Ravid,R., Doxiadis,G.M., Bontrop,R.E., & Paabo,S. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340-343 (2002).

Bibliografía

- Farmer,A.E., McGuffin,P., & Gottesman,I.I. Twin concordance for DSM-III schizophrenia. Scrutinizing the validity of the definition. *Arch. Gen. Psychiatry* **44**, 634-641 (1987).
- Feinberg,A.P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447** , 433-440 (2007).
- Felsenfeld,S. Finding susceptibility genes for developmental disorders of speech: the long and winding road. *J. Commun. Disord.* **35**, 329-345 (2002).
- Ferland,R.J., Cherry,T.J., Preware,P.O., Morrissey,E.E., & Walsh,C.A. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *J. Comp Neurol.* **460**, 266-279 (2003).
- Ferrari,P.F., Gallese,V., Rizzolatti,G., & Fogassi,L. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* **17**, 1703-1714 (2003).
- Feuk,L., Carson,A.R., & Scherer,S.W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85-97 (2006).
- Fisher,S.E., Vargha-Khadem,F., Watkins,K.E., Monaco,A.P., & Pembrey,M.E. Localisation of a gene implicated in a severe speech and language disorder. *Nat. Genet.* **18**, 168-170 (1998).
- Fisher,S.E. & Marcus,G.F. The eloquent ape: genes, brains and the evolution of language. *Nat. Rev. Genet.* **7**, 9-20 (2006).
- Flaum,M., Andreasen,N.C., Swayze,V.W., O'Leary,D.S., & Alliger,R.J. IQ and brain size in schizophrenia. *Psychiatry Res.* **53**, 243-257 (1994).
- Folstein,S. & Rutter,M. Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiatry* **18**, 297-321 (1977).
- Folstein,S.E. & Mankoski,R.E. Chromosome 7q: where autism meets language disorder? *Am. J. Hum. Genet.* **67**, 278-281 (2000).
- Fombonne,E. The prevalence of autism. *JAMA* **289**, 87-89 (2003).
- Fraga,M.F., Ballestar,E., Paz,M.F., Ropero,S., Setien,F., Ballestar,M.L., Heine-Suner,D., Cigudosa,J.C., Urioste,M., Benitez,J., Boix-Chornet,M., Sanchez-Aguilera,A., Ling,C., Carlsson,E., Poulsen,P., Vaag,A., Stephan,Z., Spector,T.D., Wu,Y.Z., Plass,C., & Esteller,M. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A* **102**, 10604-10609 (2005).
- Francks,C., MacPhie,I.L., & Monaco,A.P. The genetic basis of dyslexia. *Lancet Neurol.* **1**, 483-490 (2002).
- Frankland,P.W., Bontempi,B., Talton,L.E., Kaczmarek,L., & Silva,A.J. The involvement of the anterior cingulate cortex in remote contextual fear memory. *Science* **304**, 881-883 (2004).
- Fraser,H.B., Hirsh,A.E., Steinmetz,L.M., Scharfe,C., & Feldman,M.W. Evolutionary rate in the protein interaction network. *Science* **296**, 750-752 (2002).
- Frech,K., Danescu-Mayer,J., & Werner,T. A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. *J. Mol. Biol.* **270**, 674-687 (1997).

- Freedman,R., Leonard,S., Olincy,A., Kaufmann,C.A., Malaspina,D., Cloninger,C.R., Svrakic,D., Faraone,S.V., & Tsuang,M.T. Evidence for the multigenic inheritance of schizophrenia. *Am. J. Med. Genet.* **105**, 794-800 (2001).
- Friedman,J.I., Vrijenhoek,T., Markx,S., Janssen,I.M., van der Vliet,W.A., Faas,B.H., Knoers,N.V., Cahn,W., Kahn,R.S., Edelmann,L., Davis,K.L., Silverman,J.M., Brunner,H.G., van Kessel,A.G., Wijmenga,C., Ophoff,R.A., & Veltman,J.A. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol. Psychiatry* **13**, 261-266 (2008).
- Fritsch,E.F., Sambrook,J., & Maniatis,T. **Molecular Cloning**. (1989).
- Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L., & Paul,C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A* **89**, 1827-1831 (1992).
- Fusar-Poli,P. & Broome,M.R. Conceptual issues in psychiatric neuroimaging. *Curr. Opin. Psychiatry* **19**, 608-612 (2006).
- Fusar-Poli,P., Perez,J., Broome,M., Borgwardt,S., Placentino,A., Caverzasi,E., Cortesi,M., Veggiotti,P., Politi,P., Barale,F., & McGuire,P. Neurofunctional correlates of vulnerability to psychosis: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **31**, 465-484 (2007).
- Gatchel,J.R. & Zoghbi,H.Y. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* **6**, 743-755 (2005).
- Gauderman,W.J. Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* **155**, 478-484 (2002).
- Gauthier,J., Jooper,R., Mottron,L., Laurent,S., Fuchs,M., De,K., V, & Rouleau,G.A. Mutation screening of FOXP2 in individuals diagnosed with autistic disorder. *Am. J. Med. Genet. A* **118A**, 172-175 (2003).
- Gecz,J., Gedeon,A.K., Sutherland,G.R., & Mulley,J.C. Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat. Genet.* **13**, 105-108 (1996).
- Glahn,D.C., Laird,A.R., Ellison-Wright,I., Thelen,S.M., Robinson,J.L., Lancaster,J.L., Bullmore,E., & Fox,P.T. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. *Biol. Psychiatry* **64**, 774-781 (2008).
- Glatt,S.J., Faraone,S.V., & Tsuang,M.T. CAG-repeat length in exon 1 of KCNN3 does not influence risk for schizophrenia or bipolar disorder: a meta-analysis of association studies. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **121B**, 14-20 (2003).
- Glazko,G.V. & Nei,M. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424-434 (2003).
- Gong,X., Jia,M., Ruan,Y., Shuang,M., Liu,J., Wu,S., Guo,Y., Yang,J., Ling,Y., Yang,X., & Zhang,D. Association between the FOXP2 gene and autistic disorder in Chinese population. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **127B**, 113-116 (2004).
- Gonzalez,J.C., Sanjuan,J., Canete,C., Echanove,M.J., & Leal,C. [Evaluation of auditory hallucinations: the PSYRATS scale]. *Actas Esp. Psiquiatr.* **31**, 10-17 (2003).
- Goodman,M., Porter,C.A., Czelusniak,J., Page,S.L., Schneider,H., Shoshani,J., Gunnell,G., & Groves,C.P. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**, 585-598 (1998).

Bibliografía

- Gopnik, M. & Crago, M.B. Familial aggregation of a developmental language disorder. *Cognition* **39**, 1-50 (1991).
- Gottesman, I. I. Schizophrenia Genesis. The Origin of Madness. 1991. Freeman. New York.
Ref Type: Generic
- Gottesman, I.I. & Gould, T.D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* **160**, 636-645 (2003).
- Grayson, D.R., Jia, X., Chen, Y., Sharma, R.P., Mitchell, C.P., Guidotti, A., & Costa, E. Reelin promoter hypermethylation in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A* **102**, 9341-9346 (2005).
- Green, M.F., Hugdahl, K., & Mitchell, S. Dichotic listening during auditory hallucinations in patients with schizophrenia. *Am. J. Psychiatry* **151**, 357-362 (1994).
- Gruenbaum, Y., Cedar, H., & Razin, A. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* **295**, 620-622 (1982).
- Gu, J. & Gu, X. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* **19**, 63-65 (2003).
- Gu, Y., Shen, Y., Gibbs, R.A., & Nelson, D.L. Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat. Genet.* **13**, 109-113 (1996).
- Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696-704 (2003).
- Gur, R.E., Keshavan, M.S., & Lawrie, S.M. Deconstructing psychosis with human brain imaging. *Schizophr. Bull.* **33**, 921-931 (2007).
- Gurling, H.M., Kalsi, G., Brynjolfson, J., Sigmundsson, T., Sherrington, R., Mankoo, B.S., Read, T., Murphy, P., Blaveri, E., McQuillin, A., Petursson, H., & Curtis, D. Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. *Am. J. Hum. Genet.* **68**, 661-673 (2001).
- Haddock, G., McCarron, J., Tarrier, N., & Faragher, E.B. Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychol. Med.* **29**, 879-889 (1999).
- Haesler, S., Wada, K., Nshdejan, A., Morrisey, E.E., Lints, T., Jarvis, E.D., & Scharff, C. FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* **24**, 3164-3175 (2004).
- Hafner, H., Riecher-Rossler, A., Maurer, K., Fatkenheuer, B., & Loffler, W. First onset and early symptomatology of schizophrenia. A chapter of epidemiological and neurobiological research into age and sex differences. *Eur. Arch. Psychiatry Clin. Neurosci.* **242**, 109-118 (1992).
- Hapgood, J.P., Riedemann, J., & Scherer, S.D. Regulation of gene expression by GC-rich DNA cis-elements. *Cell Biol. Int.* **25**, 17-31 (2001).
- Harrison, P.J. & Weinberger, D.R. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry* **10**, 40-68 (2005).
- Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K.D., & Wray, G.A. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat. Genet.* **39**, 1140-1144 (2007).

- Heinz,A., Romero,B., Gallinat,J., Juckel,G., & Weinberger,D.R. Molecular brain imaging and the neurobiology and genetics of schizophrenia. *Pharmacopsychiatry* **36 Suppl 3**, S152-S157 (2003).
- Heston,L.L. Psychiatric disorders in foster home reared children of schizophrenic mothers. *Br. J. Psychiatry* **112**, 819-825 (1966).
- Higgins,J., Gore,R., Gutkind,D., Mednick,S.A., Parnas,J., Schulsinger,F., & Cannon,T.D. Effects of child-rearing by schizophrenic mothers: a 25-year follow-up. *Acta Psychiatr. Scand.* **96**, 402-404 (1997).
- Hirschhorn,J.N. & Daly,M.J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95-108 (2005).
- Hirschhorn,J.N. Genetic approaches to studying common diseases and complex traits. *Pediatr. Res.* **57**, 74R-77R (2005).
- Hoffman,R.E., Stopek,S., & Andreasen,N.C. A comparative study of manic vs schizophrenic speech disorganization. *Arch. Gen. Psychiatry* **43**, 831-838 (1986).
- Hoffman,R.E., Hampson,M., Wu,K., Anderson,A.W., Gore,J.C., Buchanan,R.J., Constable,R.T., Hawkins,K.A., Sahay,N., & Krystal,J.H. Probing the pathophysiology of auditory/verbal hallucinations by combining functional magnetic resonance imaging and transcranial magnetic stimulation. *Cereb. Cortex* **17**, 2733-2743 (2007).
- Holmes,S.E., O'Hearn,E.E., McInnis,M.G., Gorelick-Feldman,D.A., Kleiderlein,J.J., Callahan,C., Kwak,N.G., Ingersoll-Ashworth,R.G., Sherr,M., Sumner,A.J., Sharp,A.H., Ananth,U., Seltzer,W.K., Boss,M.A., Viera-Saecker,A.M., Epplen,J.T., Riess,O., Ross,C.A., & Margolis,R.L. Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat. Genet.* **23**, 391-392 (1999).
- Honea,R., Crow,T.J., Passingham,D., & Mackay,C.E. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am. J. Psychiatry* **162**, 2233-2245 (2005).
- Horrobin,D.F. Schizophrenia: the illness that made us human. *Med. Hypotheses* **50**, 269-288 (1998).
- Hsieh,J. & Gage,F.H. Chromatin remodeling in neural development and plasticity. *Curr. Opin. Cell Biol.* **17**, 664-671 (2005).
- Hulshoff,P.H., Schnack,H.G., Mandl,R.C., Cahn,W., Collins,D.L., Evans,A.C., & Kahn,R.S. Focal white matter density changes in schizophrenia: reduced inter-hemispheric connectivity. *Neuroimage.* **21**, 27-35 (2004).
- Hurst,J.A., Baraitser,M., Auger,E., Graham,F., & Norell,S. An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.* **32**, 352-355 (1990).
- HUXLEY,J., MAYR,E., OSMOND,H., & HOFFER,A. SCHIZOPHRENIA AS A GENETIC MORPHISM. *Nature* **204**, 220-221 (1964).
- Impagnatiello,F., Guidotti,A.R., Pesold,C., Dwivedi,Y., Caruncho,H., Pisu,M.G., Uzunov,D.P., Smalheiser,N.R., Davis,J.M., Pandey,G.N., Pappas,G.D., Tueting,P., Sharma,R.P., & Costa,E. A decrease of reelin expression as a putative vulnerability factor in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A* **95**, 15718-15723 (1998).
- Ingraham,L.J. & Kety,S.S. Adoption studies of schizophrenia. *Am. J. Med. Genet.* **97**, 18-22 (2000).

Bibliografía

International Schizophrenia Consortium Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-241 (2008).

Jablensky,A., Sartorius,N., Ernberg,G., Anker,M., Korten,A., Cooper,J.E., Day,R., & Bertelsen,A. Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychol. Med. Monogr Suppl* **20**, 1-97 (1992).

Jaenisch,R. & Bird,A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33 Suppl**, 245-254 (2003).

Job,D.E., Whalley,H.C., Johnstone,E.C., & Lawrie,S.M. Grey matter changes over time in high risk subjects developing schizophrenia. *Neuroimage.* **25**, 1023-1030 (2005).

Johns,L.C. & van Os,J. The continuity of psychotic experiences in the general population. *Clin. Psychol. Rev.* **21**, 1125-1141 (2001).

Johns,L.C., Hemsley,D., & Kuipers,E. A comparison of auditory hallucinations in a psychiatric and non-psychiatric group. *Br. J. Clin. Psychol.* **41**, 81-86 (2002).

Johns,L.C. Hallucinations in the general population. *Curr. Psychiatry Rep.* **7**, 162-167 (2005).

Jones,M.W., Errington,M.L., French,P.J., Fine,A., Bliss,T.V., Garel,S., Charnay,P., Bozon,B., Laroche,S., & Davis,S. A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories. *Nat. Neurosci.* **4**, 289-296 (2001).

Jones,P.A. & Takai,D. The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068-1070 (2001).

Jung, S. M, Jung, B. J., Cho, J. S., and Park, J. M. FOXP2 gene possibly associated with Korean schizophrenic patients. *European Neuropsychopharmacology The Journal of the European College of Neuropsychopharmacology* 18 S4. 2008.

Kaestner,K.H., Knochel,W., & Martinez,D.E. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* **14**, 142-146 (2000).

Kaminen,N., Hannula-Jouppi,K., Kestila,M., Lahermo,P., Muller,K., Kaaranen,M., Myllyluoma,B., Voutilainen,A., Lyytinen,H., Nopola-Hemmi,J., & Kere,J. A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *J. Med. Genet.* **40**, 340-345 (2003).

Kato,C., Petronis,A., Okazaki,Y., Tochigi,M., Umekage,T., & Sasaki,T. Molecular genetic studies of schizophrenia: challenges and insights. *Neurosci. Res.* **43**, 295-304 (2002).

Katoh, M. & Katoh, M. Human FOX gene family. *Int J Oncol. Nov;*25(5):1495-500 (2004).

Kelley,J.L. & Swanson,W.J. Positive selection in the human genome: from genome scans to biological significance. *Annu. Rev. Genomics Hum. Genet.* **9**, 143-160 (2008).

Kety,S.S., Rosenthal,D., Wender,P.H., Schulsinger,F., & Jacobsen,B. Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic: a preliminary report based on psychiatric interviews. *Proc. Annu. Meet. Am. Psychopathol. Assoc.* 147-165 (1975).

Kety,S.S., Wender,P.H., Jacobsen,B., Ingraham,L.J., Jansson,L., Faber,B., & Kinney,D.K. Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark. *Arch. Gen. Psychiatry* **51**, 442-455 (1994).

- Khaitovich,P., Weiss,G., Lachmann,M., Hellmann,I., Enard,W., Muetzel,B., Wirkner,U., Ansorge,W., & Paabo,S. A neutral model of transcriptome evolution. *PLoS. Biol.* **2**, E132 (2004).
- Khaitovich,P., Enard,W., Lachmann,M., & Paabo,S. Evolution of primate gene expression. *Nat. Rev. Genet.* **7**, 693-702 (2006).
- Khaitovich,P., Lockstone,H.E., Wayland,M.T., Tsang,T.M., Jayatilaka,S.D., Guo,A.J., Zhou,J., Somel,M., Harris,L.W., Holmes,E., Paabo,S., & Bahn,S. Metabolic changes in schizophrenia and human brain evolution. *Genome Biol.* **9**, R124 (2008).
- King,M.C. & Wilson,A.C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-116 (1975).
- Knight,S.J., Flannery,A.V., Hirst,M.C., Campbell,L., Christodoulou,Z., Phelps,S.R., Pointon,J., Middleton-Price,H.R., Barnicoat,A., Pembrey,M.E., & . Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**, 127-134 (1993).
- Kohn,M.H., Fang,S., & Wu,C.I. Inference of positive and negative selection on the 5' regulatory regions of Drosophila genes. *Mol. Biol. Evol.* **21**, 374-383 (2004).
- Koob,M.D., Moseley,M.L., Schut,L.J., Benzow,K.A., Bird,T.D., Day,J.W., & Ranum,L.P. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.* **21**, 379-384 (1999).
- Kops,G.J., Medema,R.H., Glassford,J., Essers,M.A., Dijkers,P.F., Coffey,P.J., Lam,E.W., & Burgering,B.M. Control of cell cycle exit and entry by protein kinase B-regulated forkhead transcription factors. *Mol. Cell Biol.* **22**, 2025-2036 (2002).
- Kraepelin,E. Dementia praecox and paraphrenia. (R. E. Krieger Pub. Co., New York; 1971).
- Krause,J., Lalueza-Fox,C., Orlando,L., Enard,W., Green,R.E., Burbano,H.A., Hublin,J.J., Hanni,C., Fortea,J., de la,R.M., Bertranpetit,J., Rosas,A., & Paabo,S. The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.* **17**, 1908-1912 (2007).
- Krawiecka,M., Goldberg,D., & Vaughan,M. A standardized psychiatric assessment scale for rating chronic psychotic patients. *Acta Psychiatr. Scand.* **55**, 299-308 (1977).
- Kusek,J.C., Greene,R.M., Nugent,P., & Pisano,M.M. Expression of the E2F family of transcription factors during murine development. *Int. J. Dev. Biol.* **44**, 267-277 (2000).
- Lai,C.S., Fisher,S.E., Hurst,J.A., Levy,E.R., Hodgson,S., Fox,M., Jeremiah,S., Povey,S., Jamison,D.C., Green,E.D., Vargha-Khadem,F., & Monaco,A.P. The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am. J. Hum. Genet.* **67**, 357-368 (2000).
- Lai,C.S., Fisher,S.E., Hurst,J.A., Vargha-Khadem,F., & Monaco,A.P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519-523 (2001).
- Lai,C.S., Gerrelli,D., Monaco,A.P., Fisher,S.E., & Copp,A.J. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* **126**, 2455-2462 (2003).
- Lander,E.S. & Schork,N.J. Genetic dissection of complex traits. *Science* **265**, 2037-2048 (1994).
- Laroche,F., Ramoz,N., Leroy,S., Fortin,C., Rousselot-Paillet,B., Philippe,A., Colleaux,L., Bresson,J.L., Mogenet,A., Golse,B., Mouren-Simeoni,M.C., Gorwood,P., Galli,T.,

Bibliografía

Simonneau,M., Krebs,M.O., & Robel,L. Polymorphisms of coding trinucleotide repeats of homeogenes in neurodevelopmental psychiatric disorders. *Psychiatr. Genet.* **18**, 295-301 (2008).

Larsen,F., Gundersen,G., Lopez,R., & Prydz,H. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095-1107 (1992).

Lee,P.H. & Shatkay,H. BNTagger: improved tagging SNP selection using Bayesian networks. *Bioinformatics.* **22**, e211-e219 (2006).

Lehmann,O.J., Sowden,J.C., Carlsson,P., Jordan,T., & Bhattacharya,S.S. Fox's in development and disease. *Trends Genet.* **19**, 339-344 (2003).

Lencz,T., Morgan,T.V., Athanasiou,M., Dain,B., Reed,C.R., Kane,J.M., Kucherlapati,R., & Malhotra,A.K. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol. Psychiatry* **12**, 572-580 (2007).

Levitan,C., Ward,P.B., & Catts,S.V. Superior temporal gyral volumes and laterality correlates of auditory hallucinations in schizophrenia. *Biol. Psychiatry* **46**, 955-962 (1999).

Lewis,B.A., Shriberg,L.D., Freebairn,L.A., Hansen,A.J., Stein,C.M., Taylor,H.G., & Iyengar,S.K. The genetic bases of speech sound disorders: evidence from spoken and written language. *J. Speech Lang Hear. Res.* **49**, 1294-1312 (2006).

Lewis,C.M., Levinson,D.F., Wise,L.H., Delisi,L.E., Straub,R.E., Hovatta,I., Williams,N.M., Schwab,S.G., Pulver,A.E., Faraone,S.V., Brzustowicz,L.M., Kaufmann,C.A., Garver,D.L., Gurling,H.M., Lindholm,E., Coon,H., Moises,H.W., Byerley,W., Shaw,S.H., Mesen,A., Sherrington,R., O'Neill,F.A., Walsh,D., Kendler,K.S., Ekelund,J., Paunio,T., Lonqvist,J., Peltonen,L., O'Donovan,M.C., Owen,M.J., Wildenauer,D.B., Maier,W., Nestadt,G., Blouin,J.L., Antonarakis,S.E., Mowry,B.J., Silverman,J.M., Crowe,R.R., Cloninger,C.R., Tsuang,M.T., Malaspina,D., Harkavy-Friedman,J.M., Svrakic,D.M., Bassett,A.S., Holcomb,J., Kalsi,G., McQuillin,A., Brynjolfson,J., Sigmundsson,T., Petursson,H., Jazin,E., Zoega,T., & Helgason,T. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am. J. Hum. Genet.* **73**, 34-48 (2003).

Li,E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3**, 662-673 (2002).

Li,G., Wang,J., Rossiter,S.J., Jones,G., & Zhang,S. Accelerated FoxP2 evolution in echolocating bats. *PLoS. ONE.* **2**, e900 (2007).

Li,H., Yamagata,T., Mori,M., & Momoi,M.Y. Absence of causative mutations and presence of autism-related allele in FOXP2 in Japanese autistic patients. *Brain Dev.* **27**, 207-210 (2005).

Li,J.Z., Vawter,M.P., Walsh,D.M., Tomita,H., Evans,S.J., Choudary,P.V., Lopez,J.F., Avelar,A., Shokoohi,V., Chung,T., Mesarwi,O., Jones,E.G., Watson,S.J., Akil,H., Bunney,W.E., Jr., & Myers,R.M. Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum. Mol. Genet.* **13**, 609-616 (2004).

Li,S., Weidenfeld,J., & Morrisey,E.E. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol. Cell Biol.* **24**, 809-822 (2004).

Liegeois,F., Baldeweg,T., Connelly,A., Gadian,D.G., Mishkin,M., & Vargha-Khadem,F. Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat. Neurosci.* **6**, 1230-1237 (2003).

- Lipska, B.K., Deep-Soboslay, A., Weickert, C.S., Hyde, T.M., Martin, C.E., Herman, M.M., & Kleinman, J.E. Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. *Biol. Psychiatry* **60**, 650-658 (2006).
- Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., & Ranum, L.P. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293**, 864-867 (2001).
- Loberg, E.M., Jorgensen, H.A., & Hugdahl, K. Dichotic listening in schizophrenic patients: effects of previous vs. ongoing auditory hallucinations. *Psychiatry Res.* **128**, 167-174 (2004).
- Lohoff, F.W., Weller, A.E., Bloch, P.J., Buono, R.J., Doyle, G.A., Ferraro, T.N., & Berrettini, W.H. Association between polymorphisms in the vesicular monoamine transporter 1 gene (VMAT1/SLC18A1) on chromosome 8p and schizophrenia. *Neuropsychobiology* **57**, 55-60 (2008).
- Lu, M.M., Li, S., Yang, H., & Morrisey, E.E. Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Mech. Dev.* **119 Suppl 1**, S197-S202 (2002).
- MacDermot, K.D., Bonora, E., Sykes, N., Coupe, A.M., Lai, C.S., Vernes, S.C., Vargha-Khadem, F., McKenzie, F., Smith, R.L., Monaco, A.P., & Fisher, S.E. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.* **76**, 1074-1080 (2005).
- Malhotra, A.K., Goldman, D., Mazzanti, C., Clifton, A., Breier, A., & Pickar, D. A functional serotonin transporter (5-HTT) polymorphism is associated with psychosis in neuroleptic-free schizophrenics. *Mol. Psychiatry* **3**, 328-332 (1998).
- Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S., Breschel, T.S., Stine, O.C., Callahan, C., McInnis, M.G., & Ross, C.A. cDNAs with long CAG trinucleotide repeats from human brain. *Hum. Genet.* **100**, 114-122 (1997).
- Marui, T., Koishi, S., Funatogawa, I., Yamamoto, K., Matsumoto, H., Hashimoto, O., Nanba, E., Kato, C., Ishijima, M., Watanabe, K., Kasai, K., Kato, N., & Sasaki, T. No association of FOXP2 and PTPRZ1 on 7q31 with autism from the Japanese population. *Neurosci. Res.* **53**, 91-94 (2005).
- Matheny, A.P., Jr. & Bruggemann, C.E. Children's speech: hereditary components and sex differences. *Folia Phoniatr. (Basel)* **25**, 442-449 (1973).
- Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L., Achari, M., Pulst, S.M., Alonso, E., Noebels, J.L., Nelson, D.L., Zoghbi, H.Y., & Ashizawa, T. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet.* **26**, 191-194 (2000).
- McClellan, J.M., Susser, E., & King, M.C. Schizophrenia: a common disease caused by multiple rare alleles. *Br. J. Psychiatry* **190**, 194-199 (2007).
- McDonald, B., Highley, J.R., Walker, M.A., Herron, B.M., Cooper, S.J., Esiri, M.M., & Crow, T.J. Anomalous asymmetry of fusiform and parahippocampal gyrus gray matter in schizophrenia: A postmortem study. *Am. J. Psychiatry* **157**, 40-47 (2000).
- McGrath, J.J. Variations in the incidence of schizophrenia: data versus dogma. *Schizophr. Bull.* **32**, 195-197 (2006).
- McGuire, P.K. & Matsumoto, K. Functional neuroimaging in mental disorders. *World Psychiatry* **3**, 6-11 (2004).

Bibliografía

McKay,C.M., Headlam,D.M., & Copolov,D.L. Central auditory processing in patients with auditory hallucinations. *Am. J. Psychiatry* **157**, 759-766 (2000).

McKenna,P.J. & Oh,T. Schizophrenic speech. (Cambridge University Press,2005).

Meaburn,E., Dale,P.S., Craig,I.W., & Plomin,R. Language-impaired children: No sign of the FOXP2 mutation. *Neuroreport* **13**, 1075-1077 (2002).

Mexal,S., Berger,R., Adams,C.E., Ross,R.G., Freedman,R., & Leonard,S. Brain pH has a significant impact on human postmortem hippocampal gene expression profiles. *Brain Res.* **1106**, 1-11 (2006).

Mill,J., Tang,T., Kaminsky,Z., Khare,T., Yazdanpanah,S., Bouchard,L., Jia,P., Assadzadeh,A., Flanagan,J., Schumacher,A., Wang,S.C., & Petronis,A. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am. J. Hum. Genet.* **82**, 696-711 (2008).

Mizutani,A., Matsuzaki,A., Momoi,M.Y., Fujita,E., Tanabe,Y., & Momoi,T. Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochem. Biophys. Res. Commun.* **353**, 869-874 (2007).

Muhle,R., Trentacoste,S.V., & Rapin,I. The genetics of autism. *Pediatrics* **113**, e472-e486 (2004).

Nayani,T.H. & David,A.S. The auditory hallucination: a phenomenological survey. *Psychol. Med.* **26**, 177-189 (1996).

Newbury,D.F., Bonora,E., Lamb,J.A., Fisher,S.E., Lai,C.S., Baird,G., Jannoun,L., Slonims,V., Stott,C.M., Merricks,M.J., Bolton,P.F., Bailey,A.J., & Monaco,A.P. FOXP2 is not a major susceptibility gene for autism or specific language impairment. *Am. J. Hum. Genet.* **70**, 1318-1327 (2002).

Nielsen,R., Bustamante,C., Clark,A.G., Glanowski,S., Sackton,T.B., Hubisz,M.J., Fedel-Alon,A., Tanenbaum,D.M., Civallo,D., White,T.J., Sninsky,J., Adams,M.D., & Cargill,M. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS. Biol.* **3**, e170 (2005).

O'Brien,E.K., Zhang,X., Nishimura,C., Tomblin,J.B., & Murray,J.C. Association of specific language impairment (SLI) to the region of 7q31. *Am. J. Hum. Genet.* **72**, 1536-1543 (2003).

Oeth, P., Beaulieu, M., Park, C., Kosman, D., del Mistro, G., van den Boom, D., and Jurinke, C. iPLEX™ Assay: Increased Plexing Efficiency and Flexibility for MassARRAY. System Through Single Base Primer Extension with Mass-Modified Terminators. Sequenom Application Note . 2005.
Ref Type: Generic

Okubo, T., Harada, S., Higuchi, S., & Matsushita, S. Investigation of quantitative trait loci in the CCKAR gene with susceptibility to alcoholism. *Alcohol Clin Exp Res.* 26(8 Suppl):2S-5S (2002).

Ott,U., Reuter,M., Hennig,J., & Vaitl,D. Evidence for a common biological basis of the Absorption trait, hallucinogen effects, and positive symptoms: epistasis between 5-HT2a and COMT polymorphisms. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **137B**, 29-32 (2005).

Ovcharenko,I., Boffelli,D., & Loots,G.G. eShadow: a tool for comparing closely related sequences. *Genome Res.* **14**, 1191-1198 (2004).

- Ovcharenko,I., Loots,G.G., Giardine,B.M., Hou,M., Ma,J., Hardison,R.C., Stubbs,L., & Miller,W. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res.* **15**, 184-194 (2005).
- Owen,M.J., Craddock,N., & O'Donovan,M.C. Schizophrenia: genes at last? *Trends Genet.* **21**, 518-525 (2005).
- Paillere-Martinot,M., Caclin,A., Artiges,E., Poline,J.B., Joliot,M., Mallet,L., Recasens,C., Attar-Levy,D., & Martinot,J.L. Cerebral gray and white matter reductions and clinical correlates in patients with early onset schizophrenia. *Schizophr. Res.* **50**, 19-26 (2001).
- Pantelis,C., Velakoulis,D., McGorry,P.D., Wood,S.J., Suckling,J., Phillips,L.J., Yung,A.R., Bullmore,E.T., Brewer,W., Soulsby,B., Desmond,P., & McGuire,P.K. Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison. *Lancet* **361**, 281-288 (2003).
- Park,N., Juo,S.H., Cheng,R., Liu,J., Loth,J.E., Lilliston,B., Nee,J., Grunn,A., Kanyas,K., Lerer,B., Endicott,J., Gilliam,T.C., & Baron,M. Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia. *Mol. Psychiatry* **9**, 1091-1099 (2004).
- Perez,F.A., Ballester,G.M., Giron,G.M., & Gomez,B.M. [Reliability, validity and sensitivity to change of the psychiatric evaluation scale of Krawiecka]. *Actas Luso. Esp. Neurol. Psiquiatr. Cienc. Afines* **17**, 111-118 (1989).
- Perutz,M.F., Johnson,T., Suzuki,M., & Finch,J.T. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U. S. A* **91**, 5355-5358 (1994).
- Petronis,A. Epigenetics and bipolar disorder: new opportunities and challenges. *Am. J. Med. Genet. C. Semin. Med. Genet.* **123C**, 65-75 (2003).
- Petronis,A. The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol. Psychiatry* **55**, 965-970 (2004).
- Phillips,K. & Luisi,B. The virtuoso of versatility: POU proteins that flex to fit. *J. Mol. Biol.* **302**, 1023-1039 (2000).
- Pinker,S. in *Language Evolution* (Oxford University Press,2003).
- Poepfel,D. & Hickok,G. Towards a new functional anatomy of language. *Cognition* **92**, 1-12 (2004).
- Pollard,K.S., Salama,S.R., Lambert,N., Lambot,M.A., Coppens,S., Pedersen,J.S., Katzman,S., King,B., Onodera,C., Siepel,A., Kern,A.D., Dehay,C., Igel,H., Ares,M., Jr., Vanderhaeghen,P., & Haussler,D. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167-172 (2006).
- Prabhakar,S., Noonan,J.P., Paabo,S., & Rubin,E.M. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).
- Pritchard,J.K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69**, 124-137 (2001).
- Purvis,A. A composite estimate of primate phylogeny. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **348**, 405-421 (1995).

Bibliografía

- Quandt,K., Frech,K., Karas,H., Wingender,E., & Werner,T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23**, 4878-4884 (1995).
- Rajarethinam,R.P., DeQuardo,J.R., Nalepa,R., & Tandon,R. Superior temporal gyrus in schizophrenia: a volumetric magnetic resonance imaging study. *Schizophr. Res.* **41**, 303-312 (2000).
- Razi,K., Greene,K.P., Sakuma,M., Ge,S., Kushner,M., & Delisi,L.E. Reduction of the parahippocampal gyrus and the hippocampus in patients with chronic schizophrenia. *Br. J. Psychiatry* **174**, 512-519 (1999).
- Razin,A. & Cedar,H. DNA methylation and genomic imprinting. *Cell* **77**, 473-476 (1994).
- Richards,M., Iijima,Y., Kondo,H., Shizuno,T., Hori,H., Arima,K., Saitoh,O., & Kunugi,H. Association study of the vesicular monoamine transporter 1 (VMAT1) gene with schizophrenia in a Japanese population. *Behav. Brain Funct.* **2**, 39 (2006).
- Richler,E., Reichert,J.G., Buxbaum,J.D., & McInnes,L.A. Autism and ultraconserved non-coding sequence on chromosome 7q. *Psychiatr. Genet.* **16**, 19-23 (2006).
- Rivero,O., Sanjuan,J., Molto,M.D., Aguilar,E.J., Gonzalez,J.C., De Frutos,R., & Najera,C. The microcephaly ASPM gene and schizophrenia: A preliminary study. *Schizophr. Res.* **84**, 427-429 (2006).
- Rizzolatti,G., Fadiga,L., Gallese,V., & Fogassi,L. Premotor cortex and the recognition of motor actions. *Brain Res. Cogn Brain Res.* **3**, 131-141 (1996).
- Rizzolatti,G. & Craighero,L. The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169-192 (2004).
- Rosenthal,D. & Quinn,O.W. Quadruplet hallucinations. Phenotypic variations of a schizophrenic genotype. *Arch. Gen. Psychiatry* **34**, 817-827 (1977).
- Rossell,S.L. & Boundy,C.L. Are auditory-verbal hallucinations associated with auditory affective processing deficits? *Schizophr. Res.* **78**, 95-106 (2005).
- Sabeti,P.C., Schaffner,S.F., Fry,B., Lohmueller,J., Varilly,P., Shamovsky,O., Palma,A., Mikkelsen,T.S., Altshuler,D., & Lander,E.S. Positive natural selection in the human lineage. *Science* **312**, 1614-1620 (2006).
- Sanger,F., Nicklen,S., & Coulson,A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A* **74**, 5463-5467 (1977).
- Sanjuan,J., Toirac,I., Gonzalez,J.C., Leal,C., Molto,M.D., Najera,C., & De Frutos,R. A possible association between the CCK-AR gene and persistent auditory hallucinations in schizophrenia. *Eur. Psychiatry* **19**, 349-353 (2004).
- Sanjuan, J., Najera, C., De Frutos, R., and Molto, M. D. Genética de las Alucinaciones Auditivas. *Curr.Psychiatry Rep.*(edición en español) . 2005.
Ref Type: Generic
- Sanjuan,J. & Gonzalez,J.C. in *La profecía de Darwin* (Ars Medica,2005).
- Sanjuan,J., Tolosa,A., Gonzalez,J.C., Aguilar,E.J., Molto,M.D., Najera,C., & De Frutos,R. FOXP2 polymorphisms in patients with schizophrenia. *Schizophr. Res.* **73**, 253-256 (2005).

- Sanjuan,J., Rivero,O., Aguilar,E.J., Gonzalez,J.C., Molto,M.D., De Frutos,R., Lesch,K.P., & Najera,C. Serotonin transporter gene polymorphism (5-HTTLPR) and emotional response to auditory hallucinations in schizophrenia. *Int. J. Neuropsychopharmacol.* **9**, 131-133 (2006).
- Sanjuan,J., Tolosa,A., Gonzalez,J.C., Aguilar,E.J., Perez-Tur,J., Najera,C., Molto,M.D., & De Frutos,R. Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. *Psychiatr. Genet.* **16**, 67-72 (2006).
- Sanjuan, J. Etiopatogenia de las esquizofrenias. Trastornos psicóticos. 2007. Ars Medica.
- Scharff,C. & Haesler,S. An evolutionary perspective on FoxP2: strictly for the birds? *Curr. Opin. Neurobiol.* **15**, 694-703 (2005).
- Scherf,M., Klingenhoff,A., & Werner,T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* **297**, 599-606 (2000).
- Scherzinger,E., Sittler,A., Schweiger,K., Heiser,V., Lurz,R., Hasenbank,R., Bates,G.P., Lehrach,H., & Wanker,E.E. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. U. S. A* **96**, 4604-4609 (1999).
- Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8. Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. *Am. J. Med. Genet.* **67**, 580-594 (1996).
- Schneider,K. *Clinical Psychopathology.* (New York; 1959).
- Schork,N.J., Greenwood,T.A., & Braff,D.L. Statistical genetics concepts and approaches in schizophrenia and related neuropsychiatric research. *Schizophr. Bull.* **33**, 95-104 (2007).
- Schroeder,D.I. & Myers,R.M. Multiple transcription start sites for FOXP2 with varying cellular specificities. *Gene* **413**, 42-48 (2008).
- Segurado,R., Detera-Wadleigh,S.D., Levinson,D.F., Lewis,C.M., Gill,M., Nurnberger,J.I., Jr., Craddock,N., DePaulo,J.R., Baron,M., Gershon,E.S., Ekholm,J., Cichon,S., Turecki,G., Claes,S., Kelsoe,J.R., Schofield,P.R., Badenhop,R.F., Morissette,J., Coon,H., Blackwood,D., McInnes,L.A., Foroud,T., Edenberg,H.J., Reich,T., Rice,J.P., Goate,A., McInnis,M.G., McMahon,F.J., Badner,J.A., Goldin,L.R., Bennett,P., Willour,V.L., Zandi,P.P., Liu,J., Gilliam,C., Joo,S.H., Berrettini,W.H., Yoshikawa,T., Peltonen,L., Lonqvist,J., Nothen,M.M., Schumacher,J., Windemuth,C., Rietschel,M., Propping,P., Maier,W., Alda,M., Grof,P., Rouleau,G.A., Del Favero,J., Van Broeckhoven,C., Mendlewicz,J., Adolfsson,R., Spence,M.A., Luebbert,H., Adams,L.J., Donald,J.A., Mitchell,P.B., Barden,N., Shink,E., Byerley,W., Muir,W., Visscher,P.M., Macgregor,S., Gurling,H., Kalsi,G., McQuillin,A., Escamilla,M.A., Reus,V.I., Leon,P., Freimer,N.B., Ewald,H., Kruse,T.A., Mors,O., Radhakrishna,U., Blouin,J.L., Antonarakis,S.E., & Akarsu,N. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am. J. Hum. Genet.* **73**, 49-62 (2003).
- Shen,L., Kondo,Y., Guo,Y., Zhang,J., Zhang,L., Ahmed,S., Shu,J., Chen,X., Waterland,R.A., & Issa,J.P. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS. Genet.* **3**, 2023-2036 (2007).
- Shi,P., Bakewell,M.A., & Zhang,J. Did brain-specific genes evolve faster in humans than in chimpanzees? *Trends Genet.* **22**, 608-613 (2006).
- Shriberg,L.D., Austin,D., Lewis,B.A., McSweeney,J.L., & Wilson,D.L. The speech disorders classification system (SDCS): extensions and lifespan reference data. *J. Speech Lang Hear. Res.* **40**, 723-740 (1997).

Bibliografía

Shriberg,L.D., Tomblin,J.B., & McSweeney,J.L. Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *J. Speech Lang Hear. Res.* **42**, 1461-1481 (1999).

Shu,W., Yang,H., Zhang,L., Lu,M.M., & Morrisey,E.E. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *J. Biol. Chem.* **276**, 27488-27497 (2001).

Shu,W., Cho,J.Y., Jiang,Y., Zhang,M., Weisz,D., Elder,G.A., Schmeidler,J., De Gasperi,R., Sosa,M.A., Rabidou,D., Santucci,A.C., Perl,D., Morrisey,E., & Buxbaum,J.D. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proc. Natl. Acad. Sci. U. S. A* **102**, 9643-9648 (2005).

Shu,W., Lu,M.M., Zhang,Y., Tucker,P.W., Zhou,D., & Morrisey,E.E. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* **134**, 1991-2000 (2007).

Sigmundsson,T., Suckling,J., Maier,M., Williams,S., Bullmore,E., Greenwood,K., Fukuda,R., Ron,M., & Toone,B. Structural abnormalities in frontal, temporal, and limbic regions and interconnecting white matter tracts in schizophrenic patients with prominent negative symptoms. *Am. J. Psychiatry* **158**, 234-243 (2001).

Slade P.D. and Bentall R.P. Sensory deception: A scientific analysis of hallucinations. 1988. Croom-Helm. London.

SLI Consortium A genomewide scan identifies two novel loci involved in specific language impairment. *Am. J. Hum. Genet.* **70**, 384-398 (2002).

SLI Consortium Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *Am. J. Hum. Genet.* **74**, 1225-1238 (2004).

Smith,S.D. Genes, language development, and language disorders. *Ment. Retard. Dev. Disabil. Res. Rev.* **13**, 96-105 (2007).

Sole,X., Guino,E., Valls,J., Iniesta,R., & Moreno,V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics.* **22**, 1928-1929 (2006).

Soriano, S. Estudio filogenético del promotor del gen FOXP2. Tesis de Máster. Máster de biología molecular, celular y genética. Especialidad de genética. Universitat de Valencia. Septiembre 2007

Spiteri,E., Konopka,G., Coppola,G., Bomar,J., Oldham,M., Ou,J., Vernes,S.C., Fisher,S.E., Ren,B., & Geschwind,D.H. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am. J. Hum. Genet.* **81**, 1144-1157 (2007).

Stan,A.D., Ghose,S., Gao,X.M., Roberts,R.C., Lewis-Amezcu,K., Hatanpaa,K.J., & Tamminga,C.A. Human postmortem tissue: what quality markers matter? *Brain Res.* **1123**, 1-11 (2006).

Stefansson,H., Rujescu,D., Cichon,S., Pietilainen,O.P., Ingason,A., Steinberg,S., Fossdal,R., Sigurdsson,E., Sigmundsson,T., Buizer-Voskamp,J.E., Hansen,T., Jakobsen,K.D., Muglia,P., Francks,C., Matthews,P.M., Gylfason,A., Halldorsson,B.V., Gudbjartsson,D., Thorgeirsson,T.E., Sigurdsson,A., Jonasdottir,A., Jonasdottir,A., Bjornsson,A., Mattiasdottir,S., Blondal,T., Haraldsson,M., Magnusdottir,B.B., Giegling,I., Moller,H.J., Hartmann,A., Shianna,K.V., Ge,D., Need,A.C., Crombie,C., Fraser,G., Walker,N., Lonqvist,J., Suvisaari,J., Tuulio-Henriksson,A., Paunio,T., Touloupoulou,T., Bramon,E., Di Forti,M.,

Murray,R., Ruggeri,M., Vassos,E., Tosato,S., Walshe,M., Li,T., Vasilescu,C., Muhleisen,T.W., Wang,A.G., Ullum,H., Djurovic,S., Melle,I., Olesen,J., Kiemenev,L.A., Franke,B., Sabatti,C., Freimer,N.B., Gulcher,J.R., Thorsteinsdottir,U., Kong,A., Andreassen,O.A., Ophoff,R.A., Georgi,A., Rietschel,M., Werge,T., Petursson,H., Goldstein,D.B., Nothen,M.M., Peltonen,L., Collier,D.A., St Clair,D., & Stefansson,K. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232-236 (2008).

Stein,C.M., Schick,J.H., Gerry,T.H., Shriberg,L.D., Millard,C., Kundtz-Kluge,A., Russo,K., Minich,N., Hansen,A., Freebairn,L.A., Elston,R.C., Lewis,B.A., & Iyengar,S.K. Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am. J. Hum. Genet.* **74**, 283-297 (2004).

Stein,C.M., Millard,C., Kluge,A., Miscimarra,L.E., Cartier,K.C., Freebairn,L.A., Hansen,A.J., Shriberg,L.D., Taylor,H.G., Lewis,B.A., & Iyengar,S.K. Speech sound disorder influenced by a locus in 15q14 region. *Behav. Genet.* **36**, 858-868 (2006).

Stephane,M., Barton,S., & Boutros,N.N. Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophr. Res.* **50**, 61-78 (2001).

Stevens,A.A., Donegan,N.H., Anderson,M., Goldman-Rakic,P.S., & Wexler,B.E. Verbal processing deficits in schizophrenia. *J. Abnorm. Psychol.* **109**, 461-471 (2000).

Strauss,J.S. Hallucinations and delusions as points on continua function. Rating scale evidence. *Arch. Gen. Psychiatry* **21**, 581-586 (1969).

Suarez,B.K., Duan,J., Sanders,A.R., Hinrichs,A.L., Jin,C.H., Hou,C., Buccola,N.G., Hale,N., Weilbaecher,A.N., Nertney,D.A., Olincy,A., Green,S., Schaffer,A.W., Smith,C.J., Hannah,D.E., Rice,J.P., Cox,N.J., Martinez,M., Mowry,B.J., Amin,F., Silverman,J.M., Black,D.W., Byerley,W.F., Crowe,R.R., Freedman,R., Cloninger,C.R., Levinson,D.F., & Gejman,P.V. Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. *Am. J. Hum. Genet.* **78**, 315-333 (2006).

Sullivan,P.F., Kendler,K.S., & Neale,M.C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187-1192 (2003).

Sullivan,P.F. The genetics of schizophrenia. *PLoS. Med.* **2**, e212 (2005).

Sullivan,P.F., Lin,D., Tzeng,J.Y., van den,O.E., Perkins,D., Stroup,T.S., Wagner,M., Lee,S., Wright,F.A., Zou,F., Liu,W., Downing,A.M., Lieberman,J., & Close,S.L. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol. Psychiatry* **13**, 570-584 (2008).

Sundquist,K., Frank,G., & Sundquist,J. Urbanisation and incidence of psychosis and depression: follow-up study of 4.4 million women and men in Sweden. *Br. J. Psychiatry* **184**, 293-298 (2004).

Takahashi,K., Liu,F.C., Hirokawa,K., & Takahashi,H. Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *J. Neurosci. Res.* **73**, 61-72 (2003).

Takahashi,T., Suzuki,M., Zhou,S.Y., Tanino,R., Hagino,H., Kawasaki,Y., Matsui,M., Seto,H., & Kurachi,M. Morphologic alterations of the parcellated superior temporal gyrus in schizophrenia spectrum. *Schizophr. Res.* **83**, 131-143 (2006).

Talkowski,M.E., Kirov,G., Bamne,M., Georgieva,L., Torres,G., Mansour,H., Chowdari,K.V., Milanova,V., Wood,J., McClain,L., Prasad,K., Shirts,B., Zhang,J., O'Donovan,M.C., Owen,M.J.,

Bibliografía

Devlin,B., & Nimgaonkar,V.L. A network of dopaminergic gene variations implicated as risk factors for schizophrenia. *Hum. Mol. Genet.* **17**, 747-758 (2008).

Tandon, R., Keshavan, M.S., & Nasrallah, H.A. Schizophrenia, "Just the Facts": what we know in 2008 part 1: overview. *Schizophr Res.* 100(1-3):4-19 (2008).

Taylor,M.S., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y., & Semple,C.A. Heterotachy in mammalian promoter evolution. *PLoS. Genet.* **2**, e30 (2006).

Teramitsu,I., Kudo,L.C., London,S.E., Geschwind,D.H., & White,S.A. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J. Neurosci.* **24**, 3152-3163 (2004).

Tienari,P., Sorri,A., Lahti,I., Naarala,M., Wahlberg,K.E., Ronkko,T., Pohjola,J., & Moring,J. The Finnish adoptive family study of schizophrenia. *Yale J. Biol. Med.* **58**, 227-237 (1985).

Tirosh,I., Weinberger,A., Bezalel,D., Kaganovich,M., & Barkai,N. On the relation between promoter divergence and gene expression evolution. *Mol. Syst. Biol.* **4**, 159 (2008).

Tochigi,M., Iwamoto,K., Bundo,M., Komori,A., Sasaki,T., Kato,N., & Kato,T. Methylation status of the reelin promoter region in the brain of schizophrenic patients. *Biol. Psychiatry* **63**, 530-533 (2008).

Toirac,I., Sanjuan,J., Aguilar,E.J., Gonzalez,J.C., Artigas,F., Rivero,O., Najera,C., Molto,M.D., & De Frutos,R. Association between CCK-AR gene and schizophrenia with auditory hallucinations. *Psychiatr. Genet.* **17**, 47-53 (2007).

Tolosa,A., Sanjuan,J., Leal,C., Costas,J., Molto,M.D., & De Frutos,R. Rapid evolving RNA gene HAR1A and schizophrenia. *Schizophr. Res.* **99**, 370-372 (2008).

Tomblin,J.B., Records,N.L., Buckwalter,P., Zhang,X., Smith,E., & O'Brien,M. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang Hear. Res.* **40**, 1245-1260 (1997).

Tomblin,J.B. & Buckwalter,P.R. Heritability of poor language achievement among twins. *J. Speech Lang Hear. Res.* **41**, 188-199 (1998).

Tsankova,N., Renthal,W., Kumar,A., & Nestler,E.J. Epigenetic regulation in psychiatric disorders. *Nat. Rev. Neurosci.* **8**, 355-367 (2007).

Tsuang, M. T. and Faraone, S. V. Schizophrenia. The Facts. 1997. Oxford University Press.

Tsutsumi,T., Holmes,S.E., McInnis,M.G., Sawa,A., Callahan,C., DePaulo,J.R., Ross,C.A., Delisi,L.E., & Margolis,R.L. Novel CAG/CTG repeat expansion mutations do not contribute to the genetic risk for most cases of bipolar disorder or schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **124B**, 15-19 (2004).

Vallender,E.J. & Lahn,B.T. Positive selection on the human genome. *Hum. Mol. Genet.* **13 Spec No 2**, R245-R254 (2004).

van Os,J., Hanssen,M., Bijl,R.V., & Ravelli,A. Strauss (1969) revisited: a psychosis continuum in the general population? *Schizophr. Res.* **45**, 11-20 (2000).

Vargha-Khadem,F., Watkins,K., Alcock,K., Fletcher,P., & Passingham,R. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc. Natl. Acad. Sci. U. S. A* **92**, 930-933 (1995).

- Vargha-Khadem,F., Watkins,K.E., Price,C.J., Ashburner,J., Alcock,K.J., Connelly,A., Frackowiak,R.S., Friston,K.J., Pembrey,M.E., Mishkin,M., Gadian,D.G., & Passingham,R.E. Neural basis of an inherited speech and language disorder. *Proc. Natl. Acad. Sci. U. S. A* **95**, 12695-12700 (1998).
- Verkerk,A.J., Pieretti,M., Sutcliffe,J.S., Fu,Y.H., Kuhl,D.P., Pizzuti,A., Reiner,O., Richards,S., Victoria,M.F., Zhang,F.P., & . Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905-914 (1991).
- Vernes,S.C., Nicod,J., Elahi,F.M., Coventry,J.A., Kenny,N., Coupe,A.M., Bird,L.E., Davies,K.E., & Fisher,S.E. Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum. Mol. Genet.* **15**, 3154-3167 (2006).
- Vernes,S.C., Spiteri,E., Nicod,J., Groszer,M., Taylor,J.M., Davies,K.E., Geschwind,D.H., & Fisher,S.E. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am. J. Hum. Genet.* **81**, 1232-1250 (2007).
- Vernes,S.C., Newbury,D.F., Abrahams,B.S., Winchester,L., Nicod,J., Groszer,M., Alarcon,M., Oliver,P.L., Davies,K.E., Geschwind,D.H., Monaco,A.P., & Fisher,S.E. A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.* **359**, 2337-2345 (2008).
- Vincent,J.B., Paterson,A.D., Strong,E., Petronis,A., & Kennedy,J.L. The unstable trinucleotide repeat story of major psychosis. *Am. J. Med. Genet.* **97**, 77-97 (2000).
- Vorstman,J.A., Staal,W.G., van Daalen,E., van Engeland,H., Hochstenbach,P.F., & Franke,L. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry* **11**, 1, 18-1, 28 (2006).
- Wall,J.D. & Pritchard,J.K. Haplotype blocks and linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **4**, 587-597 (2003).
- Walsh,T., McClellan,J.M., McCarthy,S.E., Addington,A.M., Pierce,S.B., Cooper,G.M., Nord,A.S., Kusenda,M., Malhotra,D., Bhandari,A., Stray,S.M., Rippey,C.F., Roccanova,P., Makarov,V., Lakshmi,B., Findling,R.L., Sikich,L., Stromberg,T., Merriman,B., Gogtay,N., Butler,P., Eckstrand,K., Noory,L., Gochman,P., Long,R., Chen,Z., Davis,S., Baker,C., Eichler,E.E., Meltzer,P.S., Nelson,S.F., Singleton,A.B., Lee,M.K., Rapoport,J.L., King,M.C., & Sebat,J. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539-543 (2008).
- Warburton,P., Baird,G., Chen,W., Morris,K., Jacobs,B.W., Hodgson,S., & Docherty,Z. Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. *Am. J. Med. Genet.* **96**, 228-234 (2000).
- Warner R. and Girolamo G. Epidemiología de los trastornos mentales y de los problemas psicosociales:esquizofrenia. 1995. Organización Mundial de la Salud. Ginebra. Madrid.
- Wassink,T.H., Piven,J., Vieland,V.J., Pietila,J., Goedken,R.J., Folstein,S.E., & Sheffield,V.C. Evaluation of FOXP2 as an autism susceptibility gene. *Am. J. Med. Genet.* **114**, 566-569 (2002).
- Watkins, K. Hearing voices: A common human experience. 1998. Hill of content. Melbourne.

Bibliografía

- Watkins,K.E., Dronkers,N.F., & Vargha-Khadem,F. Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain* **125**, 452-464 (2002a).
- Watkins,K.E., Vargha-Khadem,F., Ashburner,J., Passingham,R.E., Connelly,A., Friston,K.J., Frackowiak,R.S., Mishkin,M., & Gadian,D.G. MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain* **125**, 465-478 (2002b).
- Wayland, M., Khaitovich, P., Ryan, M., Paabo, S. & Bahn, S. The molecular basis of cognitive impairment on schizophrenia: evidence from comparative transcriptomics. *Schizophr. Res.* **81**, s17 (2006).
- Webb,D.M. & Zhang,J. FoxP2 in song-learning birds and vocal-learning mammals. *J. Hered.* **96**, 212-216 (2005).
- Wei,J. & Hemmings,G.P. The CCK-A receptor gene possibly associated with auditory hallucinations in schizophrenia. *Eur. Psychiatry* **14**, 67-70 (1999).
- Weiser,M., van Os,J., & Davidson,M. Time for a shift in focus in schizophrenia: from narrow phenotypes to broad endophenotypes. *Br. J. Psychiatry* **187**, 203-205 (2005).
- Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
- Wilke, M., Kaufmann, C., Grabner, A., Pütz, B., Wetter, T.C., & Auer, D.P. Gray matter-changes and correlates of disease severity in schizophrenia: a statistical parametric mapping study. *Neuroimage* 13(5):814-24 (2001)
- Williams,N.M., Norton,N., Williams,H., Ekholm,B., Hamshere,M.L., Lindblom,Y., Chowdari,K.V., Cardno,A.G., Zammit,S., Jones,L.A., Murphy,K.C., Sanders,R.D., McCarthy,G., Gray,M.Y., Jones,G., Holmans,P., Nimgaonkar,V., Adolfson,R., Osby,U., Terenius,L., Sedvall,G., O'Donovan,M.C., & Owen,M.J. A systematic genomewide linkage study in 353 sib pairs with schizophrenia. *Am. J. Hum. Genet.* **73**, 1355-1367 (2003).
- Williamson,S.H., Hubisz,M.J., Clark,A.G., Payseur,B.A., Bustamante,C.D., & Nielsen,R. Localizing recent adaptive evolution in the human genome. *PLoS. Genet.* **3**, e90 (2007).
- Wilmont,G.R. & Warren,S.T. in Genetic instabilities and hereditary neurological diseases. ed. Warren ST,W.R. (San Diego:Academic Press,1998).
- Woodruff,P.W. Auditory hallucinations: Insights and questions from neuroimaging. *Cognit. Neuropsychiatry* **9**, 73-91 (2004).
- Wright,A.F. & Hastie,N.D. Complex genetic diseases: controversy over the Croesus code. *Genome Biol.* **2**, COMMENT2007 (2001).
- Wyatt,R.J., Alexander,R.C., Egan,M.F., & Kirch,D.G. Schizophrenia, just the facts. What do we know, how well do we know it? *Schizophr. Res.* **1**, 3-18 (1988).
- Yeargin-Allsopp,M., Rice,C., Karapurkar,T., Doernberg,N., Boyle,C., & Murphy,C. Prevalence of autism in a US metropolitan area. *JAMA* **289**, 49-55 (2003).
- Zhang,J., Webb,D.M., & Podlaha,O. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* **162**, 1825-1835 (2002).
- Zhao,H., Pfeiffer,R., & Gail,M.H. Haplotype analysis in population genetics and association studies. *Pharmacogenomics.* **4**, 171-178 (2003).

Zhou,B., Zhong,Q., Minoo,P., Li,C., Ann,D.K., Frenkel,B., Morrisey,E.E., Crandall,E.D., & Borok,Z. Foxp2 inhibits Nkx2.1-mediated transcription of SP-C via interactions with the Nkx2.1 homeodomain. *Am. J. Respir. Cell Mol. Biol.* **38**, 750-758 (2008).