

**ANÁLISIS DE LA VARIABILIDAD GENÉTICA DE LOS GENES *MUC* Y
TFF DE PROTECCIÓN DE LA MUCOSA GÁSTRICA EN RELACIÓN AL
CÁNCER GÁSTRICO Y LAS LESIONES PRECURSORAS.**

**TESIS DOCTORAL
Fátima Marín Nieto
2013**

Análisis de la variabilidad genética de los genes MUC y TFF de protección de la mucosa gástrica en relación al cáncer gástrico y las lesiones precursoras.

Memoria presentada por
Fátima Marín Nieto

para optar al grado de Doctor por
la Universitat Autònoma de Barcelona

Este trabajo se ha realizado bajo la dirección de la Dra. Núria Sala Serra en el Laboratori de Recerca Translacional y la Unitat de Nutrició, Ambient i Càncer del Programa de Recerca en Epidemiologia del Càncer en el Insitut Català d'Oncologia (ICO)-Institut d'Investigació Biomèdica de Bellvitge (IDIBELL

Tesis adscrita al Departament de Genètica i Microbiologia de la Facultat de Biociències de la Universitat Autònoma de Barcelona.

Tutor: Dr. Ricard Marcos Dauder

Dra. Núria Sala i Serra

Fátima Marín Nieto

agradecimientos

En primer lugar, quiero agradecer a Nria por darme la oportunidad de incorporarme en su grupo e introducirme en el mundo de la investigacin. Gracias por la confianza que siempre has depositado en m y por guiarme hasta aqu.

Gracias a todas las personas que han formado parte del grupo GEM a lo largo de estos aos. A Nadia muchas gracias por haber estado ah siempre, desde el primer hasta el ltimo instante, y ofrecerme tu apoyo. A Begoa por acogerme, por ensearme y compartir tantos momentos de trabajo y tambin de distensin. Gracias por tu entusiasmo. Tambin a Xavi, por tu ayuda e incorregible sentido del humor.

Agradecerle a Carlos Alberto y a todos los miembros de la unidad de Nutricin, Ambiente y Cncer que han hecho posible la realizacin de este trabajo. Gracias Cati por tu asesoramiento estadstico que siempre es muy didctico.

Y, como no, dar las gracias, a todos los compaeros, los que estn y los que ya se han ido, del LRT, de Diagnstico del ICO y del Departamento de Gentica y del COM del IDIBELL. Gracias a Gemma, Sandra, Ruth, Carol, EVA LT, Eva Pros, Eva Serrano, Juanjo; Montse, Vane.... y tantsimas otras personas, imposible mencionarlas a todas, con las que he compartido infinidad de momentos durante las largas jornadas en el laboratorio.

Gracias a todos mis amigos y esos momentos de distraccin cuando los necesitaba. Gracias Beln por entenderme y compartir momentos de estrs y tambin de ocio.

Y, especialmente, gracias Xavo por ser paciente conmigo, animarme y haber recorrido todo este camino a mi lado. Tambin gracias a tu familia por su respaldo. I finalmente, y sobretodo, gracias a mi familia por su apoyo. Raquel agradezco tu tranquilidad y serenidad que siempre me hacen tanta falta, y a tus nios que hacen que me olvide de todo y me llenan de alegra. Y gracias a mis padres por estar siempre a mi lado, ser comprensivos y respetar y respaldar todas las decisiones que me han llevado hasta aqu. A vosotros va dedicada esta tesis.

A mi familia

Al Xavo

ÍNDICE

ÍNDICE DE TABLAS Y FIGURAS	1
ABREVIATURAS Y SÍMBOLOS.....	7
PRESENTACIÓN	9
INTRODUCCIÓN	13
1. CÁNCER GÁSTRICO	15
1.1. Incidencia y mortalidad	15
1.2. Anatomía e histología del estómago	16
1.3. Clasificación histológica y anatómica.....	17
1.3.1. Adenocarcinoma de tipo intestinal	18
1.3.2. Adenocarcinoma de tipo difuso	19
1.3.2.1 Cáncer gástrico difuso hereditario	20
1.3.3. Subtipos anatómicos del adenocarcinoma gástrico	20
1.4. Carcinogénesis gástrica y lesiones precursoras.....	21
1.4.1. Gastritis crónica activa	21
1.4.2. Gastritis atrofica.....	22
1.4.3. Metaplasia intestinal	22
1.4.4. Displasia.....	24
1.5. Etiología	24
1.5.1. <i>Helicobacter pylori</i>	25
1.5.1.1 Colonización de la mucosa gástrica	25
1.5.1.2 Factores de virulencia.....	26
1.5.1.3 Respuesta inmunitaria e inflamatoria a la infección	28
1.5.2. Otros factores ambientales	29
1.5.3. Susceptibilidad genética.....	30
1.5.3.1 Genes candidatos	31
1.5.3.2 GWAs.....	33
2. TREFOIL FACTOR FAMILY	35
2.1. Estructura de los péptidos trefoil.....	35
2.2. Estructura de los genes <i>TFF</i>	36
2.3. Expresión de los TFFs	37
2.3.1. Localización tisular.....	37
2.3.2. Regulación de la expresión.....	38
2.4. Funciones de los TFFs en el mantenimiento de la mucosa	40
2.5. TFFs y la carcinogénesis gástrica	43
2.5.1. Expresión de TFFs en condiciones patológicas.....	43
2.5.2. Papel de los TFFs en la carcinogénesis gástrica	44
2.5.3. Variabilidad genética en TFFs y la carcinogénesis gástrica	45

3.	MUCINAS.....	45
3.1.	Estructura de las mucinas	45
3.1.1.	Mucinas de membrana	46
3.1.2.	Mucinas secretadas	48
3.2.	Estructura de los genes <i>MUCs</i>	49
3.3.	Regulación de la expresión	51
3.4.	Funciones de las mucinas	53
3.5.	Papel y expresión de mucinas en la carcinogénesis gástrica	56
3.5.1.	Mucinas en lesiones precursoras.....	56
3.5.1.1	Gastritis inducida por <i>H.pylori</i>	56
3.5.1.2	Metaplasia intestinal.....	56
3.5.2.	Mucinas en cáncer gástrico.....	57
3.5.3.	Variabilidad genética de mucinas y carcinogénesis gástrica	59
	OBJETIVOS	61
	MATERIALES Y MÉTODOS	65
1.	PACIENTES Y CONTROLES	67
1.1.	Muestras procedentes de un estudio español de seguimiento de pacientes con lesiones gástricas preneoplásicas	67
1.2.	Muestras procedentes de un estudio caso-control de cáncer gástrico anidado en la cohorte europea EPIC	69
2.	BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICAS	70
2.1.	Selección de polimorfismos.....	70
2.1.1.	Selección de polimorfismos potencialmente funcionales.....	70
2.1.2.	Selección de TagSNPs a partir de los datos de HapMap.....	71
2.2.	Selección de miRNAs.....	73
2.3.	Análisis de secuencias	74
3.	EXTRACCIÓN Y CUANTIFICACIÓN DE DNA.	75
3.1.	Extracción de DNA genómico a partir de sangre periférica	75
3.2.	Extracción de DNA genómico a partir de tejido incluido en parafina	75
3.3.	Cuantificación del DNA.....	75
4.	REACCIÓN EN CADENA DE LA DNA POLIMERASA (PCR).....	76
4.1.	Purificación de productos de PCR	81
5.	SECUENCIACIÓN DEL DNA	81
6.	ELECTROFORESIS	82
6.1.	Electroforesis en geles de agarosa	82
6.1.1.	Purificación del DNA de geles de agarosa.....	82
6.2.	Electroforesis en geles de poliacrilamida	83
7.	ANÁLISIS DE POLIMORFISMOS.....	83

7.1.	Análisis de las variantes con enzimas de restricción	84
7.2.	Polimorfismo de conformación de cadena sencilla y heterodúplex (SSCP/HD)	84
7.3.	Genotipado de polimorfismos mediante PCR a tiempo real	85
7.3.1.	Genotipado con sondas de hibridación FRET	85
7.3.2.	Genotipado con sondas de hidrólisis TaqMan®	88
7.3.3.	Análisis de las variantes polimórficas con <i>High Resolution Melting</i> (HRM)	90
7.4.	Genotipado de polimorfismos en plataformas de alto rendimiento.....	92
7.4.1.	MassArray™ (iPlex-Gold) de Sequenom.....	92
7.4.2.	Illumina GoldenGate	93
8.	CLONAJE DE DNA DENTRO DE UN PLÁSMIDO	94
8.1.	Vectores utilizados en la clonación	94
8.2.	Obtención del inserto y del vector	95
8.3.	Clonaje mediante extremos cohesivos.....	97
8.3.1.	Extremos cohesivos generados por digestión con una sola enzima de restricción .	97
8.3.2.	Clonaje direccional.....	98
8.4.	Ligación.....	98
9.	TRANSFORMACIÓN DE CÉLULAS COMPETENTES.....	99
10.	COMPROBACIÓN DE LOS PLÁSMIDOS RECOMBINANTES	99
11.	MUTAGÉNESIS DIRIGIDA	100
12.	EXTRACCIÓN DEL DNA PLASMÍDICO DE BACTERIAS.....	102
12.1.	Minipreparación	102
12.2.	Midi y Maxipreparación.....	103
13.	ENSAYOS CON LÍNEAS CELULARES	103
13.1.	Cultivo y mantenimiento de líneas celulares	103
13.2.	Transfecciones transitorias	104
13.3.	Ensayos luciferasa.....	105
14.	ANÁLISIS ESTADÍSTICOS	106
RESULTADOS		109
1.	VARIABILIDAD GENÉTICA DE <i>TFFs</i> Y <i>MUCs</i>	111
1.1.	Selección de SNPs potencialmente funcionales descritos en las bases de datos.	111
1.2.	Validación de SNPs potencialmente funcionales e identificación de nuevas variantes.....	112
1.2.1.	<i>TFF1</i>	112
1.2.2.	<i>TFF2</i>	114
1.2.3.	<i>TFF3</i>	117
1.2.4.	<i>MUC1</i>	119
1.3.	Análisis de desequilibrio de ligamiento, identificación de haplotipos y selección de tagSNPs	121

1.3.1.	<i>TFFs</i>	121
1.3.2.	<i>MUCs</i>	123
1.4.	Análisis funcional del efecto de polimorfismos en <i>TFFs</i>	127
1.4.1.	Análisis funcional <i>in silico</i>	127
1.4.1.1	Predicción de sitios diana de miRNAs en la región 3'UTR de <i>TFFs</i>	129
1.4.2.	Análisis funcional de los polimorfismos de la región 3'UTR de <i>TFFs</i>	131
1.4.2.1	Efecto de los polimorfismos en la expresión de la luciferasa	131
1.4.2.2	Interacción de miRNAs con <i>TFFs</i> en la región 3'UTR.....	133
1.5.	Polimorfismos seleccionados para los estudios de asociación	134
2.	ASOCIACIÓN DE POLIMORFISMOS DE <i>TFFs</i> Y <i>MUCs</i> CON LA CARCINOGENÉISIS GÁSTRICA.....	138
2.1.1.	Descriptiva de la población y asociación de covariables con la evolución de lesiones	138
2.1.2.	Genotipado de los polimorfismos.....	142
2.1.4.	Asociación polimorfismos con lesiones precursoras	147
2.1.4.1	<i>TFFs</i>	147
2.1.4.2	<i>MUCs</i>	151
2.1.4.3	Genes involucrados en la respuesta a la infección por <i>H.pylori</i>	156
2.2.	Estudio caso-control de cáncer gástrico anidado en la cohorte EPIC	161
2.2.1.	Descriptiva de la población y asociación de covariables con cáncer gástrico	161
2.2.2.	Genotipado de los polimorfismos.....	165
2.2.3.	Análisis de LD y determinación de haplotipos.....	166
2.2.4.	Asociación polimorfismos con cáncer gástrico	171
2.2.4.1	<i>TFFs</i>	171
2.2.4.2	<i>MUCs</i>	180
2.2.5.	Asociación de haplotipos con cáncer gástrico	188
	DISCUSIÓN	193
	CONCLUSIONES	221
	BIBLIOGRAFIA	225
	APÉNDICE I: PROTOCOLOS	249
	APÉNDICE II: TABLAS	277
	APÉNDICE III: FIGURAS	311
	APÉNDICE IV: PUBLICACIONES	319

Índice de tablas

Tabla M-1. Cebadores utilizados para amplificar el DNA genómico de los genes TFF.....	78
Tabla M-2. Cebadores utilizados para amplificar el DNA genómico del gen <i>MUC1</i>	79
Tabla M-3. Condiciones específicas para la amplificación de diferentes fragmentos de gDNA de los genes <i>TFF1</i> , <i>TFF2</i> , <i>TFF3</i> , <i>MUC1</i>	80
Tabla M-4. Polimorfismos analizados utilizando enzimas de restricción	84
Tabla M-5. Cebadores, sondas FRET y condiciones de amplificación y determinación de la curva de fusión en el genotipado de SNPs con el LC480	87
Tabla M-6. Polimorfismos en genes de respuesta inflamatoria genotipados mediante el uso de sondas TaqMan en el termociclador a tiempo real, ABI 7900HT.	89
Tabla M-7. Cebadores utilizados en la clonación de la región 3'UTR de TFFs y de miRNAs y del promotor de TFF2.....	96
Tabla M-8. Cebadores utilizados para la comprobación de los plásmidos recombinantes en la clonación de la región 3'UTR de TFFs, miRNAs y el promotor de TFF2.	100
Tabla M-9. Cebadores específicos para los vectores utilizados en este trabajo	100
Tabla M-10. Cebadores diseñados para los experimentos de mutagénesis dirigida	102
Tabla R-1. Validación de SNPs seleccionados y nuevas variantes identificadas en TFFs.	118
Tabla R-2. Selección de tagSNPs en el clúster de TFFs	123
Tabla R-3. Selección de tagSNPs en el clúster de MUCs	125
Tabla R-4. Efecto del SNP c.-450G>T en la interacción de NFκB y Sp1 según regSNP.	127
Tabla R-5. Selección de miRNAs en la región 3'UTR polimórfica de los genes TFF.....	130
Tabla R-6. Características generales de todos los polimorfismos incluidos en esta tesis en los genes MUCs, TFFs, ILs y genes relacionados con la infección por <i>H.pylori</i>	135
Tabla R-7. Principales características y diagnóstico histopatológico de pacientes con lesiones gástricas en el inicio del reclutamiento.....	138
Tabla R-8. Principales características y evolución de los pacientes con lesiones gástricas (N=453).	139
Tabla R-9. Asociación de posibles factores de riesgo con la evolución de las lesiones gástricas	141
Tabla R-10. Asociación del SNP rs225358 de TFF1 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por <i>H.pylori</i>	148
Tabla R-11. Asociación de los tagSNPs de la región genómica a 5' de TFF1 con la regresión de lesiones gástricas precursoras de cáncer y estratificación por consumo de AINES.	150
Tabla R-12. Asociación de los SNPs de MUC2 rs10794293, rs3924453 y rs4077759 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por <i>H.pylori</i>	152

Tabla R-13. Asociación de los SNPs de <i>MUC2</i> rs10902073, rs10794281, rs2071174 y 7944723 con la regresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por <i>H.pylori</i>	153
Tabla R-14. Asociación de los SNPs de <i>MUC2</i> rs10902073, rs10794281 y rs2071174 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por <i>H.pylori</i>	155
Tabla R-15. Asociación de los SNPs de <i>PTPN11</i> , <i>TNF</i> , <i>CD14</i> , <i>TLR4</i> y <i>CDH1</i> asociados con la evolución de lesiones gástricas.	157
Tabla R-16. Haplotipos de <i>TFFs</i> y asociación con la evolución de lesiones gástricas.	158
Tabla R-17. Haplotipos de la región genómica a 5' de <i>TFF1</i> y asociación con la evolución de lesiones gástricas	159
Tabla R-18. Haplotipos de genes <i>MUC</i> y asociación con la evolución de lesiones gástricas.	160
Tabla R-19. Características principales de la población caso-control de cáncer gástrico.	162
Tabla R-20. Asociación de los factores de riesgo con cáncer gástrico	163
Tabla R-21. Descriptiva de las principales características de los casos de cáncer gástrico según la localización anatómica y tipo histológico.	164
Tabla R-22. SNPs en el gen <i>TFF1</i> que mostraron asociación con cáncer gástrico y/o con alguno de sus subtipos histológicos.	173
Tabla R-23. SNPs en el gen <i>TFF3</i> que mostraron asociación con cáncer gástrico y/o con sus subtipos histológicos.	177
Tabla R-24. SNPs en el gen <i>TFF3</i> que mostraron asociación con los subtipos anatómicos de cáncer gástrico.	179
Tabla R-25. SNPs en <i>MUC1</i> que mostraron asociación significativa con cáncer gástrico y sus subtipos histológicos.	181
Tabla R-26. SNPs en <i>MUC1</i> que mostraron asociación significativa con los subtipos anatómicos de cáncer gástrico.	182
Tabla R-27. SNPs en <i>MUC2</i> que mostraron asociación significativa con cáncer gástrico y sus subtipos histológicos.	184
Tabla R-28. Asociación de los SNPs en <i>MUC2</i> que mostraron asociación significativa con los subtipos anatómicos de cáncer gástrico.	185
Tabla R-29. Asociación de los SNPs en <i>MUC6</i> que mostraron asociación significativa con cáncer gástrico en la estratificación por sexo.	187
Tabla R-30. Haplotipos en los genes <i>TFF</i> y asociación con cáncer gástrico y sus subtipos histológicos.	189
Tabla R-31. Haplotipos en los genes <i>MUC</i> y asociación con cáncer gástrico y sus subtipos anatómicos.	191
Tabla A-1. Polimorfismos potencialmente funcionales descritos en las bases de datos en el momento de su selección (dbSNP v126).	279
Tabla A-2. Principales características de la población de pacientes con lesiones gástricas (N=387).	281

Tabla A-3. Resultados del genotipado de los polimorfismos seleccionados para el estudio de seguimiento de lesiones precursoras de cáncer gástrico.	281
Tabla A-4. Valores p de significación de la asociación entre polimorfismos y evolución de lesiones gástricas bajo los diferentes modelos de herencia.	284
Tabla A-5. Asociación de los SNPs de <i>MUC2</i> rs10794293, rs3924453, rs4077759, rs10794281 y rs2071174 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por el consumo de AINES.	286
Tabla A-6. Asociación de los SNPs de <i>MUC2</i> rs10794293 y rs4077759 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por el por sexo.	287
Tabla A-7. Asociación de los SNPs de <i>MUC2</i> rs10794281, rs2071174 y 7944723 con la regresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por <i>H.pylori</i>	288
Tabla A-8. Descriptiva de la edad, consumo de alcohol y dieta de la población caso-control de cancer gástrico.	289
Tabla A-9. Resultados del genotipado de los polimorfismos en los genes <i>MUC</i> y <i>TFF</i> en la población caso-control de cáncer gástrico.	290
Tabla A-10. Valores p de significación de la asociación entre polimorfismos en genes <i>MUC</i> y <i>TFF</i> con cáncer gástrico y su subtipo histológico bajo los diferentes modelos de herencia. ..	291
Tabla A-11. Valores p de significación de la asociación entre polimorfismos en genes <i>MUC</i> y <i>TFF</i> con los subtipos anatómicos de cáncer gástrico bajo los diferentes modelos de herencia.	293
Tabla A-12. SNPs en genes <i>TFF</i> que mostraron asociación significativa con cáncer gástrico estratificando por sexo.	294
Tabla A-13. SNPs de <i>TFF1</i> y región adyacente que mostraron asociación con cáncer gástrico de tipo intestinal estratificando por sexo.	296
Tabla A-14. Asociación de SNPs en genes <i>TFF</i> que mostraron asociación significativa con cáncer gástrico difuso estratificando por sexo.	298
Tabla A-15. Asociación de los SNPs en genes <i>TFF</i> que mostraron asociación significativa con cáncer gástrico no cardias estratificando por sexo.	299
Tabla A-16. SNPs de <i>TFF3</i> que mostraron asociación significativa con cáncer gástrico del cardias estratificando por sexo.	300
Tabla A-17. SNPs de <i>TFFs</i> que mostraron asociación con cáncer gástrico no cardias, estratificando por la infección por <i>H.pylori</i>	301
Tabla A-18. SNPs en genes <i>MUC</i> que mostraron asociación con cáncer gástrico estratificando por sexo.	303
Tabla A-19. SNPs en genes <i>MUC</i> que mostraron asociación con cáncer gástrico del tipo intestinal estratificando por sexo.	304
Tabla A-20. SNPs en <i>MUC2</i> que mostraron asociación significativa con cáncer gástrico del cardias en la estratificación por sexo.	305
Tabla A-21. SNPs en genes <i>MUC</i> que mostraron asociación significativa con cáncer gástrico no cardias en la estratificación por la infección por <i>H.pylori</i>	306

Tabla A-22. SNPs en genes <i>MUC</i> que mostraron asociación significativa con cáncer gástrico no cardias en los infectados por <i>H. pylori</i> cuando se estratificó por la presencia de CagA...	307
Tabla A-23. Haplotipos en genes <i>TFFs</i> y asociación con los subtipos anatómicos de cancer gástrico.....	309
Tabla A-24. Haplotipos en genes <i>MUC</i> y asociación con los subtipos histológicos de cancer gástrico.....	310

Índice de figuras

Figura I-1. A. Anatomía del estómago. B. Histología de las glándulas gástricas.....	16
Figura I-2. Etapas de la carcinogénesis gástrica	19
Figura I-3. Estructura de los péptidos trefoil (Thim 1997, (175)).....	36
Figura I-4. Esquema de la estructura proteica de mucinas	47
Figura I-5. Esquema de la estructura de los genes de mucina.....	50
Figura M-1. Diagrama de flujo de selección de pacientes participantes en el estudio observacional longitudinal de lesiones preneoplásicas gástricas en la población de Soria.	67
Figura M-2. Índice de Correa para la clasificación de lesiones preneoplásicas gástricas	68
Figura M-3. Distribución a lo largo de los genes <i>TFF1</i> , <i>TFF2</i> , <i>TFF3</i> y <i>MUC1</i> de los cebadores diseñados	77
Figura M-4. A. Transmisión de Energía Fluorescente, FRET. B. Análisis de un polimorfismo mediante sondas de hibridación FRET en el LC480.....	86
Figura M-5. Funcionamiento de las sondas de hidrólisis TaqMan.....	88
Figura M-6. A. Ejemplo de gráficos de amplificación. B. Diagrama de puntos de un experimento de genotipado mediante sondas TaqMan®.	88
Figura M-7. Análisis de una variante mediante HRM.	91
Figura R-1. Digestión enzimática con <i>BsrDI</i> de <i>TFF1</i> c.-2C>T.....	113
Figura R-2. A. SSCPs del exón 1 de <i>TFF1</i> . B. Secuencia correspondiente al heterocigoto del SNP c.-29G>A, cadena reversa.....	113
Figura R-3. A. SSCPs del exón 3 de <i>TFF1</i> . B. Secuencias correspondientes a los polimorfismos detectados en la figura A.....	114
Figura R-4. A. SSCPs de la región 5' del gen <i>TFF2</i> . B. Secuencias correspondientes al SNP c.-450G>T.	114
Figura R-5. A. SSCPs del exón 1 de <i>TFF2</i> . B. Secuencia correspondiente al heterocigoto del SNP c.20A>C.	115
Figura R-6. A. SSCPs del exón 4 de <i>TFF2</i> . B. Secuencias correspondientes a los polimorfismos detectados en la figura A.....	115
Figura R-7. Análisis mediante HRM de los polimorfismos c.*68_69insCTT y c.*82T>C de <i>TFF2</i> . A. Representación gráfica de la de curva de fusión	116

Figura R-8. A. SSCPs del exón 2 del gen <i>TFF2</i> . B. Secuencias correspondientes a los polimorfismos detectados en la figura A.....	116
Figura R-9. A. Digestión enzimática con <i>Avall</i> de c.-3A>G, en el exón 1 de <i>TFF3</i>	117
Figura R-10. A. SSCPs del exón 1 de <i>TFF3</i> . B. SSCPs del exón 3 de <i>TFF3</i> . C. Secuencia correspondiente al heterocigoto del SNP c.*41G>A.	118
Figura R-11. A. SSCPs del exón 2 de <i>TFF3</i> . B. Secuencia correspondiente al heterocigoto del SNP c.238G>A.	118
Figura R-12. SSCPs de los fragmentos correspondientes a los exones 1, 3, 5 y 8 de <i>MUC1</i>	119
Figura R-13. A. SSCPs del exón 2 de <i>MUC1</i> . B. Secuencia correspondiente al SNP c.93G>A. 1- Homocigoto CC, cadena reversa. 2- Homocigoto TT, cadena reversa.....	120
Figura R-14. A. Perfil de la curva de fusión del HRM del exón 7 de <i>MUC1</i> . B. Perfiles de la curva de fusión del HRM del exón 2 de <i>MUC1</i>	120
Figura R-15. Desequilibrio de ligamiento en el clúster de <i>TFFs</i> en la región genómica 21q22.3, según los datos obtenidos para población CEU en HapMap.	122
Figura R-16. Representación gráfica del LD y bloques de haplotipos seleccionados en la región de <i>MUC1</i> , en 1q21, según los datos obtenidos para población CEU en HapMap.	124
Figura R-17. A. Representación gráfica del LD y bloques de haplotipos seleccionados en la región de <i>MUC6</i> , en 11p15, según los datos obtenidos para población CEU en HapMap.....	124
Figura R-18. A. Representación gráfica del LD y bloques de haplotipos seleccionados en la región 11p15 que incluye <i>MUC2</i> , según los datos obtenidos para población CEU en HapMap.....	126
Figura R-19. A. Alineamiento múltiple de la secuencia proteica de <i>TFF3</i> de diferentes especies. La primera fila muestra la secuencia de <i>TFF3</i> en humanos. B. Estructura en 3D de la forma dimérica del péptido <i>TFF3</i>	128
Figura R-19. A. Alineamiento múltiple de la secuencia nucleotídica de la región 3'UTR de <i>TFF1</i> de diferentes especies. B. Alineamiento múltiple de las secuencias nucleotídicas de las regiones 3'UTR de los tres genes <i>TFF</i>	129
Figura R-20. Análisis de la interacción del miRNA hsa-miR-615-3p con la región 3'UTR de <i>TFFs</i> donde se localiza el SNP c.*41G>A.	131
Figura R-21. Efecto de los SNPs en 3'UTR de genes <i>TFF</i> , en la expresión de la luciferasa en células HEK-293T. L	132
Figura R-22. Efecto de los SNPs en 3'UTR de genes <i>TFF</i> , en la expresión de la luciferasa en células AGS, de adenocarcinoma gástrico.	132
Figura R-23. Efecto de los miRNAs, en función de los SNPs en 3'UTR de los genes <i>TFF</i> , en la expresión de la luciferasa en células HEK-293T.	134
Figura R-24. Comparativas de los resultados de genotipado del SNP rs225334 de <i>TFF2</i> , mediante distintas tecnologías. S	143
Figura R-25. Análisis del desequilibrio de ligamiento entre polimorfismos de las regiones en 21q22.3 donde se localizan <i>TFF1</i> , <i>TFF2</i> y <i>TFF3</i> , a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico.....	144

Figura R-26. Análisis de desequilibrio de ligamiento entre polimorfismos de las regiones donde se localizan <i>MUC1</i> , en 1q21, y <i>MUC2</i> y <i>MUC6</i> , en 11p15, a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico.	146
Figura R-27. Análisis de desequilibrio de ligamiento entre polimorfismos de las regiones en 21q22.3 donde se localizan <i>TFF1</i> , <i>TFF2</i> y <i>TFF3</i> , a partir de los datos de genotipado en la población caso-control de cáncer gástrico.	167
Figura R-28. Análisis de desequilibrio de ligamiento entre polimorfismos de las regiones donde se localizan <i>MUC1</i> , en 1q21, y <i>MUC2</i> y <i>MUC6</i> , en 11p15, a partir de los datos de genotipado la población caso-control de cáncer gástrico.	169
Figura D-1. Regiones genómicas que están sujetas a una alteración funcional por causa de una sustitución nucleotídica.	197
Figura A-1A. Representación gráfica del coeficiente D' de LD en la región 21q22.3 donde se localizan <i>TFF1</i> , <i>TFF2</i> y <i>TFF3</i> , calculada a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico.	313
Figura A-1B. Representación gráfica del coeficiente r^2 de LD en la región 21q22.3 donde se localizan <i>TFF1</i> , <i>TFF2</i> y <i>TFF3</i> , calculada a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico. E.....	314
Figura A-2. Representación gráfica de los coeficientes D' y r^2 de LD entre los 14 SNPs localizados en el gen <i>MUC2</i> y regiones adyacentes, en 11p15.	315
Figura A-3A. Representación gráfica del coeficiente D' de LD en la región 21q22.3 donde se localizan <i>TFF1</i> , <i>TFF2</i> y <i>TFF3</i> , calculada a partir de los datos de genotipado en la población caso-control de cáncer gástrico. E.....	316
Figura A-3B. Representación gráfica del coeficiente r^2 de LD en la región 21q22.3 donde se localizan <i>TFF1</i> , <i>TFF2</i> y <i>TFF3</i> , calculada a partir de los datos de genotipado en la población caso-control de cáncer gástrico..	317
Figura A-4. Representación gráfica de los coeficientes D' y r^2 de LD entre los 12 SNPs localizados en el gen <i>MUC2</i> y regiones adyacentes, en 11p15. E	318

Abreviaturas y símbolos

AGS	Células epiteliales adherentes derivadas de un adenocarcinoma gástrico
AINES	antiinflamatorios no esteroideos
AP1	factor de transcripción, del inglés, <i>activator protein 1</i>
cagA	del inglés <i>cytotoxin-associated gene A</i> , cepa <i>H.pylori</i>
CDX	del inglés <i>caudal-related homeobox</i>
CD14	antígeno CD14 (del inglés <i>cluster of differentiation</i>)
CDH1	cadherina tipo 1 o E-cadherina
C/EBP β	del inglés <i>CCAAT/enhancer-binding protein beta</i>
CEU	Población de residentes americanos con ascendencia europea
CK	del inglés <i>cystein-knot</i> , dominio de MUCs
COX-2	ciclooxigenasa o prostaglandina-endoperóxido sintasa
EC	células <i>enterochromaffin</i>
ECL	células <i>enterochromaffin-like</i>
EPIC	del inglés <i>European Prospective Investigation into Cancer and Nutrition</i>
D	dominio de dimerización de MUCs
D'	medida de desequilibrio de ligamiento
dNTP	desoxinucleótido trifosfato
EGF	del inglés <i>epidermal growth factor</i>
ER	receptor del estrógeno
ERCC	del inglés <i>excision repair cross-complementing</i>
ERE	elemento de respuesta a estrógeno
FRET	del inglés <i>fluorescence resonance energy transfer</i>
GATA6	factor de transcripción, <i>GATA binding protein 6</i>
GCSI	del inglés <i>globet cell silencer inhibitor</i>
GWA	del inglés <i>genome-wide association study</i>
<i>H.pylori</i>	<i>Helicobacter pylori</i>
HEK-293T	Células embrionarias de riñón, del inglés <i>Human Embryonic Kidney 293</i>
HIF1	del inglés <i>hypoxia inducing factor 1</i>
HRM	del inglés <i>high resolution melting</i>
IC	intervalo de confianza
IL	interleucina
IM	metaplasia intestinal
Kb, pb	kilobase, par de bases
KGF	del inglés <i>keratinocyte growth factor</i>
LD	desequilibrio de ligamiento
Le ^b	antígeno Lewis b

LTA	del inglés <i>lymphotoxin alpha</i>
MAF	del inglés <i>minor allele frequency</i> , frecuencia alélica del alelo menor
MAG	atrofia gástrica multifocal
miRNA	microRNA
MMP	del inglés <i>matrix metalloproteinase</i>
MUC	mucina
NAG	gastritis no atrófica
NAT	del inglés <i>natural antisense transcripts</i>
NGX6	del inglés <i>nasopharyngeal carcinoma-associated gene 6</i>
NOD	del inglés, <i>nucleotide-binding oligomerization domain</i>
NCBI	del inglés <i>National Center for Biotechnology Information</i>
NFκB	del inglés <i>nuclear factor of kappa light polypeptide gene enhancer in B-cells</i>
NLR	del inglés, <i>NOD-like receptor</i>
OR	del inglés <i>odd ratio</i>
PTPN11	del inglés <i>protein tyrosine phosphatase, non-receptor type 11</i> , también SHP2
PCR	reacción en cadena de la polimerasa
PSCA	del inglés <i>prostate stem cell-antigen</i>
r ²	medida de desequilibrio de ligamiento
RFLP	del inglés <i>restriction fragment length polymorphism</i>
SEA	del inglés <i>sea urchin sperm protein, enterokinase, and agrin</i> , dominio de MUCs
SHP2	del inglés <i>Src homology region 2 domain-containing phosphatase-2</i> , ver PTPN11
SNP	polimorfismo de cambio de nucleótido, del inglés <i>single nucleotide polymorphism</i>
Sp1/2/3	factor de transcripción, del inglés <i>specificity protein-1/2/3</i> .
SSCP	del inglés <i>single strand conformation polymorphism</i>
SPEM	del inglés <i>spasmolytic polypeptide expressing metaplasia</i>
STAT	del inglés <i>signal transducers and activators of transcription</i>
tagSNPs	SNPs marcadores
TFF	<i>trefoil factor</i>
Th1	del inglés <i>T helper 1</i>
TLR	del inglés <i>toll-like receptor</i>
Tm	temperatura de fusión, del inglés <i>melting temperatura</i>
TNF	del inglés <i>tumor necrosis factor</i>
TR	Repeticiones en tándem
UTR	del inglés <i>untranslated region</i> , región transcrita pero no traducida,
VacA	del inglés <i>vacuolating cytotoxin A</i>
VEGF	del inglés <i>vascular endothelial growth factor</i>
VNTR	repeticiones en tándem de número variable
XRCC	del inglés <i>X-ray repair complementing</i>
x ²	test chi-cuadrado

Presentación

El cáncer gástrico es el resultado final de un largo proceso carcinogénico multifactorial que se desarrolla en distintas etapas y en el que intervienen un elevado número de factores ambientales y genéticos.

El grupo de investigación donde se ha realizado este trabajo se ha centrado tradicionalmente en el estudio de los factores genéticos de riesgo de diversas enfermedades complejas, con una especial atención al cáncer gástrico. Durante los años previos a este trabajo, ha participado en distintos estudios sobre factores ambientales, infección por *H.pylori* y susceptibilidad genética asociados a la carcinogénesis gástrica, identificando algunos factores genéticos de riesgo, como determinados SNPs en genes relacionados con la síntesis y reparación del DNA, con la vía de señalización de *H.pylori* o en el gen PSCA, entre otros.

Los genes *TFFs* y *MUCs*, codifican para las principales proteínas de formación y mantenimiento de la mucosa gástrica, mucinas y péptidos trefoil, que ejercen un papel importante en la protección del epitelio contra los patógenos y otras agresiones externas, así como intervienen en la regulación de diversos procesos celulares durante la homeostasis gástrica normal. Sin embargo, su implicación en la carcinogénesis gástrica todavía es poco conocida aunque cada vez hay más evidencias de su importancia.

Este trabajo profundiza en el conocimiento de los factores de susceptibilidad genética al cáncer gástrico a través del análisis de la variabilidad polimórfica de los genes *MUC* y *TFF* de la mucosa gástrica y su relación con el cáncer gástrico y la evolución de las lesiones precursoras.

La memoria de esta tesis se estructura siguiendo el esquema clásico de Introducción al tema, materiales y métodos, resultados, discusión y conclusiones. Para facilitar su lectura, en el apartado de materiales y métodos se resumen las metodologías utilizadas y en el apéndice I se detallan los protocolos de cada una de las técnicas. Referente al apartado de resultados, en el mismo solamente se incluyen las tablas y figuras que exponen e ilustran los más relevantes. El resto de tablas y figuras con resultados poco relevantes, adicionales o complementarios, se incluyen en los apéndices II y III. Finalmente, el apéndice 4 incluye las publicaciones obtenidas como consecuencia directa (una de ellas) o indirecta (otras dos) de este trabajo.

Introducción

1. CÁNCER GÁSTRICO

De la variedad de neoplasmas malignos que pueden afectar el estómago, el adenocarcinoma gástrico, tumor originado en el epitelio glandular de la mucosa gástrica, es el más frecuente (90-95%). Aunque existen otros tipos menos frecuentes como linfomas (3-5%) o tumores del estroma gastrointestinal (1-3%), en esta tesis hablaremos de cáncer gástrico para referirnos específicamente al adenocarcinoma gástrico.

El cáncer gástrico es el resultado final de un largo proceso carcinogénico multifactorial que se desarrolla en distintas etapas y en el que intervienen elevado número de factores ambientales y genéticos.

1.1. Incidencia y mortalidad

La incidencia y mortalidad de cáncer gástrico, especialmente el no cardias (FIGURA I-1A) y APARTADO 1.3), ha declinado considerablemente durante las últimas décadas en todo el mundo. Aun así, la distribución geográfica de la incidencia de este tumor es muy característica, existiendo importantes diferencias regionales de hasta 10 veces., desde una índice de incidencia estandarizado por edad (ASR, *age-standardized rate*) en hombres de 3.9 por cada 100.000 personas en el norte de África hasta 42.4 en el este asiático (1). Mientras que el descenso en la incidencia es notable en países como Norteamérica y Europa del Norte y Occidental, existen otras regiones geográficas como Asia Oriental, América del Sur, Centroamérica y Europa del Este, donde las tasas de incidencia y mortalidad permanecen elevadas. En conjunto el cáncer gástrico representa la segunda causa más frecuente de muerte por cáncer en el mundo, después del cáncer de pulmón, y la cuarta en Europa (2,3). En este continente la mayor incidencia de cáncer gástrico corresponde, sobretodo, a los países de Europa Oriental (1,3). En España, y a pesar de la disminución de la mortalidad que se ha producido en las últimas décadas en prácticamente todas las provincias, existe un marcado patrón geográfico que afecta a ambos sexos y que se ha mantenido consistente, con pequeños cambios, durante los últimos años. Las mayores tasas de mortalidad corresponden a determinadas zonas del centro y norte del país, estando en la cabeza Castilla-León con una tasa de 20.14 ARS por cada 100.000 personas en el periodo 2000-2005, y se encuentran entre las más altas de Europa Occidental (4,5).

Una de las características epidemiológicas a destacar del cáncer gástrico es que su incidencia es aproximadamente dos veces mayor en hombres que en mujeres en todas las regiones geográficas. Las causas por las que existen estas diferencias todavía son poco conocidas, aunque se ha propuesto que las hormonas femeninas, como el estrógeno, pueden tener un efecto protector (6-8).

Dado que la mayoría de pacientes se diagnostican en fases avanzadas, la supervivencia de este cáncer a los 5 años es baja, un 30% de los casos diagnosticados en España. Por ello, el conocimiento y control de los principales factores de riesgo para el desarrollo del cáncer gástrico son el método más efectivo para su prevención (9).

1.2. Anatomía e histología del estómago

Desde el punto de vista anatómico el estómago se divide en tres regiones principales: fundus, cuerpo y antro. Además, posee dos esfínteres: el cardias, que separa el estómago del esófago, y el píloro, que separa el estómago de la primera porción del intestino delgado o duodeno (FIGURA I-1A).

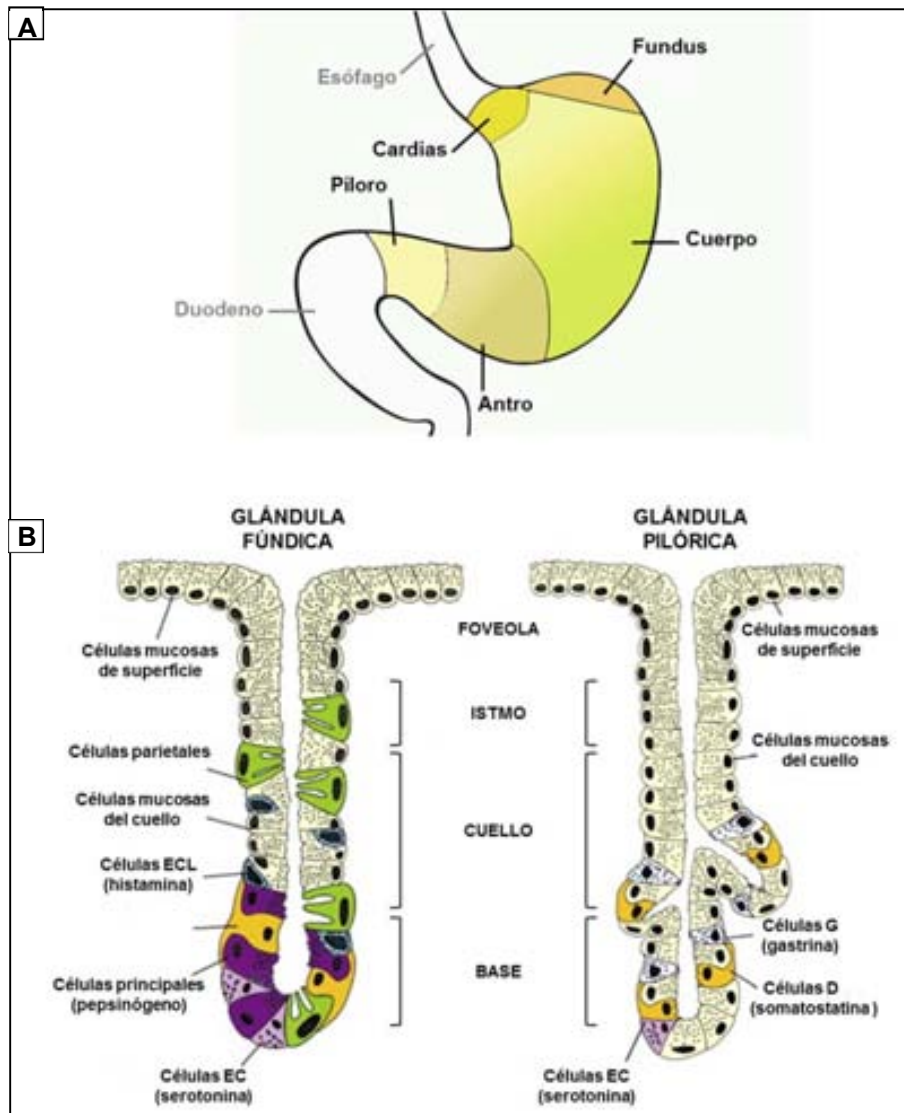


Figura I-1. A. Anatomía del estómago. B. Histología de las glándulas gástricas. Adaptado de Schubert et al. (10)

Histológicamente, la pared del estómago está formada por las siguientes capas superpuestas: mucosa (la capa más interna asociada a numerosas glándulas secretoras), submucosa, muscular, subserosa y serosa, que es el estrato más externo y alejado de la luz del estómago. A su vez, la mucosa gástrica está formada por tres capas: el epitelio superficial, la lámina propia que contiene los vasos sanguíneos que inervan el epitelio y una capa fina de músculo liso denominada muscularis mucosa (11). Una capa gruesa de mucus recubre la mucosa gástrica y actúa como barrera física separándola del lumen del estómago y protegiéndola de las

secreciones ácidas y otras agresiones externas, como la infección por patógenos. Esta barrera de mucus está compuesta por dos capas, una exterior, más laxa y que puede dispersarse fácilmente, y una interior, constituida por mucus firmemente adherido a la mucosa y que difícilmente se desprende por métodos mecánicos (12).

La mucosa gástrica se invagina formando foveolas gástricas donde desembocan una media de 4-5 glándulas secretoras. Se pueden diferenciar tres áreas histológicas en función de la estructura glandular presente:

- Una región compuesta por el fundus y el cuerpo, donde predominan las glándulas oxínticas o fúndicas responsables primordialmente de la secreción ácida del estómago, constituida por ácido clorhídrico y diversas enzimas, y secretoras también del mucus protector que recubre el epitelio. Estas glándulas presentan una estructura tubular recta con una luz muy estrecha, son muy profundas y están conectadas con foveolas largas. Las glándulas oxínticas (FIGURA I-1B) están compuestas por una población mixta de células:
 - ◆ Células mucosas de superficie, secretoras del mucus que cubre y protege el epitelio gástrico
 - ◆ Células parietales o oxínticas productoras de ácido clorhídrico y distribuidas a lo largo de toda la glándula pero más abundantes en la zona intermedia
 - ◆ Células mucosas del cuello, en la zona de transición hacia la base de la glándula, responsables también de la secreción de mucus pero de composición diferente al producido por las células mucosas de superficie.
 - ◆ Células zimogénicas o principales localizadas en la base de la glándula y responsables de la secreción de pepsinógeno.
 - ◆ Células endocrinas, en su mayoría células ECL (*enterochromaffin-like*) productoras y secretoras de histamina, un potente estimulador de la producción de ácido clorhídrico en las células parietales, y células EC (*enterochromaffin*) secretoras de serotonina.
- Una región antro-pilórica donde predominan las glándulas de estructura tubular ramificada y enrollada en su base, y conectadas por foveolas cortas, compuestas por una población de células menos compleja que las de las glándulas oxínticas (FIGURA I-1B). La principal función de estas glándulas es la secreción de mucus por células especializadas, similares a las células secretoras de mucus del cuerpo y fundus, y la secreción de las hormonas gastrina y somatostatina por células endocrinas G y D, respectivamente. La gastrina induce la secreción de ácido clorhídrico y pepsinógeno en las glándulas oxínticas.
- La región del cardias, una pequeña zona de glándulas fundamentalmente secretoras de mucus y similares a las de la región antro-pilórica.

1.3. Clasificación histológica y anatómica.

Las dos clasificaciones histológicas más utilizadas en el cáncer gástrico son la propuesta por la Organización Mundial de la Salud (OMS) (13) y la clasificación de Lauren (14).

Según la OMS, los cuatro patrones histológicos predominantes, basados principalmente en la morfología, son:

- Adenocarcinoma papilar: constituido por proyecciones epiteliales digitiformes dispuestas alrededor de ejes fibrovasculares y conectivos.
- Adenocarcinoma tubular: constituido principalmente por túbulos ramificados de diferente diámetro rodeados de estroma fibroso.
- Adenocarcinoma mucinoso: caracterizado por contener más del 50% del tumor depósitos mucosos extracelulares.
- Carcinoma de células en anillo de sello: con un componente predominante de células tumorales aisladas o en pequeños grupos que contienen importantes cantidades de mucina intracitoplasmática, desplazando el núcleo a la periferia.

La clasificación de Lauren ha sido ampliamente utilizada, sobretodo en estudios científicos, y describe dos tipos histológicos mayoritarios, intestinal y difuso, con diferentes características clínico-patológicas y epidemiológicas. Además, estos dos tipos difieren en su etiología, incluyendo la referida a factores genéticos. Esta clasificación es la que se ha utilizado en esta tesis ya que es muy útil para evaluar parámetros como su asociación a factores ambientales, tendencias de incidencia y lesiones precursoras del cáncer gástrico, entre otros. Cuando no es posible tipificar la neoplasia en ninguna de las dos categorías anteriores, se clasifican como de tipo indeterminada, que representan un 10% de todos los carcinomas de estómago.

1.3.1. Adenocarcinoma de tipo intestinal

El adenocarcinoma de tipo intestinal es un tumor bien diferenciado caracterizado por la formación de estructuras glandulares semejantes a las del intestino. Este tipo de adenocarcinoma es el resultado de unos cambios histológicos muy bien definidos que dan lugar a una serie de etapas diferenciadas (15) (FIGURA I-2).

El proceso se inicia con una inflamación de la mucosa gástrica, o gastritis, causada habitualmente por la infección de la bacteria Gram-negativa *Helicobacter pylori* (*H.pylori*) (ver APARTADO 1.5.1). La persistencia de la infección puede provocar una gastritis superficial crónica activa, que posteriormente puede progresar a gastritis atrófica, metaplasia intestinal (IM) y, finalmente, la lesión puede evolucionar a displasia, carcinoma y la subsiguiente diseminación metastásica. Esta secuencia de lesiones precursoras se describen más en detalle en el APARTADO 1.4

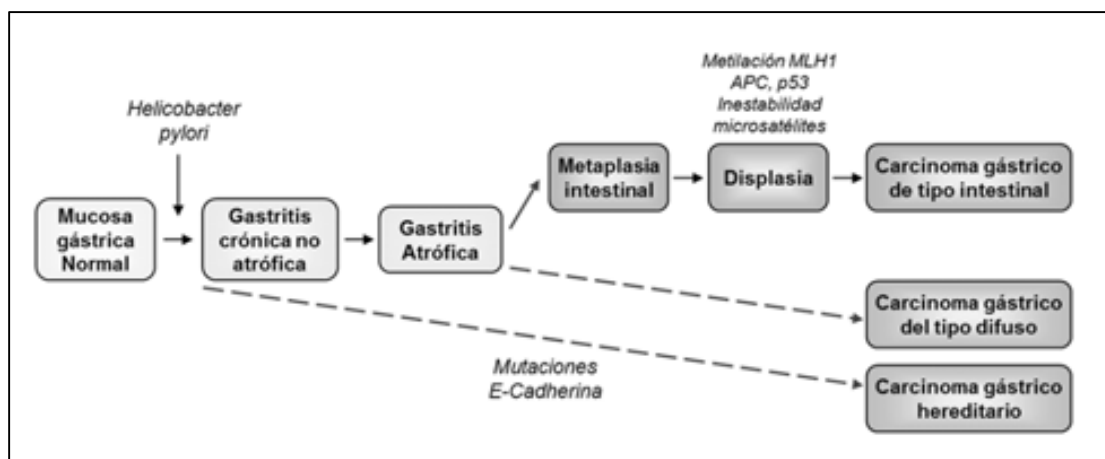


Figura I-2. Etapas de la carcinogénesis gástrica

Algunas de las alteraciones genéticas frecuentemente asociadas a la transformación maligna de las lesiones precursoras a un cáncer gástrico de tipo intestinal son la inestabilidad de microsatélites, mutaciones en p53, expresión reducida de p27, incremento en la expresión de la ciclina E y la presencia de transcritos aberrantes de c-met generados por procesamiento alternativo. Además, la inactivación de genes supresores de tumores como *DCC* y *APC* por pérdida de heterocigosidad o mutación, la expresión reducida del receptor tipo I de TGF- β (*TGFB1*) o la amplificación de *HER2* (o *ERBB2*) se han relacionado con un estadio más avanzado del carcinoma gástrico del tipo intestinal (16).

El tipo intestinal es el predominante en los países de alto riesgo, es más común en hombres que en mujeres y su incidencia incrementa significativamente con la edad (17).

1.3.2. Adenocarcinoma de tipo difuso

El adenocarcinoma de tipo difuso es histológicamente poco diferenciado. Consiste en células neoplásicas individualmente infiltradas que no forman estructuras glandulares y no está asociado con una secuencia de lesiones previas preneoplásicas definidas, como la metaplasia intestinal, exceptuando la gastritis crónica que es común a ambos tipos histológicos y característica de la infección por *H. pylori* (18).

Las células se infiltran de manera dispersa en la pared gástrica sin formar estructuras glandulares. Estas células son de apariencia redonda y pequeña, dispuestas individualmente o agrupadas en formaciones reticulares. Estos tumores se asemejan a aquellos clasificados como carcinomas de células en anillo de sello en la clasificación de la OMS. En comparación con el tipo intestinal, en el carcinoma difuso la desmoplasia, producción excesiva de tejido conectivo que acompaña el tumor, es más pronunciada, la inflamación asociada es menos evidente y la tasa mitótica también es menor (13). El fenotipo difuso de cáncer gástrico, tanto el hereditario como el esporádico, se ha relacionado con una expresión reducida de E-cadherina. La deficiencia de esta molécula de adhesión célula-célula es probablemente un defecto fundamental para el desarrollo de este tipo de carcinoma gástrico ya que proporciona una explicación para la

morfología observada en el mismo de células poco cohesivas con pérdida de la polaridad y de la arquitectura glandular. Aunque actualmente se desconocen lesiones precursoras de cáncer gástrico difuso esporádico, se ha propuesto un modelo de progresión similar al descrito en casos de cáncer gástrico familiares a partir de análisis histológicos, en el que la pérdida de E-cadherina provoca la dispersión de células de anillo de sello en el epitelio glandular hasta que invaden la lamina propia (18-20). Otra alteración genética implicada en el desarrollo del cáncer gástrico difuso es la pérdida de heterocigosidad en la región cromosómica 17p, donde se localiza p53. Además, la pérdida de p27, inhibidor dependiente de ciclina que controla la transición de G1 a S en la célula, y la amplificación génica de los proto-oncogenes c-met y K-sam contribuyen a la progresión del carcinoma y la dispersión metastática (16).

El cáncer gástrico difuso afecta más comúnmente a pacientes jóvenes, afecta por igual a hombres y mujeres y tiene una distribución geográfica más uniforme que el cáncer gástrico intestinal. Además, se considera que en la etiología de este tipo histológico la susceptibilidad genética pueda desempeñar un papel importante, aunque la infección por *H.pylori* sea el mayor factor de riesgo identificado hasta ahora (21).

1.3.2.1 Cáncer gástrico difuso hereditario

Aunque la mayoría de casos de cáncer gástrico son esporádicos, en aproximadamente el 10% de los casos existe agregación familiar (22,23). En 1998, Guilford *et al.* reportaron tres casos similares en familias Maories con cáncer gástrico difuso poco diferenciado, multigeneracional y de inicio temprano, en los que se identificaron mutaciones germinales en el gen de la E-cadherina (*CDH1*) mediante análisis de ligamiento y búsqueda de mutaciones (24). Este hallazgo condujo a la identificación de un nuevo síndrome de cáncer hereditario, el único hasta ahora caracterizado, denominado cáncer gástrico difuso hereditario y que representa entre el 1 y 3% de los casos de cáncer gástrico (25).

En aproximadamente el 30-40% de casos de cáncer gástrico hereditario se identifica una mutación germinal en uno de los alelos de *CDH1*, de las cuales se han reportado más de 100 hasta la actualidad. La inactivación del segundo alelo ocurre por mutación somática o hipermetilación y se estima que las mutaciones en *CDH1* tienen una penetrancia del 75-80% (25,26).

1.3.3. Subtipos anatómicos del adenocarcinoma gástrico

Desde el punto de vista anatómico, el cáncer gástrico puede clasificarse en cardias o no cardias según su localización. El cáncer gástrico del cardias incluye a los adenocarcinomas que se originan en la zona próxima del estómago, cerca o incluyendo la unión gastroesofágica. El cáncer gástrico no cardias comprende el resto de tumores, los que se originan en la parte distal del estómago como el cuerpo, el fundus o en la región antro-pilórica (27). Existen cada vez más evidencias de que estos dos subtipos anatómicos difieren tanto en su epidemiología como en la etiología y patogénesis. Como se ha comentado anteriormente, la incidencia de cáncer gástrico ha disminuido considerablemente, sin embargo, la incidencia del adenocarcinoma gástrico del

cardias ha incrementado rápidamente en muchos países durante las últimas décadas (17). Aunque es posible que este incremento se deba, en parte, a una mejora en la clasificación anatómica del cáncer que, actualmente, es mucho más precisa, no explica las grandes diferencias observadas con el cáncer gástrico no cardias. De hecho, se estima que un 50% de los adenocarcinomas gástricos que se diagnostican actualmente en países occidentales se originan en la pequeña localización anatómica del cardias cuando hasta ahora han representado un porcentaje mucho menor del total de cánceres de estómago diagnosticados, entre el 16-37% (17,28). El cáncer gástrico del cardias presenta un ratio entre hombres y mujeres mucho más pronunciado que el tipo no cardias y es mucho más común en los países occidentales. Histológicamente, el cáncer gástrico del cardias es del tipo intestinal en la mayoría de los casos y se origina probablemente a partir de metaplasia intestinal (27).

En relación a la etiología, existe una fuerte asociación entre el cáncer gástrico no cardias y la infección por *H.pylori*. La mayoría de estos tipos de cáncer son una consecuencia de la gastritis provocada por la infección de la bacteria que puede progresar a lesiones más severas. La relación entre *H.pylori* y el cáncer gástrico de la región del cardias es mucho más compleja y diversos estudios sugieren que *H.pylori* no está involucrado en la etiología de este tipo de cáncer (29,30). No obstante, la asociación de *H.pylori* con el cáncer del cardias varía entre distintas regiones geográficas, mientras que en países occidentales existe una fuerte tendencia negativa en la asociación, en algunos países orientales, como China, existe una tendencia positiva. Hay que tener en cuenta también que algunos adenocarcinomas gástricos del cardias comparten algunas características etiológicas y patogénicas con el cáncer esofágico, para el cual el reflujo gastroesofágico es un importante factor de riesgo (27,31,32)}.

1.4. Carcinogénesis gástrica y lesiones precursoras

1.4.1. Gastritis crónica activa

La infección por *H.pylori* es la principal causante de inflamación de la mucosa del epitelio gástrico, o gastritis crónica. Aunque la inflamación se desencadena acompañada de una respuesta inmunitaria contra la infección, *H.pylori* ha desarrollado diversos mecanismos que le permiten evadir el sistema inmune, que no es capaz de eliminar la bacteria de manera efectiva. En consecuencia, la infección puede perdurar toda la vida en ausencia de una terapia antimicrobiana adecuada (33).

Una de las características principales de la gastritis crónica es la infiltración de células mononucleares representativas de la inflamación crónica, como linfocitos, células plasmáticas y macrófagos, en la porción superficial de la lámina propia de la mucosa gástrica (34). Adicionalmente pueden observarse eosinófilos dispersos y mastocitos. La gastritis se denomina activa cuando se encuentran neutrófilos polimorfonucleares característicos de una inflamación aguda. La infiltración de células mononucleares se asocia a la expresión de citocinas proinflamatorias del tipo T helper-1 (IL-1 β , TNF- α e IFN- γ). La respuesta inflamatoria puede

persistir durante todo el proceso precancerígeno, pero su intensidad tiende a disminuir a medida que el proceso avanza (35).

La gastritis crónica activa no muestra pérdida de glándulas, por lo que se define como no atrófica, y se localiza mayoritariamente en el antro. Habitualmente, la gastritis no atrófica antral se asocia a úlcera duodenal, la cual no incrementa el riesgo de cáncer gástrico (34,35).

La evolución de la gastritis crónica dependerá de la interacción entre diferentes factores etiológicos: los relacionados con el causante de la infección (factores de virulencia en las cepas de *H.pylori*), la susceptibilidad del huésped y los factores ambientales (ver APARTADO 1.5). Si no se trata eliminando la infección por *H.pylori*, la gastritis crónica puede evolucionar de dos maneras: permanecer estable como no atrófica o progresar en severidad hacia una gastritis atrófica.

1.4.2. Gastritis atrófica

La gastritis atrófica se caracteriza por la pérdida del tejido glandular especializado del estómago, como las glándulas antrales o las glándulas oxínticas del cuerpo gástrico, que contienen las células parietales responsables de la secreción ácida. La pérdida de estas glándulas está acompañada por la fibrosis de la lámina propia y la aparición de nuevas glándulas metaplásicas, similares a las del intestino y/o del tipo secretor de mucus (35).

La gastritis iniciada en el antro tiende a extenderse de manera pilorocardial, es decir, desde el piloro hacia el cardias, hasta que la atrofia ocupa todo el estómago y se pierde la práctica totalidad de las glándulas propias de la mucosa gástrica. Este tipo de atrofia se denomina multifocal y se considera una lesión precursora de cáncer gástrico. Al desaparecer las células parietales y las células secretoras de pepsinógeno, disminuye la secreción ácida del estómago,. Por esta razón la atrofia de la mucosa en el corpus comporta con frecuencia la hipoclorhidria y, en ocasiones, aclorhidria (libre de ácido) del estómago. Por lo contrario, si la gastritis se mantiene en el antro, generalmente se asocia a hipergastrinemia, exceso de secreción de gastrina, e hiperclorhidria, factores de riesgo para úlcera duodenal (36,37).

La atrofia y posterior aclorhidria del estómago puede provocar la colonización masiva por parte de bacterias y hongos representativos de flora normal del tracto gastrointestinal. El crecimiento de esta flora contribuye a potenciar la formación endógena de acetaldehído y de compuestos N-nitroso potencialmente carcinogénicos (38,39). El acetaldehído, abundante en el tracto gastrointestinal superior, especialmente en el estómago, ha sido recientemente clasificado como carcinógeno del grupo I por la IARC en 2009. Este compuesto está asociado con las bebidas alcohólicas y se forma endógenamente a partir del etanol. El acetaldehído también es uno de los carcinógenos más abundantes en el humo del tabaco, otro de los factores de riesgo conocidos de cáncer gástrico (40).

1.4.3. Metaplasia intestinal

La metaplasia intestinal se considera una lesión preneoplásica del estómago, en la que la mucosa normal gástrica se reemplaza por una mucosa histológicamente similar a la del intestino

humano y se caracteriza, especialmente, por la presencia de células caliciformes secretoras de mucus. Esta lesión puede progresar a displasia y adenocarcinoma gástrico.

La metaplasia intestinal es una lesión heterógena para la cual se han propuesto diversas clasificaciones. Actualmente, la clasificación más ampliamente adoptada por patólogos y clínicos es la propuesta inicialmente por Jass y Filipe (41) que discrimina entre metaplasia intestinal completa e incompleta en base a la morfología observada en secciones teñidas con hematoxilina-eosina y el tipo de mucinas detectadas en tinciones histoquímicas y que distinguía tres tipos: completa o tipo I, incompleta tipo II e incompleta tipo III. La metaplasia intestinal completa, o tipo I, se caracteriza por la presencia de enterocitos con microvillosidades, que son las células absortivas del intestino, células Paneth y células caliciformes secretoras de sialomucinas. En este tipo de IM también se caracteriza por una disminución de la expresión de las mucinas gástricas MUC1, MUC5AC y MUC6 y la expresión *de novo* de la mucina intestinal MUC2. Estas características corresponden al fenotipo del intestino delgado. En la metaplasia intestinal incompleta se adquiere una morfología similar a la del colon. Se caracteriza por la presencia de células columnares y caliciformes que secretan sialo y/o sulfomucinas y la coexpresión de las mucinas gástricas MUC1, MUC5AC y MUC6 con MUC2. La diferencia entre la metaplasia intestinal incompleta tipo II y III está en la clase de mucinas que se produce en cada caso, neutrales y sialomucinas en el tipo II y sulfomucinas en el tipo III. Otra de las características del tipo III es la ausencia de células Paneth y una importante desestructuración glandular. Debido a la desorganización celular que presenta el tipo III se considera esta lesión prácticamente displásica y la fase más avanzada de metaplasia intestinal (42,43).

Algunos estudios han indicado que la metaplasia intestinal incompleta es la que presenta un riesgo más elevado de cáncer gástrico (44,45) aunque la asociación entre este subtipo de metaplasia y mayor riesgo de cáncer gástrico ha sido refutada por otros autores (46). A pesar de la controversia existente al respecto, una reciente revisión de todos los estudios realizados hasta ahora ha demostrado la utilidad del subtipaje de esta lesión como marcador del riesgo de cáncer gástrico (47). Se considera que el patrón de distribución de la metaplasia, antro o corpus, también proporciona un valor predictivo de cáncer gástrico, siendo el riesgo más elevado a medida que la metaplasia se extiende por el corpus. La asociación entre metaplasia intestinal y riesgo de cáncer gástrico se ha reportado principalmente para el tipo intestinal, no existiendo tan claras evidencias en el caso del tipo difuso (48:Sakitani, 2011 #1241,49)

Durante las diferentes etapas de la metaplasia intestinal se suceden numerosas alteraciones moleculares y genéticas, aunque todavía son poco conocidos los mecanismos por los cuales éstas se relacionan con la progresión de la lesión. Se han descrito inestabilidad de microsátelites y pérdida de heterocigosidad, aunque estos eventos son más frecuentes una vez se ha desarrollado el cáncer. Otras alteraciones genéticas observadas en esta lesión precursora son mutaciones en p53 y un aumento en la activación de la telomerasa, que incrementa la longitud del DNA telomérico permitiendo que la célula continúe replicándose (48). Además se desregulan diversos genes con un importante papel en la transformación del epitelio gástrico a un epitelio de fenotipo intestinal. En metaplasia intestinal aumenta la expresión de *CDX1* y *CDX2*, dos

factores de transcripción que regulan la proliferación y diferenciación de las células epiteliales intestinales y disminuye la de *SHH*, factor implicado en la organogénesis de los vertebrados y que se expresa normalmente en el fundus y corpus del estómago (50,51). También se ha reportado la alteración de la expresión de genes como *RUNX3*, que se ha descrito como supresor de tumores para cáncer gástrico, *PDX1*, *OCT-1* o diversos miembros de la familia de las mucinas (*MUCs*) y de los *trefoil factors* (*TFFs*), entre otros (48,52).

La infección por *H.pylori* también se ha asociado más recientemente a una forma alternativa de metaplasia denominada SPEM (*spasmolytic polypeptide expressing metaplasia*) que crece en relevancia (49). El linaje SPEM se caracteriza especialmente por un incremento en la expresión de *TFF2* (*trefoil factor family 2*), que se expresa normalmente en la mucosa intestinal (53). En este tipo de metaplasia, *TFF2* es expresado por glándulas localizadas en la mucosa oxíntica del corpus pero que presentan la morfología de las glándulas del antro (atrofia oxíntica), proceso conocido tradicionalmente como “antralización” del cuerpo (54). Es de destacar que la metaplasia tipo SPEM es la lesión preneoplásica predominante en diversos modelos animales (55).

1.4.4. Displasia

La displasia es un precursor directo del adenocarcinoma gástrico y también un importante indicador del riesgo de cáncer. Esta lesión se caracteriza por un fenotipo neoplásico, en términos de morfología y organización tisular. Una de sus características es la atipia celular, con células de formas irregulares, hipertróficas e hiper cromáticas, ya que sus núcleos presentan una tinción más oscura debido a una mayor concentración de proteínas asociadas al DNA, y que carecen de polaridad. El tejido displásico también presenta glándulas desestructuradas, de diverso tamaño y formas irregulares y, ocasionalmente, bifurcadas o con ramificaciones. Esta transformación y desdiferenciación del epitelio es progresiva y se convierte en carcinoma invasivo cuando la atipia celular alcanza la membrana basal. Actualmente, la displasia se denomina también neoplasia intraepitelial, o no invasiva, y se clasifica como de bajo o alto grado reflejando, en cada caso, el riesgo de cáncer (13,56). En la mucosa displásica también se suceden numerosas alteraciones genéticas, muchas de ellas presentes también en otras lesiones precursoras previas y que se mantienen en el carcinoma gástrico aunque, en este último caso, en mayor frecuencia. Algunas de las alteraciones moleculares detectadas son mutaciones en supresores de tumores, especialmente *p53* y *APC*, y oncogenes, como *c-erbB2*, *c-met*, *K-ras* y *K-sam*. También se ha descrito inestabilidad de microsatélites e hipermetilación en genes involucrados en el ciclo celular, reparación del DNA e invasión y metástasis, entre otros (57,58).

1.5. Etiología

Durante muchas décadas, la etiología del cáncer gástrico permaneció muy poco conocida. Inicialmente, los esfuerzos se concentraron en evaluar la hipótesis de que compuestos químicos ambientales potencialmente cancerígenos, especialmente nitrosaminas, fueran factores causales de cáncer gástrico. El descubrimiento posterior de *H.pylori*, a principios de los 80 (59),

fue un momento crucial para el conocimiento de la patogénesis del cáncer gástrico, centrando la atención de los investigadores en la etiología bacteriana. Más recientemente, se ha puesto de relieve la influencia de la susceptibilidad genética individual en el proceso carcinogénico y su interacción con la infección por *H.pylori* y otros factores ambientales, un campo que ha ocupado el primer plano en la investigación de los últimos años.

1.5.1. Helicobacter pylori

H.pylori es una bacteria Gram negativa, de forma espiral y microaerófila (requiere bajas concentraciones de oxígeno para su desarrollo) que coloniza el estómago humano. La infección causada por este microorganismo, clasificado desde 1994 como un carcinógeno del grupo 1 para humanos (60), es uno de los factores más importantes para entender la patogénesis del cáncer gástrico. La colonización del estómago por *H.pylori* causa gastritis crónica, la cual, durante las décadas siguientes a la infección inicial, puede permanecer silente o evolucionar a enfermedades más severas, como la gastritis atrófica, la úlcera péptica, o el linfoma gástrico. Numerosos estudios epidemiológicos también confirman que la infección por *H.pylori* es el principal factor de riesgo de desarrollar cáncer gástrico, especialmente del tipo no cardias, el cual aumenta entre 2 y 30 veces en individuos infectados en comparación con los no infectados. La variabilidad de los OR observados depende de la técnica utilizada para la detección de la bacteria: ELISA o Western Blot, mucho más sensible (30,61). De hecho, se considera que *H.pylori* es responsable de aproximadamente el 75% de los casos de cáncer gástrico del tipo no cardias y del 63.4% de todos los casos de cáncer de estómago a nivel mundial (29).

La infección causada por *H.pylori* es de las más comunes en el ser humano. Entre un 10 y 50 % de la población adulta de países desarrollados está infectada por la bacteria, mientras que esa proporción puede llegar hasta el 80 y 90 % en países en vía de desarrollo. Sin embargo, menos del 1% de los de los infectados desarrollarán un cáncer gástrico en su vida, indicando que existen otros factores involucrados en la carcinogénesis (62). Se considera que la combinación de factores de virulencia de la propia bacteria, factores de susceptibilidad genética del huésped, que condicionan la colonización de la mucosa por parte de la bacteria y la consecuente respuesta inmunoinflamatoria, y factores ambientales, relacionados principalmente con la dieta y el consumo de tabaco, influyen la progresión y el destino final de la infección.

1.5.1.1 Colonización de la mucosa gástrica

En el polo apical de las células epiteliales que forman la mucosa gástrica aparece una gruesa capa de mucus viscoelástico consistente en agua, mayoritariamente, mucinas, y otras moléculas como péptidos TFF que sirve de protección del desgaste y de la invasión por patógenos. *H.pylori* contiene diversos factores que le permiten colonizar su hábitat natural, la mucosa gástrica, en un ambiente de condiciones variables y expuesto a agentes lesivos como el ácido gástrico (63).

En el estómago existe un gradiente de pH, desde un pH casi neutro (entre 4.5 y 6.5) en las superficie epitelial, hasta un pH ácido en el lumen, que puede variar entre 5.0, en la fase

digestiva, hasta valores cercanos a 1. *H. pylori* ha desarrollado mecanismos para protegerse tanto de la acidez gástrica durante la colonización como de los fuertes cambios bruscos de pH que suceden en su entorno, comportamiento descrito como una respuesta adaptativa al ácido, mediada esencialmente por la enzima ureasa y su complejo sistema de regulación para prevenir la alcalinización excesiva del medio, que sería letal para la bacteria. La ureasa es una enzima que se acumula principalmente en el citoplasma de la bacteria y cataliza la hidrólisis de la urea presente en el estómago en amonio y dióxido de carbono. La proteína transportadora ureI, dependiente del pH, es la responsable de activar la entrada de urea en el citoplasma de la bacteria. Cuando se alcanza un pH neutro ureI se inactiva, disminuyendo la tasa de urea que entra en el citoplasma y, por tanto, la actividad de la ureasa se reduce. Para conservar energía y recursos, *H. pylori* intenta conseguir un nicho donde no tenga que hacer uso constantemente de su capacidad de resistencia y adaptación al entorno ácido. Por esta razón, su colonización es máxima inicialmente en la parte antral del estómago, una región con un pH más elevado que la mucosa del corpus (64,65).

H. pylori dispone de diversas adhesinas que facilitan la unión de la bacteria con las células epiteliales gástricas a través de múltiples receptores. Esta capacidad de adherencia permite a la bacteria colonizar el epitelio gástrico y liberar eficientemente los factores de virulencia, como CagA, en las células huésped (66). De entre las más de 30 proteínas de la membrana exterior de *H. pylori* conocidas, BabA es una de las mejor descritas. Esta adhesina se une al antígeno de grupo sanguíneo Lewis B en las células gástricas, promoviendo posiblemente la patogenicidad de la bacteria (67). El gen babA está presente en el 40-95% de las cepas de *H. pylori*, dependiendo de la región geográfica. Las cepas que contienen babA están asociadas con una mayor incidencia de cáncer gástrico y, además, la presencia de este gen correlaciona con la presencia de la cag PAI y el alelo s1 de vacA, asociados a un mayor riesgo de cáncer gástrico (21). La mucosa gástrica de pacientes infectados con cepas portadoras de BabA también presenta una mayor densidad de colonización de la bacteria y mayores niveles de la citocina proinflamatoria IL-8 (68). Otra de las adhesinas bien conocidas de *H. pylori* es SabA, que se une al antígeno Lewis x. Lewis x se sobreexpresa como consecuencia a la infección por *H. pylori* y la respuesta inflamatoria desencadenada (69).

1.5.1.2 Factores de virulencia

Las poblaciones de *H. pylori* son extremadamente diversas a nivel genético. Además, un mismo huésped puede presentar diversas cepas de la bacteria y éstas pueden cambiar a lo largo del tiempo debido a mutaciones, reordenamientos cromosómicos o recombinación entre cepas. Esta plasticidad del genoma de *H. pylori* le confiere una importante capacidad adaptativa al huésped (70-72). Aunque esta diversidad ha dificultado la búsqueda de factores de la bacteria asociados a cáncer, dos de ellos, la isla de patogenicidad *cag* y la toxina VacA, han sido ampliamente estudiados y considerados como factores clave en la patogenicidad de *H. pylori*. Estos factores serán determinantes, especialmente, en el desarrollo y evolución de la gastritis crónica.

El factor de virulencia mejor caracterizado de *H. pylori* es la isla de patogenicidad *cag* (*cag* PAI), un locus de aproximadamente 40 kb que contiene 31 genes. Un gen de esta isla, *cagA* (*cytotoxin-associated gene A*), que codifica la proteína CagA de 120-145 kDa, se utiliza habitualmente como marcador del locus (73). Más del 95% de las cepas aisladas en países de Asia Oriental son portadores de la *cag* PAI y, por tanto, *cagA* positivas, mientras que en países occidentales el 40% de las cepas, aproximadamente, son *cagA* negativas. Las cepas CagA+ aumentan el riesgo de desarrollar gastritis severa, gastritis atrófica, úlcera peptídica y cáncer gástrico comparadas con las cepas CagA- (74,75). La *cag* PAI también contiene genes que codifican para el llamado sistema de secreción tipo IV, a través del cual CagA y otros productos bacterianos, como fragmentos de peptidoglicano, son inyectados en las células huésped (76). Una vez la proteína CagA es translocada al citoplasma de la célula epitelial del huésped, diferentes miembros de la familia de quinasas Src la fosforilan en los residuos tirosina de la región de repeticiones de 5 aminoácidos EPIYA (77). CagA fosforilada interacciona con SHP2, activando la cascada de las ERK/MAPK e induciendo cambios morfológicos, como el reordenamiento del citoesqueleto y la elongación y dispersión celular (78-80). CagA también puede alterar diversas funciones celulares independientemente a su estatus de fosforilación. CagA interacciona con componentes de los complejos de unión celular, como la proteína de adhesión E-cadherina, causando la ruptura de las uniones célula-célula, esenciales para la integridad del epitelio gástrico. De hecho, las células que expresan CagA pierden la polaridad y la adhesión celular, y son capaces de invadir a través de la matriz extracelular. Además, la desestabilización de los complejos E-cadherina/ β -catenina provocan la acumulación nuclear y citoplásmica de la β -catenina que, a su vez, induce genes específicamente intestinales en las células epiteliales gástricas como *CDX1* y *MUC2* (81-83). CagA no fosforilada también potencia la activación del factor de transcripción NF- κ B, que a su vez induce la producción de la citocina proinflamatoria IL-8 (84). Más recientemente se ha confirmado CagA como una oncoproteína bacteriana, debido a su papel como inhibidor de la apoptosis (85,86) e inductor de la proliferación de células epiteliales gástricas y desarrollo de carcinoma en ratones transgénicos que expresan CagA (87).

Otro de los factores que participa en la colonización crónica de la mucosa gástrica y que desempeña un papel importante en la patogénesis del cáncer gástrico es el factor de virulencia **VacA** (*vacuolating cytotoxin A*). VacA es una proteína de secreción, codificada por el gen *vacA*, inicialmente identificada por inducir la formación de vacuolas citoplasmáticas en diversos tipos celulares, incluyendo células epiteliales gástricas (88,89).

La proteína VacA es secretada a través del sistema de secreción tipo V en el espacio extracelular, donde forma agregados dodecaméricos. El oligómero, al exponerse a condiciones de pH ácidas o alcalinas, se desensambla en monómeros que se insieren en la bicapa lipídica de las células huésped formando canales selectivos de aniones, que posteriormente son endocitados por la célula. La actividad de estos canales aumenta la concentración aniónica intraluminal, especialmente de cloruro, que trata de compensarse con el incremento de bombeo de protones al interior del compartimento endocítico, provocando su acidificación. A través de la membrana

se difunden bases débiles, como el amonio producido por la actividad ureasa de *H.pylori*, que son protonadas en el interior creando un desequilibrio osmótico que causará la captación de agua por el endosoma, originando la vacuolización de la célula (90,91).

Además de su función vacuolizante, VacA se considera una toxina multifuncional, capaz de producir numerosas alteraciones en las células epiteliales, y contribuye significativamente en la patogenicidad y persistencia de la infección por *H.pylori*. VacA altera la dinámica mitocondrial, induciendo la liberación del citocromo c y, consecuentemente, la apoptosis de las células epiteliales gástricas (92,93). VacA puede actuar también como modulador de la respuesta inmune, inhibiendo la activación y proliferación de los linfocitos T e interfiriendo entre la señalización del receptor de linfocitos T y IL-2. Al ser los linfocitos T uno de los mayores efectores de la respuesta inmune adaptativa, se considera que la acción inmunomoduladora de VacA puede contribuir a la capacidad de *H.pylori* de infectar de manera crónica la mucosa gástrica (94). Además, VacA también incrementa la permeabilización de la membrana epitelial que, en células parietales gástricas, provoca la ruptura de las interacciones membrana-citoesqueleto y, como consecuencia, potencia la hipoclorhidria inducida por *H.pylori* (95).

El gen que codifica VacA, *vacA*, presenta una considerable diversidad genética entre las distintas cepas de *H.pylori*. Las regiones con mayor variabilidad de secuencia son la localizada en el extremo 5' de *vacA* (alelos s1a, s1b, s1c o s2), la región media (alelos m1 o m2) y la región intermedia (alelos i1 y i2) (96,97). Aunque todas las cepas contienen el gen *vacA*, los diferentes alelos presentan variaciones en la actividad funcional de VacA *in vitro*, desde el tipo s1/i1/m1, que es la forma totalmente activa, hasta el tipo s2/i2/m2 que carece de actividad vacuolizante. El tipo s1/m1, con mayor actividad citotóxica, se ha asociado consistentemente con un mayor riesgo de cáncer gástrico (98).

1.5.1.3 Respuesta inmunitaria e inflamatoria a la infección

La respuesta inmune innata está mediada, en gran parte, por receptores que se expresan en las células epiteliales gástricas y células presentadoras de antígenos como los receptores *toll-like* (TLRs), en el epitelio gástrico TLR2 y TLR4 mayoritariamente, y los receptores *NOD-like* (NLRs), especialmente NOD1. Estos receptores reconocen componentes de la pared bacteriana o péptidos derivados del patógeno iniciando la señalización que provocará la activación de las defensas del huésped. Mientras que los TLRs están asociados a la membrana plasmática actuando como receptores celulares de superficie de componentes de la pared bacteriana (lipoproteínas y lipopolisacáridos), NOD1 se expresa en el citosol de la célula, donde reconoce péptidos derivados de la degradación de los peptidoglicanos inyectados por *H.pylori*. La activación de estos receptores incrementa la expresión de la citocina proinflamatoria IL-8 vía NF- κ B, mediando la respuesta inflamatoria causada por patógenos (99,100).

La respuesta innata inducida por la infección de la bacteria se desarrolla acompañada de una respuesta inmune adaptativa predominantemente del tipo Th1 (*T helper 1*), la cual está asociada a un incremento en la producción de diversas citocinas proinflamatorias como TNF α , IFN- γ , IL-1 β e IL-8. Un nuevo subgrupo de efectores definidos como Th17, identificados por la

secreción de IL-17 y que se distinguen de las células Th1 y Th2 en su diferenciación y función, también se han asociado a la respuesta celular inducida por la infección por *H.pylori* (101). Las células Th1 y Th17 podrían actuar sinérgicamente en la inducción de la gastritis durante la infección, promoviendo el reclutamiento en la mucosa gástrica de células inflamatorias mediante la inducción de quimoquinas. La polarización de la respuesta inmunitaria, que es predominantemente Th1/Th17, se debe en gran parte a la inducción de la secreción de IL-12 y IL-6, por neutrófilos, macrófagos y células dendríticas, estimulada por la infección por *H.pylori* (91).

Las citocinas, además de su papel en inflamación, han sido implicadas como importantes mediadores de los cambios fisiológicos que acompañan la infección por *H.pylori*. Las citocinas IFN- γ , IL-1 β , IL8 y TNF- α se han asociado a un incremento en la producción de la hormona gastrina, potenciadora de la secreción de ácido, por parte de las células G en la región antro-pilórica. Además, TNF- α e IFN- γ inhiben la secreción de somatostatina por las células D, que regula negativamente la gastrina. La gastrina y IL-1 β actúan aumentando la expresión de ligandos para el receptor EGF, modificando la proliferación y supervivencia de las células epiteliales gástricas. De hecho, la producción de gastrina se ha asociado con una respuesta hiperproliferativa de las células epiteliales gástricas durante la infección (102). Por otro lado, las citocinas proinflamatorias IL-1 β y TNF- α son potentes inhibidores de la secreción de ácido por parte de las células parietales del fundus y cuerpo. IL-1 β también disminuye la liberación de histamina, estimulador de la producción de ácido clorhídrico, por parte de las células ECL (103). También se ha reportado que la infección de *H.pylori* aumenta los niveles de IL-21 en las células de la mucosa gástrica. El incremento de la expresión de IL-21 está asociada con la producción de las metaloproteinasas de matriz MMP-2 y MMP-9, mediante un mecanismo dependiente de NF- κ B. La estimulación de la producción de MMPs puede favorecer la diseminación del patógeno y la infección persistente, promoviendo la destrucción del tejido gástrico (104).

H.pylori también aumenta la expresión de COX-2, enzima responsable de la producción de prostaglandina que puede incrementar la proliferación celular e inhibir la apoptosis en la mucosa gástrica infectada (36). La activación de la vía de señalización mediada por COX2 es también responsable de la inducción de la expresión de VEGF en las células gástricas infectadas por *H.pylori*. Este efecto se relaciona específicamente con la toxina VacA, y se asocia a un incremento significativo de la formación de vasos sanguíneos, sugiriendo que la neoangiogénesis puede contribuir al crecimiento tumoral en la carcinogénesis gástrica ligada a *H.pylori* (105,106).

1.5.2. Otros factores ambientales

Además de *H.pylori*, otras exposiciones ambientales y hábitos de vida pueden influir en la carcinogénesis gástrica. Diversos estudios prospectivos han demostrado una significativa relación dosis dependiente entre el consumo de tabaco y el riesgo de cáncer gástrico. Este riesgo es mayor para los fumadores y ex-fumadores, incrementa con la intensidad y duración del tabaquismo, y está asociado principalmente al riesgo de cáncer gástrico del cardias (107,108). Un meta-análisis y un estudio en la cohorte EPIC (*European Prospective Investigation into Cancer*

and Nutrition) recientes sugieren que el consumo excesivo de alcohol también aumenta el riesgo de cáncer gástrico no cardias (109,110). La obesidad también se ha reportado como uno de los factores de riesgo para cáncer gástrico, especialmente del cardias, aunque los diferentes estudios epidemiológicos presentan resultados conflictivos (111,112). Los datos epidemiológicos también revelan que los antecedentes familiares de cáncer gástrico también son un importante factor de riesgo, con un incremento del riesgo entre 1.5 y 3 veces entre los parientes de primer grado de pacientes con cáncer gástrico. En cuanto a la dieta, existen evidencias de que el consumo de frutas y vegetales frescos disminuye significativamente el riesgo de cáncer gástrico, especialmente del tipo intestinal (113-116). Diversos estudios caso-control también han reportado la asociación entre un elevado consumo de sal o ingesta de alimentos conservados en sal con un mayor riesgo de cáncer gástrico, aunque esta evidencia en estudios prospectivos es inconsistente. El consumo de alimentos salados o conservados en sal conlleva cambios en la viscosidad de la mucosa gástrica, aumentando la exposición a compuestos carcinogénicos. Una elevada ingesta de sal también puede dañar el epitelio gástrico, potenciando la respuesta inflamatoria y promoviendo la proliferación celular que aumentaría la probabilidad de mutaciones (117). El consumo de carnes curadas también se ha asociado con un mayor riesgo de cáncer gástrico, especialmente del no cardias, aunque en este caso las evidencias tampoco son concluyentes (118). Se considera que el descenso a nivel mundial en la incidencia de cáncer gástrico puede ser atribuido, en parte, a la mejora de las condiciones higiénicas y al uso de la refrigeración para preservar los alimentos, disminuyendo, así, el consumo de alimentos conservados en sal e incrementando la ingesta de frutas y vegetales frescos (119).

1.5.3. Susceptibilidad genética

Existe una considerable variación en el riesgo individual de desarrollar cáncer gástrico en pacientes infectados por *H.pylori*. Este hecho se hace evidente cuando se observa la elevada proporción global de infección, entre el 10 y 90% dependiendo de la región geográfica, y la incidencia global de cáncer gástrico, que es mucho menor. Además de los factores de virulencia de *H.pylori* y los factores ambientales que pueden modular la infección, se considera que los factores genéticos también pueden tener un papel importante en la carcinogénesis gástrica. Un amplio estudio con gemelos realizado en el norte de Europa estimó la heredabilidad del cáncer gástrico, determinada por factores genéticos, en un 28% (120). La susceptibilidad genética individual se ha asociado a polimorfismos genéticos específicos, mayoritariamente polimorfismos de un simple cambio de nucleótido o SNPs (*Single Nucleotide Polymorphism*), que podrían modificar el efecto de las exposiciones ambientales, explicando, en parte, la gran variabilidad en la incidencia de cáncer gástrico a nivel mundial. El estudio de los polimorfismos genéticos en la carcinogénesis gástrica se ha ido incrementando durante los últimos años, probablemente debido a los avances en las tecnologías de análisis del DNA y en el conocimiento del genoma humano. Se han publicado numerosos estudios de asociación entre polimorfismos y cáncer gástrico, caso-control o de cohorte, basados inicialmente en la selección individual de SNPs en genes candidatos y, a medida que aumentaba el conocimiento de la variabilidad genética en el

genoma humano y los avances tecnológicos, incorporando la selección de polimorfismos marcadores (tagSNPs) hasta los recientes estudios de rastreo completo del genoma (GWA, del inglés *genome-wide association study*).

Los tagSNPs son representativos de la variabilidad genética de una determinada región del genoma ya que se encuentran en elevado desequilibrio de ligamiento (LD, *linkage disequilibrium*) con otros SNPs. De esta manera, la selección de tagSNPs en estudios de asociación es muy útil porque permite reducir el número de SNPs analizados manteniendo la potencia para detectar la variante de riesgo asociada. Gran parte de la mejora en el conocimiento de la variabilidad genética y su aplicación en los estudios de asociación se debe al proyecto internacional HapMap (121), que tiene como objetivo mapear los patrones de diversidad genética en el genoma y, de esta manera, proporcionar a la comunidad científica una descripción del desequilibrio de ligamiento permitiendo la optimización del proceso de selección de SNPs. En una primera fase, que finalizó en 2005 (122), se genotiparon aproximadamente 1 millón de SNPs en 269 muestras de DNA correspondientes a trios de individuos procedentes de 4 poblaciones diferentes de ascendencia europea, asiática y africana. Posteriormente, en una segunda fase que finalizó en 2007, se aumentó considerablemente el número de SNPs genotipados a 3.1 millones, resultando en una densidad de, aproximadamente, 1 SNP por kilobase (123). Actualmente, el proyecto está en la tercera fase, donde se han genotipado 1.6 millones de SNPs comunes en más de 1184 individuos procedentes de 11 poblaciones diferentes y se han secuenciado diez regiones de 100 kb en 692 de estos individuos para identificar e incluir también variantes raras (124).

1.5.3.1 Genes candidatos

Los factores de susceptibilidad genética incluyen polimorfismos en genes involucrados en procesos críticos de la carcinogénesis gástrica como: la respuesta inflamatoria e inmunitaria a la infección, incluidos los genes responsables del reconocimiento de *H.pylori* en las células epiteliales gástricas y de la activación de vías de señalización que activan respuestas proinflamatorias y apoptóticas, la protección de la mucosa gástrica, la metabolización y detoxificación de compuestos cancerígenos, la reparación del daño oxidativo, la proliferación y adhesión celular y la reparación del DNA. Los genes de estos mecanismos y, muy especialmente, los relacionados con la respuesta inmunoinflamatoria, han sido pues los más analizados en las estrategias de análisis de genes candidatos (125,126).

Las primeras evidencias epidemiológicas indicando un mayor riesgo de cáncer gástrico y polimorfismos en genes proinflamatorios se reportaron para el cluster de genes *IL1* en la región cromosómica 2q14, que contiene los genes *IL1A*, *IL1B* y *IL1RN*, codificando para IL-1 α , IL-1 β y IL-1Ra (receptor antagonista de la IL1), respectivamente. Los genotipos proinflamatorios *IL1B* c.-31TT, en la secuencia de la caja TATA del promotor del gen, e *IL1RN* 2/2 (o alelo IL1RN*2 (2 repeticiones)) portador de los alelos con 2 repeticiones de un polimorfismo VNTR (*variable number tandem repeat*) pentaalélico de 86 pb, se asociaron con un mayor riesgo de cáncer gástrico y a un aumento de la respuesta hipoclorhídrica crónica a la infección por *H.pylori* (127).

En el mismo estudio, el polimorfismo *IL1B* c.-511C>T, también en el promotor del gen, se asoció a mayor riesgo de cáncer gástrico, aunque este efecto puede ser debido al desequilibrio de ligamiento existente entre este polimorfismo y *IL1B* c.-31C>T. Las citocinas son importantes mediadores de la fisiología gástrica y desempeñan un papel clave en la etiología del cáncer gástrico. Polimorfismos en estos genes pueden modificar la respuesta inflamatoria y contribuir a las variaciones individuales en las consecuencias clínicas de la infección, por lo que se han estudiado extensamente. Posteriormente, se han publicado numerosos estudios de asociación entre riesgo de cáncer gástrico o lesiones precursoras y polimorfismos en genes que codifican para estas y otras citocinas y en otros genes relacionados con la respuesta inflamatoria: *IL1B*, *IL1RN*, *IL4*, *IL4R*, *IL8*, *IL10*, *IL12A*, *IL12B*, *IL10*, *LTA* y *TNFA*, entre otros. En algunos casos se reportaron asociaciones positivas, especialmente con el riesgo de cáncer gástrico del tipo intestinal, aunque existen importantes discrepancias entre estudios (128,129). Más recientemente se han llevado a cabo una serie de meta-análisis en un intento por consolidar la información existente entre los factores de riesgo genético relacionados con los genes de la respuesta inflamatoria y la carcinogénesis gástrica. Los resultados de estos análisis también son heterogéneos, debido en muchos casos a la población estudiada, caucásica o asiática, a la existencia de desequilibrio de ligamiento entre algunos de los polimorfismos, al tamaño de la muestra o al tipo de estratificación por subtipo realizada en cada estudio individual. A pesar de esta heterogeneidad, alguno de los polimorfismos se han asociado más consistentemente con un mayor riesgo de cáncer gástrico, especialmente en población caucásica: *IL1B* c.-511C>T, en desequilibrio de ligamiento con *IL1B* c.-31C>T, el VNTR de *IL1RN* y *TNFA* c.-308G>A (130-134). Un metaanálisis reciente también ha confirmado la asociación entre el genotipo *IL1RN* 2/2 y lesiones precursoras de cáncer gástrico (135). Diversos estudios epidemiológicos también han centrado su atención en la variabilidad genética en genes involucrados en la reparación del DNA y su asociación con la carcinogénesis gástrica. Algunos procesos pueden inducir mutaciones en el DNA como la exposición a carcinógenos, una replicación inadecuada, la pérdida de bases debido a la rotura espontánea de las uniones químicas y la exposición a ciertos compuestos endógenos, como las especies reactivas al oxígeno. Si las células no son capaces de reparar el daño provocado, estas mutaciones pueden contribuir al desarrollo carcinogénico. Polimorfismos en genes de las diversas vías de reparación del DNA, *MSH2*, *MLH1*, *XRCC1*, *OGG1*, *ERCC1* y *ERCC2*, se han asociado significativamente con cáncer gástrico, aunque, como en el caso anterior, los resultados son en muchas ocasiones conflictivos (136). En los últimos años se han reportado diversos estudios de asociación discrepantes entre los polimorfismos potencialmente funcionales de *XRCC1*, Arg194Trp, Arg280His y Arg399Gln, y el riesgo de cáncer gástrico. Sin embargo, dos metaanálisis recientes no han encontrado una asociación significativa, concluyendo que estos factores genéticos no alteran la susceptibilidad al cáncer gástrico (137,138). Otro reciente metaanálisis de los polimorfismos de *ERCC2*, Lys751Gln y Asp312Asn., desvela que, en este, caso el homocigoto para la variante 312Asn es un factor de riesgo para el desarrollo de cáncer gástrico (139) en población asiática. Las diferencias en el origen étnico de las poblaciones incluidas en los diferentes estudios de asociación podrían ser, en parte, la causa de los

resultados discrepantes que recurrentemente se obtienen. Otro grupo de polimorfismos que se han analizado son los localizados en genes responsables del reconocimiento de *H.pylori* y mediadores de la respuesta inmune innata, *TLRs* y *NODs*. Estos polimorfismos pueden modificar la capacidad de los receptores para reconocer *H.pylori*, o sus derivados y, consecuentemente, provocar defectos en la señalización celular que transmiten en respuesta. Los polimorfismos de cambio de aminoácido Asp299Gly y Thr399Ile en el gen *TLR4* se han asociado a cáncer gástrico y también a lesiones precursoras, especialmente a la gastritis inducida por *H.pylori*, únicamente en población caucásica, ya que ninguno de los dos polimorfismos está presente en población asiática (140-142). Estos SNPs alteran la estructura normal de la región extracelular de *TLR4* y, por tanto, se ha propuesto que pueden disminuir capacidad de unión y respuesta a lipopolisacáridos del patógeno (143,144). En cuanto a polimorfismos en los genes *NOD1* y *NOD2*, pocos estudios han abordado su posible relación con la carcinogénesis gástrica. Un SNP de cambio de aminoácido, *NOD1* Glu266Lys, se ha asociado con las consecuencias clínicas de la infección por *H.pylori*, úlcera péptica, atrofia antral y metaplasia intestinal. Cuatro polimorfismos en el gen *NOD2* también se han asociado significativamente con las lesiones precursoras de cáncer gástrico, aunque los resultados de los diferentes estudios son contradictorios. El alelo menos frecuente de uno de estos polimorfismos, *NOD2* Arg702Trp, resulta en una inducción reducida de NF- κ B en respuesta a *H.pylori* (145,146). Recientemente, dos tagSNPs de *NOD2* se han asociado como factores protectores de cáncer gástrico del no cardias (147). Los GSTs (*glutathione S-transferases*) son enzimas metabólicas de fase II que desempeñan un papel muy importante en la protección contra el cáncer mediante la detoxificación de numerosos compuestos potencialmente cito o genotóxicos. Algunos de los genes que codifican para las distintas isoenzimas de GST (*GSTM1*, *GSTT1* y *GSTP1*) presentan unos polimorfismos de delección, que resultan en la pérdida de la actividad de la enzima. Estos polimorfismos se han asociado consistentemente con un mayor riesgo de cáncer gástrico aunque algunos resultados son discrepantes entre diferentes poblaciones y áreas geográficas (148-151). No obstante, un reciente meta-análisis ha confirmado la asociación del genotipo de *GSTT1* que presenta la delección con un mayor riesgo de cáncer gástrico en población asiática y europea (152).

La variabilidad genética en genes de protección de la mucosa gástrica (*MUCs* y *TFFs*) también puede ser un factor influyente en el riesgo de cáncer gástrico. Estos genes codifican para las principales proteínas de formación y mantenimiento de la mucosa gástrica, mucinas y péptidos trefoil, que ejercen un papel importante en la protección del epitelio contra los patógenos y otras agresiones externas. Dado que estos genes son el objetivo de estudio de esta tesis, la variabilidad genética asociada a los mismos se trata en un los APARTADOS 2.5.3, para *TFFs*, y 3.5.3, para *MUCs*.

1.5.3.2 GWAs

Además de los estudios de asociación caso-control o de cohorte que utilizan una estrategia de gen candidato, incluyendo polimorfismos potencialmente funcionales o tagSNPs,

recientemente se han publicado cuatro análisis de asociación de genoma completo relacionados con el cáncer gástrico, todos ellos en población asiática (153-156). Los GWAs, en los últimos años, han emergido como una importante herramienta para identificar regiones en el genoma asociadas a diversas enfermedades. Estos estudios una serie de SNPs marcadores, cientos de miles, en un fuerte desequilibrio de ligamiento con muchas otras variantes a lo largo de todo el genoma. Generalmente, se dividen en dos fases: una primera en que se incluyen todos los polimorfismos marcadores y una segunda donde se replican las asociaciones más significativas en otras poblaciones (157,158). El primer GWA en cáncer gástrico se publicó en 2008 e identificó el polimorfismo rs2976392, que marcaba un bloque de desequilibrio de ligamiento en la región cromosómica 8q24.3, como fuertemente asociado al cáncer gástrico de tipo difuso, incluso después de la corrección por Bonferroni. Concretamente, este SNP se localizaba en la región intrónica del gen que codifica para PSCA (*prostate stem cell-antigen*), una proteína ancorada a la membrana de células epiteliales y relacionada anteriormente con otros tipos de cáncer, aunque su función es todavía poco conocida (153). La resecuenciación del gen y genotipado de alta densidad permitió identificar otros polimorfismos de PSCA fuertemente asociados, entre ellos un SNP, rs2294008 (C/T), en la región 5'UTR. El alelo T de este polimorfismo, asociado con un mayor riesgo de cáncer gástrico, introduce una metionina de inicio nueve codones antes que la variante C. No obstante, la única funcionalidad conocida para la variante T es una reducción de la actividad transcripcional del gen (153). La asociación de este polimorfismo con cáncer gástrico se ha validado en otras poblaciones asiáticas (159-163) y, recientemente, también en población caucásica, en la que también se ha encontrado asociación con lesiones precursoras (164,165). El mismo GWAS también reveló dos polimorfismos en otra región, 1q22, fuertemente asociados con cáncer gástrico de tipo difuso, aunque su análisis se abordó en un estudio posterior (166). Mediante un mapeo de alta densidad en la región 1q22, marcada por los dos SNPs asociados en el GWA, se identificó un bloque de desequilibrio de ligamiento de 49 kb que incluía 8 tagSNPs asociados significativamente con el tipo difuso. Este bloque incluía 5 genes, entre ellos *MUC1*, y uno de los SNPs asociados era rs4072037, cuyo efecto funcional ya se había descrito (167). La asociación para este polimorfismo se replicó también en otras poblaciones asiáticas y europeas (166,168). De esta manera, el SNP rs4072037 de *MUC1* se confirma como un posible factor riesgo para cáncer gástrico por las evidencias tanto de los estudios de gen candidato como de los GWAs. Posteriormente, se realizaron dos GWA paralelos en carcinoma de esófago de células escamosas, que comparte muchas características con el cáncer gástrico del cardias en términos de distribución geográfica y factores de riesgo ambientales (154,155). Algunos de los polimorfismos asociados a mayor riesgo de carcinoma de esófago en la primera fase se replicaron en un grupo de casos de cáncer gástrico en población asiática y europea (168,169). De esta manera se detectaron dos loci, no descritos anteriormente, de riesgo compartido por ambos tipos de cáncer, *PCLE1* en la región cromosómica 10q23 y *C20orf54* en 20p13. *PCLE1* se ha asociado con procesos de gran importancia biológica como la regulación del crecimiento celular, diferenciación, apoptosis y angiogénesis. El homólogo de *C20orf54*, conocido también como *RFT2*, presenta gran homología con el *Rtf2* de rata que tiene un importante papel en la absorción

intestinal de la riboflavina, cuya deficiencia se asoció con un incremento en el riesgo tanto de carcinoma de esófago de células escamosas como de cáncer gástrico del cardias (154,155). Finalmente, un último GWA llevado a cabo en población china ha identificado también dos nuevos locus de susceptibilidad genética para cáncer gástrico no cardias. Los dos polimorfismos significativamente asociados fueron el rs13361707, localizado en el gen *PRKAA*, en la región 5p13.1, aunque la señal significativa también incluiría el gen *PTGER4*, y el rs9841504 en el gen *ZBTB20*, en la región 3q13.31 (156). Todas las asociaciones encontradas se replicaron en estudios posteriores en población asiática (170,171). A pesar de los nuevos locus identificados mediante GWAs, todavía se conoce muy poco de los mecanismos por los cuales gran parte de esta variabilidad contribuye al riesgo de cáncer gástrico.

2. TREFOIL FACTOR FAMILY

Los componentes de la familia de los péptidos trefoil o *trefoil factors* (TFFs) tienen un papel esencial en procesos de restitución epitelial en el tracto gastrointestinal y están involucrados en el mantenimiento y reparación de la mucosa gástrica. Esta familia está comprendida por los péptidos gástricos TFF1, o pS2, y TFF2, o polipéptido espasmolítico (SP), y el péptido intestinal TFF3, o ITF. Los tres miembros de la familia son pequeñas proteínas (7-14 kDa) resistentes a proteasas que se secretan abundantemente en la superficie mucosa del tracto gastrointestinal por células especializadas de la misma. Generalmente se expresan en asociación con las mucinas, el mayor componente del gel de mucus que recubre la mucosa.

2.1. Estructura de los péptidos trefoil

Los TFFs comparten un motivo muy particular altamente conservado de seis residuos de cisteína, denominado dominio trefoil o dominio P. Actualmente, con todos los datos estructurales disponibles, este se define como una secuencia constituida por 42-43 residuos, de manera que, cuando las seis cisteínas son numeradas desde el extremo N-terminal, se forman tres enlaces disulfuro intracatenarios en una configuración 1-5, 2-4 y 3-6 (172). El resultado final es la formación de una estructura de 3 bucles compacta (FIGURA I-3) que contribuye a la resistencia de los TFFs a digestión proteolítica y degradación por ácidos. Esta propiedad les permite mantener la funcionalidad en un entorno desfavorable como es el lumen del tracto gastrointestinal (173). Una séptima cisteína no emparejada, a 3 residuos del extremo C-terminal de TFF1 y TFF3 (FIGURA I-3), media la formación de homo y heterodímeros que son de gran importancia para la actividad de estos dos péptidos (174).

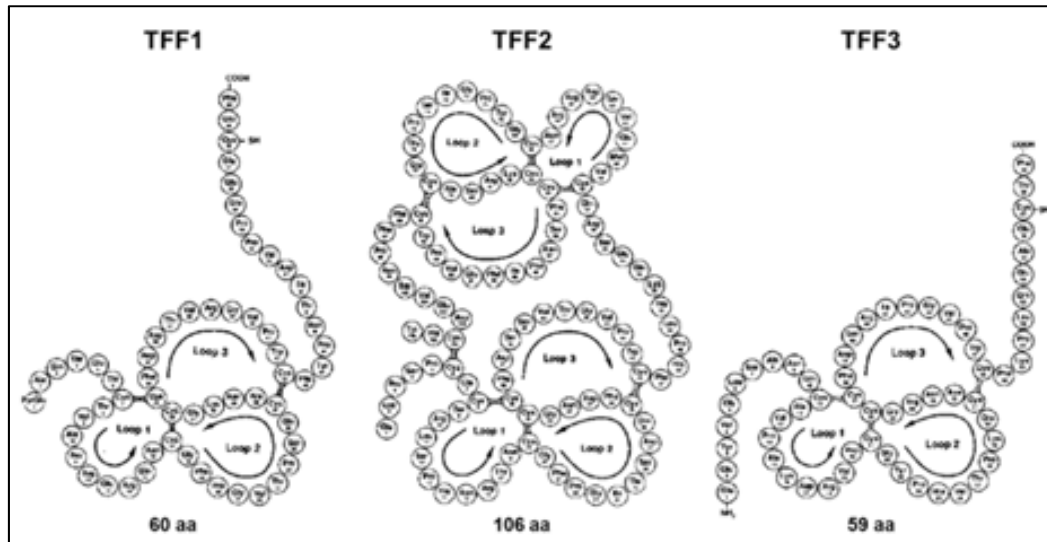


Figura I-3. Estructura de los péptidos trefoil (Thim 1997, (175))

En la mucosa gástrica normal TFF1 se encuentra habitualmente en su forma heterodimérica unido covalentemente a la proteína TFIZ1 (176). TFF2, que tiene dos dominios trefoil unidos a través de un enlace disulfuro entre dos cisteínas adicionales, puede realizar eficientemente su función en su forma monomérica. TFF2 se diferencia de las formas diméricas de TFF1 y TFF3 porque sus dominios trefoil no sólo están unidos por los enlaces disulfuro sino también por un enlace peptídico entre un residuo de leucina y uno de ácido aspártico localizados entre los dos dominios trefoil, confiriéndole una estructura muy compacta y extremadamente resistente contra el ácido y proteasas como la tripsina (177). Para todos los TFFs de mamíferos, existe una tendencia de un mayor grado de conservación en los dominios trefoil que en otras partes de la molécula, con la excepción del extremo carboxi-terminal de TFF3. Además de los seis residuos de cisteína, otros residuos en el dominio trefoil están conservados entre TFF1, TFF2 y TFF3 de la mayoría de especies de mamíferos secuenciadas, indicando que estos aminoácidos están restringidos por importantes limitaciones estructurales o funcionales. Por otra parte, el número de aminoácidos en el péptido maduro de diferentes especies es el mismo: 60 en TFF1, 106 en TFF2 y 59 en TFF3. Los tres péptidos se sintetizan con una secuencia señal de entre 12 y 27 aminoácidos que se elimina por proteólisis durante su paso por el retículo endoplásmico (172).

2.2. Estructura de los genes *TFF*

Los genes que codifican para los péptidos trefoil se denominan *TFF1* (GeneID:7031), *TFF2* (GeneID:7032), y *TFF3* (GeneID:7032). Estos genes se localizan en un cluster de 55 kb en la región cromosómica 21q22.3 (178). Los genes *TFF1* y *TFF3* tienen 3 exones y abarcan una longitud de 4254 y 3929 pb, respectivamente (FIGURA M-3 del APARTADO 4 de materiales y métodos). Los exones 1 y 3 codifican para los extremos amino y carboxiterminal y el exón 2 codifica para el dominio *TFF*. El gen *TFF2* tiene una longitud total de 4743 bp y contiene 4 exones y dos dominios trefoil, codificados por los exones 2 y 3, mientras que los extremos amino

y carboxiterminal están codificados por los exones 1 y 4, respectivamente. Los dos dominios trefoil homólogos de *TFF2* probablemente deriven de duplicación genómica. Estos genes se transcriben en pequeños mRNA de 0,5-0,6 kb (177).

La organización genómica de los tres *TFFs* de humanos y ratones es homóloga y se han conservado elementos reguladores en las regiones proximales del promotor de todos los *TFFs* entre estas especies. Además, los miembros *TFF* están altamente conservados evolutivamente con más del 70% de identidad en la secuencia aminoacídica entre estas dos especies. Aunque la identidad de secuencia entre los miembros de *TFF* en una misma especie es menor, y se ha propuesto que estos tres genes podrían haber evolucionado de un mismo ancestro común por duplicación. De hecho, los tres genes tienen una estructura exónica similar y la región 5'UTR de todos los *TFFs* contiene secuencias con elementos reguladores muy similares, sugiriendo que tienen una regulación común o coordinada (179).

2.3. Expresión de los TFFs

2.3.1. Localización tisular

Los *TFFs* se expresan esencialmente en el epitelio gastrointestinal, siendo el estómago el único órgano donde se expresan los tres miembros de la familia. No obstante, estos péptidos, especialmente *TFF3*, se han observado en prácticamente todos los tejidos que contienen células secretoras de mucus (180). En los tejidos normales tanto las mucinas como *TFFs* presentan un modelo de expresión celular y tisular específicos sugiriendo una relación en las funciones de estas proteínas. En el tracto gastrointestinal cada *TFF* se coexpresa con una mucina específica, *TFF1* con *MUC5AC*, *TFF2* con *MUC6* y *TFF3* con *MUC2* (181). Aunque existen algunas excepciones, el patrón de expresión está altamente conservado entre especies. *TFF1* se expresa abundantemente en las células epiteliales superficiales incluyendo las foveolas gástricas y, en menor medida, en la parte superior de las glándulas de Brunner del duodeno (182). Otras localizaciones más minoritarias de *TFF1* son el esófago y el intestino delgado y grueso (177). *TFF2* también se ha detectado ocasionalmente en la superficie epitelial del estómago, pero su mayor nivel de expresión se encuentra en las células mucosas del cuello de las glándulas oxínticas, en las glándulas pilóricas y en las glándulas de Brunner (183). En condiciones normales, *TFF3* se expresa abundantemente en la región apical de las células caliciformes, en el intestino delgado y grueso (184) y, por el contrario, su expresión es muy escasa en el estómago, donde está confinada a las células mucosas de superficie, específicamente en el cardia y antro. Esta expresión parece estar relacionada con el proceso de maduración del epitelio superficial gástrico (185).

Además de en el tracto gastrointestinal, se ha descrito niveles variables de expresión de *TFFs* en otras localizaciones, particularmente en tejidos y glándulas que contienen células secretoras de mucus como el respiratorio y el ocular, glándulas salivales, próstata y mama. También se ha descrito la expresión de estos péptidos en órganos como páncreas e hígado, entre otros, y en localizaciones mucho más minoritarias como el cerebro, la tiroides y el músculo

(177,180). Por tanto, la amplia distribución de la expresión de los TFFs, especialmente de TFF1 y TFF3, sugiere que estos péptidos ejercen múltiples funciones, aunque su relevancia fisiológica en la mayoría de localizaciones se desconoce. El hecho de que la expresión de los TFFs sea complementaria en el tracto gastrointestinal donde, además, cada miembro se coexpresa con una única mucina sugiere que cada TFF puede tener un papel distinto en la protección del epitelio. En condiciones patológicas del tracto gastrointestinal o daño de la mucosa, como enfermedades inflamatorias, ulceración o cáncer, el patrón de expresión de estos péptidos se altera con frecuencia (177) (ver APARTADO 2.5.1).

2.3.2. Regulación de la expresión

Los *TFFs* siguen un patrón de expresión específico en diferentes tejidos donde cada *TFF* puede tener mecanismos individuales de regulación a nivel transcripcional. Además de la expresión específica de tejido de cada TFF individualmente en condiciones normales, existen diversos datos que sugieren una expresión de TFFs coordinada, como su localización contigua en una distancia corta (55 kb) en el cromosoma, que incrementan su expresión rápidamente cuando se produce una lesión en la mucosa o la correlación de expresión que existe en ratones *knockout* para estos genes (ver apartado 2.4.4) (177).

Aunque todavía se desconocen parte de los mecanismos de regulación para cada uno de los genes TFF, el análisis de los promotores ha permitido la identificación de diversos factores de transcripción y otros elementos implicados en su regulación. El promotor del gen *TFF1* contiene una región compleja de estimuladores de la transcripción (*enhancers*) en respuesta a diversos agentes, como estrógenos y EGF (186). Además, los factores de transcripción GATA6 y HNF-3 han sido implicados en su regulación (187,188). *TFF1* es un gen clásicamente regulado por estrógenos, principalmente en mama y útero. El efecto del estrógeno está mediado por sus receptores alfa y beta y actúa regulando la expresión de *TFF1* mediante la activación de elementos de respuesta al estrógeno (ERE) en el promotor del gen, que además puede ser incrementada por la presencia de sitios de unión a Sp1 y Sp3 (189). Aunque ambos receptores del estrógeno están presentes en el estómago, no existen evidencias claras de que TFF1 sea un gen regulado por estrógeno en este tejido y existen muy pocos estudios al respecto (190). El promotor de *TFF2* también es un promotor complejo con elementos *enhancer* de respuesta a diversos agentes, como factores de crecimiento y hormonas, y a factores de transcripción basal como GATA-6, responsables de su expresión específica en el tracto gastrointestinal. Además TFF2 es capaz de autorregular autoinducir su propia expresión mediante la activación de su promotor. (188,191,192). El promotor de *TFF3* también contiene diversos elementos reguladores en *cis* como *enhancers* y regiones silenciadoras, específicos de células caliciformes (193). Estas regiones se unen específicamente a proteínas nucleares que se encuentran en células caliciformes intestinales, pudiendo ser la base para su expresión tisular y celular específica (194,195). Uno de los factores ya identificados que regulan a TFF3 mediante este mecanismo es el factor de crecimiento de los queratinocitos (KGF). KGF promueve la diferenciación de células

calciformes y la transcripción del gen *TFF3* mediante la inducción de la proteína de unión a GCSI (196).

Diversos estudios han demostrado que numerosas citocinas y otros efectores del sistema inmune pueden regular los TFFs, aunque existen datos controvertidos, probablemente a causa de las diferentes líneas celulares utilizadas, los diferentes subtipos de efectores estudiados o las diferentes consecuencias de su activación. También puede ser debido al diferente papel que puedan ejercer las citocinas y los cambios de expresión de TFFs durante las diferentes fases de la inflamación y procesos de restitución (177). Las citocinas proinflamatorias IL-1 β y, particularmente, IL-6 son importantes reguladores de la expresión de los genes trefoil. Un estudio *in vitro* llevado a cabo por Dossinger *et al.* mostró que ambas citocinas pueden reprimir sinérgicamente la actividad del promotor de estos genes y, consecuentemente, su expresión mediante la inhibición de los factores de transcripción NF κ B y C/EBP β , respectivamente (197). Por otro lado, datos procedentes de un estudio llevado a cabo por Tebbut *et al.*, que implicaba a gp130, un receptor transductor de señales que actúa conjuntamente con el receptor de IL-6, mostraron que la expresión de *TFF1* está positivamente regulada por IL-6 mediante la activación de SHP2-Ras-Erk. En el mismo estudio se mostró que IL-6 puede inducir fuertemente la transcripción de *TFF3* pero, en este caso, mediante la vía de señalización de STAT1/3 (198). La citocina TNF- α también puede inhibir la transcripción del gen *TFF3* mediante NF- κ B (199). De hecho, el promotor de *TFF3* presenta cuatro sitios potenciales de unión para NF- κ B, el que se sitúa a -97 del inicio de transcripción parece ser el más importante, y 2 para C/EBP β (200). El efecto de la citocina proinflamatoria TNF- α en *TFF1* es opuesto al de *TFF3*, ya que su expresión aumenta de manera dosis dependiente. El sitio de unión a NF- κ B situado a -231 en el promotor de *TFF1* es esencial para la respuesta del gen a TNF- α (201). En el tracto gastrointestinal las citosinas tipo Th2 se expresan como parte de la respuesta inmune y tienen como dianas importantes las células calciformes. Dos de estas citocinas, IL-13 e IL-14, incrementan la transcripción del gen *TFF3* en líneas celulares intestinales, de manera dosis dependiente y en coordinación con la mucina *MUC2*. STAT6 media este aumento de la transcripción e induce la síntesis de *novoo* de la proteína TFF3 en células calciformes (202).

Otro de los reguladores mejor definidos de TFFs es la gastrina. La gastrina es una hormona peptídica secretada principalmente por las células G endocrinas del estómago que induce la secreción de ácido y, por tanto, mantiene el pH ácido apropiado para el funcionamiento de las enzimas digestivas gástricas. La gastrina, además, actúa como factor de crecimiento en la mucosa normal oxíntica gástrica estimulando la proliferación y migración de las células parietales y de las células ECL productoras de histamina. Además, la expresión de la gastrina aumenta durante la infección por *H. pylori* y durante las últimas décadas diversos grupos han relacionado esta hormona con la promoción de la carcinogénesis gástrica (102). *TFF1* es uno de los genes sensible a la gastrina. La expresión de *TFF1* se incrementa rápida y potentemente en respuesta a la gastrina sugiriendo un efecto a nivel transcripcional. *TFF1* puede jugar un papel importante en la supresión de la proliferación celular, por lo que su regulación por la hormona gastrina parece ser un mecanismo regulador para contrarrestar los efectos

proliferativos del potente factor de crecimiento gástrico (203). La capacidad de respuesta a la gastrina se localizó en una región del promotor rica en el dinucleótido GC y fue dependiente de los elementos reguladores solapados en *cis* SP1 y MAZ. La activación de *TFF1* es dependiente de la vía MAPK/ERK. La expresión de *TFF2* también responde fuertemente a un rango fisiológico normal de gastrina circulante. La gastrina puede regular transcripcionalmente la expresión de *TFF2* mediante dos elementos en *cis* de respuesta a la hormona en el promotor y mediante la vía PKC y MEK, aunque, de la misma manera que *TFF1*, es independiente de EGFR (204).

La regulación epigenética también parece ser uno de los mecanismos responsables de la expresión específica tisular de genes trefoil, concretamente la metilación. En el estómago, los promotores proximales de *TFF1* y *TFF2*, que se expresan abundantemente en esta localización, no están prácticamente metilados mientras que en otros tejidos, donde no se expresan, sus promotores están hipermetilados. En el caso de *TFF3*, el menor grado de metilación se encuentra en el intestino y el páncreas (205). El patrón de metilación de los promotores de estos genes está alterado en condiciones patológicas, ver APARTADO 2.5.

2.4. Funciones de los TFFs en el mantenimiento de la mucosa

En el tracto gastrointestinal existen numerosos agentes lesivos como el ácido gástrico o las bacterias. El epitelio gastrointestinal forma una barrera física contra estos agentes aunque frecuentemente la capa superficial de células puede verse dañada por lo que se necesitan mecanismos para restituir rápidamente la continuidad epitelial. Uno de los procesos de reparación requeridos es la restitución celular. Este proceso promueve la migración de las células epiteliales hacia la zona dañada para poder cubrir nuevamente la superficie después de la agresión. Este mecanismo involucra un rápido desprendimiento de las adhesiones entre células y entre la célula y el sustrato, desdiferenciación y extensión de las células superficiales. Durante la restitución, las células epiteliales se propagan y migran a través de la membrana basal para restablecer la continuidad de la superficie celular, un proceso independiente la proliferación celular. Esta rápida restitución después de la agresión limita las pérdidas de fluidos y electrolitos y evita la inflamación que causaría la exposición directa de la lámina propia a bacterias y otros agentes exteriores del lumen gastrointestinal (173).

Los TFFs son unos péptidos cruciales en el proceso de **protección y restitución de la mucosa durante la homeostasis gástrica normal** aunque, a pesar de los numerosos estudios funcionales realizados, el modelo de acción de los TFFs *in vivo* todavía no está completo. No obstante, los datos sugieren que los péptidos trefoil ejercen múltiples funciones en procesos de proliferación, apoptosis, migración y angiogénesis. Además se ha demostrado que los TFFs interactúan con las mucinas, contribuyendo a la formación de la barrera de mucus, e interfieren en los procesos inflamatorios. La expresión y liberación de trefoil factors se incrementa rápidamente en respuesta a una lesión epitelial actuando en determinadas circunstancias como genes de expresión rápida (173,206).

Algunas de las evidencias que relacionan los TFFs con la protección, mantenimiento y reparación de la mucosa provienen de los estudios en modelos animales. Los ratones *knockout* de

TFF1 desarrollan hiperplasia y displasia severa de la mucosa gástrica provocando la pérdida casi completa del mucus y la aparición de adenomas y carcinomas (207). Además, la expresión gástrica de TFF2 está ausente en las dos terceras partes de ratones *knockout* para TFF1, evidenciando el potencial de corregulación genética de TFFs (207). En contraste con los ratones deficientes en TFF1, los ratones *knockout* de TFF2 exhiben alteraciones gástricas más leves, como una disminución del grosor de la mucosa, menor proliferación celular, incremento de la secreción ácida y mayor susceptibilidad a las lesiones provocadas por AINES (208). Además, la mayoría de genes diferencialmente expresados en estos ratones, en comparación con ratones salvajes, corresponden al sistema inmune sugiriendo que TFF2, además de tener efectos protectores de la mucosa, también interacciona con la respuesta inmunitaria (209,210). La expresión de TFF3 en ratones *knockout* de TFF2 aumenta considerablemente en el estómago, mostrando una correlación de la expresión entre genes TFF aunque en la dirección opuesta comparado con la disminución de expresión de TFF2 en ratones deficientes en TFF1. De manera similar, los ratones *knockout* de TFF3 no tienen un fenotipo muy marcado excepto una defensa de la mucosa intestinal deficiente, donde la restitución está ausente, siendo más susceptibles a determinados agentes lesivos. Este defecto en la restitución puede revertirse mediante administración tópica de TFF3 recombinante, que reconstituye la regeneración normal de la mucosa asociada a una mayor migración celular. El ratón deficiente en TFF3 presenta una síntesis reducida tanto de TFF1 como de TFF2 (211).

Existen algunos datos que muestran la habilidad de los TFFs para **modular la progresión del ciclo celular**. Bossenmeyer-Pourie *et al.* demostró que TFF1 mantiene las células en la fase G1 del ciclo celular, tanto de manera paracrina como autocrina, y retrasa, por tanto, la entrada a la fase S reduciendo la proliferación celular en células gastrointestinales (212). También se ha reportado una disminución de la proliferación en células gastrointestinales transfectadas con TFF3 (213). El efecto antiproliferativo de estos TFFs no se puede extrapolar a otros tejidos ya que en mama, por ejemplo, se ha descrito un efecto contrario para ambos péptidos (214,215) y se consideran factores oncogénicos.

Las **propiedades antiapoptóticas** de los péptidos trefoil son cruciales para la restitución epitelial, en que las células tienen que desprenderse y migrar hacia el área lesionada. Se ha descrito un efecto antiapoptótico de TFF1 en diversas líneas celulares gastrointestinales (177). TFF1 es capaz de proteger a las células de la apoptosis inducida por diversos estímulos como químicos, la desunión de las células de su matriz o la proteína Bad, un miembro de la familia Bcl-2. En este proceso, la señalización de TFF1 tiene como diana la forma activa de la caspasa-9 (212). TFF1 representa un importante factor regulador del equilibrio entre la proliferación, muerte y diferenciación celular gastrointestinal (216). Estas observaciones proporcionan evidencias de un papel complejo de TFF1, no sólo como un potente inhibidor de la proliferación sino también como un eficiente inhibidor de la apoptosis, funciones duales y, a la vez, paradójicas un supresor de tumores. Para TFF2 también se ha observado un potencial significativo para reducir la apoptosis en células gastrointestinales, entre otras (217).

El desprendimiento de las células de la matriz extracelular o la pérdida de adherencia provoca una forma de apoptosis denominada anoikis. El anoikis es un mecanismo importante de regulación de la homeostasis epitelial y del mantenimiento de la integridad de la mucosa normal (218). Se ha demostrado que TFF3 protege a las células epiteliales del anoikis mediante la activación de NFκB (219). Este efecto antiapoptótico requiere un dímero intacto de TFF3 y la activación de las vías de señalización PI3K y EGF-R (220). Además la apoptosis espontánea de enterocitos se incrementa en los ratones deficientes en TFF3 y la exposición a TFF3 exógeno protege a las células epiteliales de la apoptosis inducida por agentes citotóxicos *in vitro* (221).

La migración es esencial para el proceso de restitución celular en que las células migran para cubrir las áreas lesionadas de la mucosa manteniendo, así, la homeostasis gastrointestinal. En concordancia con el papel de los TFFs en la restitución, todos los miembros de la familia **pueden promover la migración en líneas celulares gastrointestinales**. Este efecto ejercido por TFFs es independiente de TGF-beta, que actúa como mediador de la mayoría de otros péptidos promigratorios (222). Los primeros estudios realizados en este campo ya mostraron que la función promigratoria de TFFs es dependiente de la presencia del complejo E-caderina/B-catenina, que media la adhesión intercelular en el epitelio. Se ha reportado que tanto TFF2 como TFF3 regulan negativamente la formación del complejo E-caderina-catenina en células gastrointestinales, contribuyendo a una disminución de la adhesión célula-célula y célula-matriz extracelular que facilita la migración celular (223,224). EGFR también regula negativamente el complejo E-caderina/B-catenina y se ha observado un efecto sinérgico juntamente con TFF1 o TFF3 en la migración celular (225).

También se han descrito **propiedades proangiogénicas** para todos los TFFs, con un potencial comparable al de VEGF. Este efecto es dependiente de la señalización de COX2 y EGFR (226). La restitución *in vivo* depende de la capacidad de mantener un flujo de sangre continuo en la mucosa y la angiogénesis es un proceso que se observa habitualmente cuando el daño en la mucosa sobrepasa la superficie epitelial. La capacidad de los TFFs para actuar como factores proangiogénicos, tanto *in vitro* como *in vivo*, sugiere una interacción directa entre estos péptidos gastrointestinales y las células endoteliales y puede contribuir a procesos fisiológicos en los que la neovascularización juega un papel importante. Además, en células gástricas, los TFFs pueden ser inducidos por el factor 1 inducible por hipoxia (HIF1, *de hypoxia inducing factor 1*) en respuesta a determinadas condiciones lesivas, estableciendo otra conexión con el endotelio (227).

Los efectos de TFFs se han asociado con la **regulación de la expresión de mediadores de la inflamación y la inmunidad** como iNOS y COX. TFF2 disminuye la expresión de iNOS inducida por lipopolisacáridos y la producción de óxido nítrico en monocitos (228). Por el contrario, se ha reportado que TFF3 incrementa la producción de óxido nítrico, mediante el aumento de expresión de iNOS, un efecto que está potenciado por las mucinas (229). El óxido nítrico regula diversas funciones fisiológicas y es un mediador de la inflamación. Además, su efecto puede ser dual, mientras que en unas condiciones determinadas puede mediar las lesiones intestinales, en otras actúa como un agente citoprotector en el sistema gastrointestinal,

participando también en la restitución epitelial (63). La ciclooxigenasa es una enzima clave en la biosíntesis de la prostaglandina. COX2 se expresa normalmente a bajas concentraciones pero es rápidamente inducida por citocinas y otros estímulos inflamatorios, siendo la responsable de la elevada producción de prostaglandinas durante la inflamación. Las prostaglandinas son importantes moléculas citoprotectoras en el tracto gastrointestinal y juegan un importante papel en el mantenimiento de la integridad de la mucosa (230). TFF3 incrementa la expresión de COX2 e induce la producción de prostaglandinas en células epiteliales intestinales. Además, TFF3 también tiene un importante papel en la protección del epitelio intestinal contra las lesiones causadas por especies reactivas del oxígeno, propiedad atribuida a la inducción de COX2 por TFF3 (231).

Numerosos estudios confirman la **cooperación entre mucinas y TFFs en la formación y mantenimiento de la barrera mucosa gastrointestinal**. Como se ha comentado anteriormente, cada TFF colocaliza con una única mucina de manera específica y se ha demostrado que ambos protegen las monocapas celulares contra agentes perjudiciales, disminuyen la permeabilidad de los protones a través de la barrera mucosa e incrementan la viscosidad del mucus (232,233). Además, la interacción entre TFFs y mucinas puede conllevar otros efectos a nivel celular como la estimulación de la migración o la producción de óxido nítrico (222,229).

2.5. TFFs y la carcinogénesis gástrica

2.5.1. Expresión de TFFs en condiciones patológicas

En los tejidos normales, tanto las mucinas como los péptidos TFF presentan un modelo de expresión celular y tisular específicos (ver APARTADO 2.3). Estos patrones de expresión normal son modificados en las lesiones premalignas y malignas del estómago (234-237). Los TFFs son rápidamente inducidos cuando se produce una lesión en el epitelio gástrico, por lo que su expresión aumenta en las etapas iniciales de la lesión gástrica, durante la reparación de la mucosa. Existen numerosos estudios analizando la expresión anormal de estos genes en diversas condiciones patológicas del tracto gastrointestinal y en las distintas etapas de la carcinogénesis gástrica, aunque en algunos casos los resultados obtenidos son controvertidos, especialmente en el caso de TFF2. TFF1 disminuye significativamente su expresión en la mucosa infectada por *H.pylori*, en la mucosa preneoplásica y, en mayor medida, en cáncer gástrico (238-241). Esta disminución de la expresión esta ha asociado de TFF1 con la activación de la vía de NFκB y un incremento de la inflamación crónica durante las distintas fases de la carcinogénesis gástrica (242,243). Por el contrario, durante la progresión de la gastritis a metaplasia intestinal y cáncer gástrico aumenta la expresión del factor trefoil intestinal TFF3 (238,241,244,245). TFF2 disminuye su expresión, en ocasiones, en metaplasia intestinal y en algunos cánceres gástricos primarios, pero si se expresa en el linaje SPEM y en tumores gástricos invasivos y está fuertemente asociada con tumores de histología difusa (236,244-246)}. Aunque los TFFs tienen efectos beneficiosos en citoprotección y restitución, pueden tener consecuencias adversas cuando se sobreexpresan en tumores. Existe una correlación entre la sobreexpresión de TFF2 y

TFF3 y un peor pronóstico de carcinoma gástrico y neoangiogénesis (236,247). Además, se ha demostrado que los niveles de *TFF3* en suero son eficientes marcadores del cáncer gástrico en combinación con los de pepsinógeno (248,249).

2.5.2. Papel de los TFFs en la carcinogénesis gástrica

Muchos de los procesos en los que los TFFs participan para proteger y restituir la mucosa durante la homeostasis gástrica normal, son también procesos involucrados en el desarrollo del cáncer. Los péptidos TFFs activan diversas vías de señalización que promueven la migración celular, la supervivencia, angiogénesis y, cuando se sobreexpresan, la transformación celular (190).

Numerosas evidencias sugieren que *TFF1* podría ser un supresor de tumores específico de cáncer gástrico. Experimentos con modelos animales demostraron que el ratón *knockout* para *TFF1* desarrollaba una mucosa gástrica alterada y predisposición a cáncer gástrico (207)(ver APARTADO 2.4). Posteriormente se han descrito mutaciones somáticas, pérdida de heterocigosidad e hipermetilación como mecanismo de inactivación de la expresión de *TFF1*, observada en un elevado porcentaje de tumores gástricos (250,251). Recientemente, se ha demostrado la represión epigenética de *TFF1* es uno de los eventos tempranos que ocurre en el antro durante la carcinogénesis gástrica. Esta hipermetilación de *TFF1* podría ser revertida por la hormona gastrina ejerciendo un efecto supresor en la progresión de la carcinogénesis gástrica (240). La pérdida de expresión de *TFF1* por hipermetilación también se ha asociado con invasividad tumoral y baja supervivencia en pacientes de cáncer gástrico (252). Además, *TFF1* bloquea la transición de la fase G1 a S del ciclo celular, incrementa la expresión del gen Retinoblastoma y promueve la diferenciación gástrica (216), características que confirman su posible papel como supresor de tumores.

El papel de los otros TFFs en la carcinogénesis gástrica es mucho menos conocido. La expresión de *TFF2* se induce rápidamente en patologías gastrointestinales ulcerativas o después del tratamiento con antiinflamatorios no esteroideos (AINES) (208). Recientemente se ha demostrado la pérdida creciente de *TFF2* durante las distintas etapas de la carcinogénesis gástrica, acompañada de una progresiva hipermetilación del promotor (253). La metilación de *TFF2* se inicia probablemente durante la infección por *H.pylori* aunque no es causada directamente por la bacteria sino que es consecuencia de la inflamación dependiente de la infección, particularmente por la actividad de la citosina proinflamatoria IL-1 β (253). Se ha descrito también que p53 puede inducir la apoptosis celular e inhibir la migración celular, en parte, porque disminuye la expresión de *TFF2* a través del sitio de unión AP1. Este efecto proporcionaría un mecanismo por el cual mutaciones en p53 podrían incrementar la expresión de *TFF2* en células de cáncer gástrico (254). La inducción de *TFF3*, vía hipoxia y mediada por el factor HIF1, puede incrementar la expresión de VEGF y, consecuentemente, la angiogénesis en células de cáncer gástrico (255). También se ha demostrado que *TFF3* induce la migración de células de cáncer gástrico a través de twist, un factor de transcripción implicado en metástasis y que induce la transición epitelio-mesenquima (EMT) (256).

Los TFFs también están involucrados en el proceso de colonización por *H.pylori*. La bacteria interactúa directamente con la forma dimérica de TFF1, permitiendo la unión de la bacteria con la mucosa gástrica (257). Además TFF1 colocaliza con MUC5AC, cuyo antígeno de grupo sanguíneo Le^b se ha propuesto también como receptor de la bacteria en el estómago. Recientemente se ha identificado un componente bacteriano, un oligosacárido central del lipopolisacárido de *H.pylori*, que podría mediar la interacción con TFF1 y se ha sugerido que esta interacción podría ser un factor importante en la capacidad de *H.pylori* para colonizar la mucosa (258).

2.5.3. Variabilidad genética en TFFs y la carcinogénesis gástrica

En cuanto a la variabilidad genética en otros genes de protección de la mucosa gástrica, los TFFs, existen muy pocos trabajos. Un estudio realizado en población portuguesa analizó la posible asociación entre un VNTR dialélico de *TFF2* y el riesgo de cáncer gástrico, no encontrando una asociación significativa (259). Recientemente, se ha publicado un estudio caso-control de un SNP en el promotor de *TFF1*, c.-394 C/T, en que el genotipo TT se asoció a un mayor riesgo de cáncer gástrico (260). Otros dos SNPs (rs3814896 y rs9981660) en el promotor de *TFF2* y *TFF3*, respectivamente, se han asociado también recientemente con un menor riesgo de gastritis atrófica y/o cáncer gástrico del tipo difuso (261).

3. MUCINAS

Entre los genes que tienen un importante papel en las fases iniciales de la carcinogénesis gástrica se encuentran los de las principales proteínas de formación y mantenimiento de la mucosa gástrica como son las mucinas y los péptidos trefoil, que también están estrechamente relacionadas con la respuesta inflamatoria que se desencadena por la infección por *H.pylori*.

3.1. Estructura de las mucinas

Las mucinas son una familia de glicoproteínas de elevada masa molecular y alto contenido en carbohidratos (entre el 50 y 90% de su masa) que son sintetizadas por el epitelio secretor como proteínas de membrana o secretadas al medio, las cuales tienden a polimerizar y formar estructuras superiores de gran tamaño. Estas glicoproteínas se caracterizan por contener un dominio central de repeticiones de aminoácidos en tándem, variable en tamaño y número entre diferentes mucinas, en el que se unen cadenas de carbohidratos en los residuos Ser, Thr y Pro mediante enlaces O-glicósidos (262). Las mucinas también pueden clasificarse en neutrales o ácidas, que a la vez se categorizan en sialo o sulfomucinas dependiendo de la presencia de un grupo terminal de ácido siálico o sulfato, respectivamente, en las cadenas de oligosacáridos (263). La estructura proteica de las principales mucinas de la mucosa gástrica se muestra en la FIGURA I-4.

3.1.1. Mucinas de membrana

Las mucinas de la superficie celular se expresan en la membrana apical de las células epiteliales y son las constituyentes principales del glicocálix. Este tipo de mucinas son monoméricas y se pueden escindir y liberar en la capa mucilaginosa superficial contribuyendo a sus propiedades físicas. Las mucinas transmembrana funcionan como sensores del ambiente externo, a través de la unión de ligandos a su dominio extracelular o como consecuencia de la alteración de su conformación provocada por los cambios en las condiciones bioquímicas externas, como el pH, la composición iónica o interacciones físicas. Las señales se transmiten al interior de la célula mediante modificaciones postraduccionales del dominio citoplasmático (264).

En el tracto gastrointestinal estas mucinas son principalmente MUC1, MUC3A/B, MUC4, MUC12, MUC13, MUC15, MUC16, MUC17, MUC20 y MUC21, aunque en condiciones no patológicas en el estómago sólo se expresa MUC1. Desde el extremo N-terminal al C-terminal, la estructura general de las mucinas asociadas a membrana presenta tres regiones principales: un dominio extracelular, que contiene la región central de repeticiones en tándem flanqueada por secuencias únicas, un dominio transmembrana de paso único y el dominio citoplasmático (FIGURA I-4), que se asocia con proteínas del citoesqueleto celular y participa la transducción de señales externas al interior de la célula.

El dominio extracelular de estas mucinas, altamente O-glicosilado, forma largas estructuras filamentosas que en el caso de MUC1 alcanzan los 200-500 nm, sobresaliendo por encima de la mayoría de moléculas asociadas a la membrana celular (265). La región central de MUC1 contiene entre 20 y 100 repeticiones en tándem de la secuencia de 20 aminoácidos GSTAPPAHGVTSAPDTRPAP y su núcleo proteico extracelular varía entre 100-300 kDa, aunque la proteína madura, después de la glicosilación, puede doblarse en tamaño (266). Las mucinas asociadas a membrana pueden presentar dos tipos principales de dominios funcionales en la porción extracelular, adyacentes a la región de repeticiones en tándem: el dominio SEA (*sea urchin sperm protein, enterokinase, and agrin*) y/o el dominio tipo EGF.

Las mucinas de superficie celular son cortadas durante su biosíntesis y transportadas a la membrana celular como dos subunidades unidas por enlaces no covalentes en el dominio SEA, permitiendo la subsiguiente liberación del dominio extracelular glicosilado, proceso conocido como *shedding*. MUC4 es la única de las mucinas asociadas a membrana ya caracterizadas completamente que no posee el dominio SEA. En el caso de MUC1, después de su traducción, un único polipéptido precursor es escindido en el retículo endoplasmático en dos fragmentos, mediante autoproteólisis del dominio SEA (FIGURA I-4), en el sitio FRPG/SVVV, tal y como sugieren los estudios estructurales (267,268).

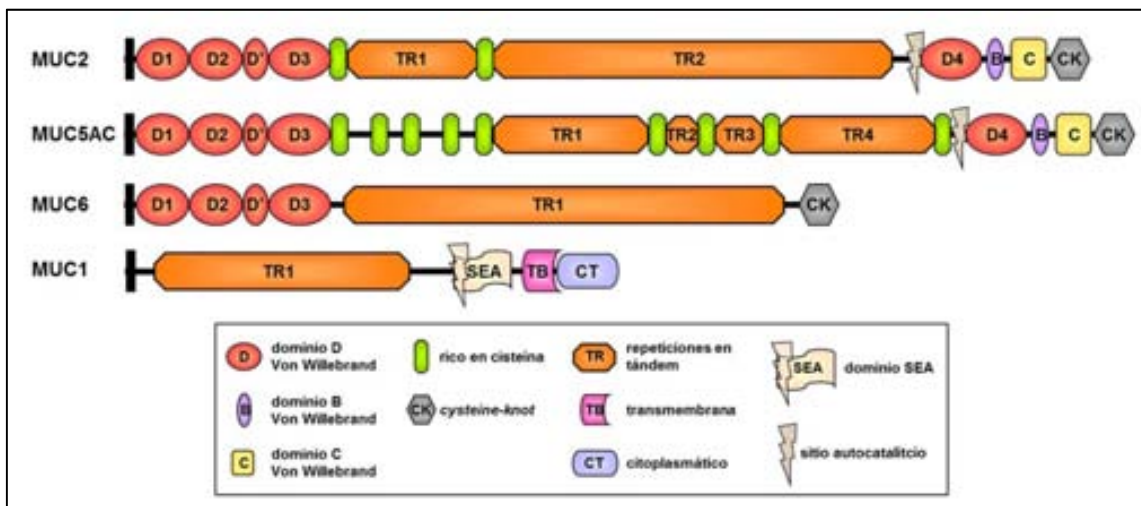


Figura I-4. Esquema de la estructura proteica de mucinas

La unidad más larga se compone de la secuencia señal N-terminal y del dominio de repeticiones, y la unidad C-terminal más pequeña contiene 58 residuos del dominio extracelular y los dominios transmembrana y citoplasmático. Estas dos unidades unidas no covalentemente pueden disociarse en respuesta a alteraciones físicas del medio extracelular, liberando en el gel mucoso una forma soluble de MUC1 que contiene la región de repeticiones glicosilada. La disociación de las dos subunidades de MUC1 también puede ser promovida por la adhesión de *H.pylori* a o por proteasas relacionadas con la familia de las metaloproteasas (269,270). Además, mediante procesamiento alternativo del gen *MUC1*, también se produce una isoforma secretada (MUC1/SEC) que carece de los dominios transmembrana y citoplasmático (271). Existen otras isoformas de MUC1 (MUC1/Y {Zrihan-Licht, 1994 #1915}, MUC1/X (272), MUC1/ZD (273)) originadas del procesamiento alternativo del gen, que por lo contrario carecen de la región de repeticiones en tándem y, por tanto, los dominios extracelulares de estas isoformas son mucho menos complejos.

La región extracelular adyacente al dominio transmembrana de la mayoría de mucinas asociadas a membrana presenta homología con la familia EGF. Se postula que su función es permitir a las mucinas interactuar con miembros de la familia del receptor de EGF, participando en la regulación y señalización relacionada con el crecimiento, motilidad, diferenciación e inflamación. Tanto el número como la disposición de este dominio están conservados entre las mucinas de membrana. Generalmente tienen 2 o 3 dominios tipo EGF, separados por un dominio SEA. Sin embargo, MUC1 no presenta dominios tipo EGF claramente definidos en su dominio extracelular (FIGURA I-4), aunque puede unirse a miembros de la familia ERBB a través de su dominio citoplasmático (274).

El dominio citoplasmático de las mucinas asociadas a membrana, que corresponde a la porción carboxi-terminal, es corto, puede contener sitios potenciales de fosforilación Ser, Thr o Tyr y está altamente conservado entre especies. La activación de los sitios Ser, Thr o Tyr por parte de quinasas puede activar múltiples vías de señalización mediante la interacción con otras

moléculas. (264,275). En MUC1 este dominio consta de 72 aminoácidos y contiene tres aminoácidos adyacentes al dominio transmembrana, Cys-Gln-Cys, que desempeñan un papel importante en dirigir a MUC1 a la membrana celular.

3.1.2. Mucinas secretadas

De las mucinas secretadas completamente caracterizadas, MUC2, MUC5B, MUC5AC, MUC6 y MUC19 son formadoras de gel, y las principales constituyentes de la capa mucilaginosa que recubre el epitelio. MUC7, una mucina localizada principalmente en la saliva, no es formadora de gel. Todas ellas se expresan en diferentes localizaciones del tracto gastrointestinal siendo MUC5AC y MUC6 las mucinas secretadas en la mucosa gástrica normal, siguiendo un patrón de expresión específico de tipo celular. MUC5AC es secretada principalmente por las células foveolares en la superficie del epitelio gástrico mientras que MUC6 es secretada por células mucosas del cuello en el cuerpo y glándulas antro-pilóricas. En condiciones patológicas, especialmente en metaplasia intestinal, se expresa *de novo* MUC2, la mucina típicamente intestinal secretada por células caliciformes (262). También se ha reportado la detección de MUC5B en un pequeño porcentaje de muestras de cáncer gástrico (276).

El núcleo proteico de las mucinas formadoras de gel es muy extenso y su estructura general muy compleja. La densa glicosilación de estas mucinas en la región central repetitiva, juntamente con diferentes niveles de sialilación y sulfatación, les proporciona protección contra enzimas proteolíticas y provoca que estas moléculas ocupen un volumen muy grande en solución, factor clave para la formación de la red gelatinosa de mucinas en el mucus (277). MUC5AC consta de cuatro regiones repetitivas (TR1, TR2, TR3 y TR4) de una secuencia de 8 aminoácidos, separadas por dominios ricos en cisteína, que muestran un grado mínimo de variabilidad en el número de repeticiones, a diferencia de otras mucinas (278,279). MUC2 también presenta una estructura compleja de dos dominios de repeticiones en tándem (TR1 y TR2) separados por una región rica en cisteínas. El primer dominio de 380 aminoácidos es invariable, mientras que el segundo se compone de unidades de repetición de 23 aminoácidos variable en número, entre 40 y más de 150, con una media de 100. Por el contrario, la región central de MUC6 se compone de un único dominio de repeticiones en tándem muy extenso, entre 15 y 30 repeticiones, aproximadamente, de una secuencia peptídica de 169 aminoácidos (280) (FIGURA I-4).

El tipo de oligosacáridos que presenta cada mucina está determinado por el patrón de expresión de glicosiltransferasas en cada localización tisular y tipo celular, y es muy variable. La expresión de estas glicosiltransferasas, que puede ser modulada por la inmunidad innata y adaptativa y la variabilidad genética en los genes que las codifican, resulta en una gran diversidad de estructuras de carbohidratos en las mucinas. Un por ejemplo de ello es la heterogeneidad en oligosacáridos del grupo de antígenos de sangre Lewis. En el estómago humano, el antígeno de grupo sanguíneo Lewis B, que une la adhesina BabA de *H.pylori*, se expresa principalmente en MUC5AC en el epitelio superficial, mientras que los antígenos Lewis Y y Lewis X colocalizan con MUC6 (281).

Una de las características de las mucinas secretadas es la presencia, en sus extremos N- y C-terminal, de dominios de homología con los dominios de dimerización D (D1, D2, D', D3 y D4), B, C y los dominios ricos en cisteína CK (*cystein-knot*), del factor von Willebrand (282) (ver FIGURA I-4). En el retículo endoplasmático se produce la N-glicosilación de la proteína y la dimerización mediante enlaces disulfido intermoleculares en el dominio CK de la región C-terminal. Seguidamente se produce la O-glicosilación en el complejo de Golgi y la posterior multimerización N-terminal de los dímeros, a través de los dominios D, en el trans-Golgi o en los gránulos preparados para la secreción, resultando en estructuras supramoleculares, lineales o ramificadas, esenciales para las propiedades físicas del gel mucilaginoso (283). En el dominio D4 C-terminal de MUC2 y MUC5AC se encuentra un sitio GDPH de proteólisis autocatalítica responsable de los cortes proteolíticos que se producen en el procesamiento normal de estas mucinas *in vivo*. La proteólisis de estas proteínas es un proceso aparentemente intracelular y dependiente de pH, y se cree que genera extremos C-terminales reactivos capaces de formar enlaces cruzados entre ellos y con otras moléculas (284,285). Otra característica de algunas mucinas secretadas, incluidas MUC2 y MUC5AC, es la presencia de dominios ricos en cisteína tanto internos como flanqueantes a la región de repeticiones y que muestra una elevada similitud entre mucinas, aunque su número es variable. La función de estos dominios no es clara aunque se ha propuesto que puedan ser potenciales sitios de interacción con otros componentes del mucus o, debido a su posición entre regiones rígidas de la proteína, que puedan ser importantes para la dinámica y/o el tráfico intracelular y empaquetamiento de estas glicoproteínas en los gránulos secretores (286). MUC6, en comparación con MUC2 y MUC5AC, carece tanto de los dominios D4, B y C y del sitio de proteólisis en el extremo C-terminal, que es más corto, como de los dominios ricos en cisteína, sugiriendo que esta mucina puede que sea procesada de una manera diferente (280).

3.2. Estructura de los genes *MUCs*

La clonación de los genes de mucinas y la caracterización de su secuencia y estructura genómica ha supuesto un gran desafío debido, principalmente, a la presencia de largas regiones de secuencias repetitivas, ricas en los nucleótidos G y C

El gen *MUC1* (4.4 kb, GenelD:4582) fue clonado y secuenciado por diversos grupos en 1990 y fue el primer de los genes de mucinas para el cual se determinó la secuencia completa del cDNA (271,287,288). Este gen, que codifica para una mucina asociada a membrana, se localiza en la región cromosómica 1q21 y su estructura es mucho menos compleja que los genes *MUC* que codifican para mucinas secretadas. *MUC1* se compone de 8 exones (FIGURA I-5). La secuencia señal de la proteína está codificada por el exón 1, mayoritariamente, y parte del 2. El exón 2 es el más largo de todos y el que codifica para el VNTR de *MUC1*, entre 25 y 125 repeticiones de 60 pb. Parte del exón 5 y el exón 6 codifican para el dominio transmembrana, de 24 aminoácidos, seguido de los exones finales que corresponderían al dominio citoplasmático. Es de destacar que *MUC1* presenta múltiples isoformas, más de 10 descritas, generadas por

procesamiento alternativo, en algunos casos determinando por la presencia de polimorfismos tipo SNP en la secuencia.

Los genes *MUC2* (~30 kb, GeneID:4583), *MUC5AC* (GeneID:4586) y *MUC6* (~25 kb, GeneID:4588), junto con *MUC5B*, se localizan en un clúster de 400 kb en el extremo telomérico de la región cromosómica 11p15.5, en una zona con diversos puntos (*hotspots*) de recombinación (289). El primer punto de recombinación se encuentra entre el gen *MUC6* y *MUC2* y ha sido suficiente para romper el desequilibrio de ligamiento entre *MUC6* y los otros genes del clúster, de manera que *MUC6* puede ser considerado completamente independiente al resto de mucinas en relación a la asociación alélica. Por otro lado, existe una clara asociación entre diferentes marcadores y regiones de repeticiones de los genes *MUC2* y *MUC5AC* (289,290).

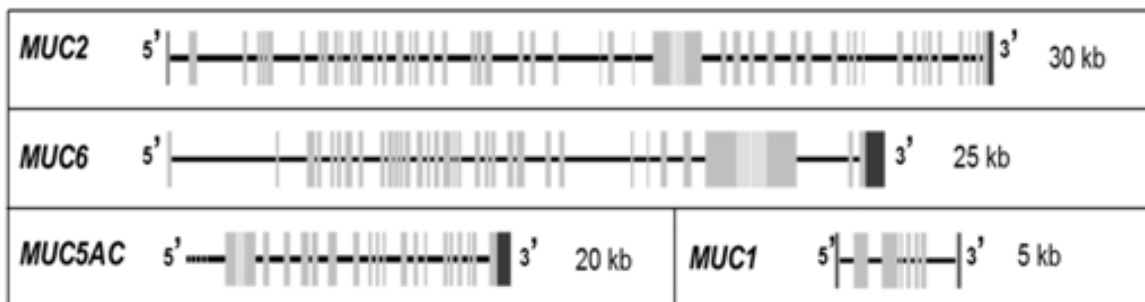


Figura I-5. Esquema de la estructura de los genes de mucina

Estos genes, que codifican para mucinas secretadas, comparten una estructura genómica similar compuesta de un exón central largo, que codifica para el dominio de repeticiones en tándem, flanqueado por las regiones 5' y 3' (FIGURA I-5). La secuencia nucleotídica también es muy similar en las regiones que no codifican para el dominio de repeticiones en tándem y se considera que estos genes comparten un ancestro común y han evolucionado por duplicación (291). La región 5' de estos genes está compuesta por 29 o 30 exones que codifican los diferentes dominios estructurales N-terminales de la proteína, con homología al factor von Willebrand, D1, D2, D' y D3. *MUC2* y *MUC6*, junto con *MUC5B*, muestran una gran conservación en los límites intrón/exón, longitud de exones y posición de los intrones (280). La organización genómica de la región 5' de *MUC5AC* no está totalmente caracterizada, aunque se cree muy similar a la de *MUC5B* y, además, presenta gran identidad de secuencia con las otras mucinas del clúster en esta región (279). Las regiones 3' de *MUC2* y *MUC5AC* también son muy similares ya que constan de 19 y 18 exones, respectivamente, que codifican para los dominios C-terminales de la proteína de homología con el factor von Williebrand D4, B, C y CK. La región 3' de *MUC6* es mucho más corta y se compone de dos únicos exones, que codifican para el único dominio de homología con el factor von Williebrand, CK, en su extremo C-terminal. *MUC6* también difiere de los otros genes del clúster en la orientación que presenta en el cromosoma, adyacente y opuesta a *MUC2*, lo que significa que las secuencias regulatorias de ambos genes se encuentran próximas en la región de 38.2 kb que los separa (278,280).

El dominio de repeticiones en tándem está codificado por un único y largo exón central en cada una de las tres mucinas. La organización de este exón central es diferente en cada caso, pero todas contienen unidades de repetición que son únicas en secuencia y número para cada gen *MUC* y son polimórficas del tipo VNTR (*variable number tandem repeat*). El exón central de *MUC5AC*, de aproximadamente 10.5 kb, muestra una organización muy compleja y codifica para diecisiete dominios importantes. Nueve de estos dominios son ricos en cisteína y su estructura se ha conservado a lo largo de la evolución, sugiriendo que tienen un importante papel en la preservación de las funciones principales de las mucinas (279). El exón central de *MUC2* se compone de cuatro regiones diferentes. La primera región corresponde a un dominio rico en cisteína, seguida de un dominio de repeticiones en tándem de 1140 pb no polimórfico, una tercera región que codifica para un segundo dominio rico en cisteína y, finalmente, la región principal de repeticiones en tándem de 69 pb que muestra un polimorfismo tipo VNTR, con alelos de entre 5 y 15 kb. El exón central de *MUC6* codifica exclusivamente para el dominio de repeticiones en tándem que también es polimórfico. El VNTR de *MUC6* está compuesto por repeticiones de 507 pb y el tamaño del exón central varía entre 10 y 15 kb, aproximadamente (280,292).

3.3. Regulación de la expresión

La expresión de mucinas, basal o inducida, está controlada principalmente a nivel transcripcional, aunque se han observado otros mecanismos adicionales de control post-transcripcional, como el procesamiento alternativo del gen *MUC1*. También existen cada vez más evidencias de la regulación epigenética de la expresión de genes *MUC*.

La caracterización de los promotores de los genes *MUC* ha permitido identificar diversos factores de transcripción implicados en la regulación de estos genes y determinar su interacción con elementos específicos del promotor. No obstante, se necesitan investigaciones adicionales para un mejor entendimiento de la regulación de estos factores de transcripción y establecer los mecanismos de activación en los diferentes tejidos y células productoras de mucinas que expliquen la diversidad en los patrones de expresión de genes *MUC*.

El promotor de *MUC1*, uno de los mejor caracterizados, contiene la típica caja TATA, 25 pb secuencia arriba del inicio de transcripción, así como otros elementos comunes como secuencias ricas en GC (*GC-box*), y las secuencias *enhancers*, E-box y CAAT-box (293). También se han identificado sitios de unión para diversos factores de transcripción como Sp1 (294), importante en la expresión constitutiva específica de tipo celular (295), STAT (296) y NF-κB (297), entre otros. De la misma manera se ha caracterizado la estructura de los promotores de los genes del cluster 11p15 que codifican para las mucinas secretadas *MUC2* (298), *MUC5AC* (299) y *MUC6* (300). Todos estos genes contienen cajas TATA entre 25 y 35 pb secuencia arriba de los sitios de inicio de transcripción y también se han identificado sitios de unión para diversos factores de transcripción, aunque sólo unos pocos han sido confirmados como importantes reguladores de mucinas. *MUC2*, *MUC5AC* y *MUC6* contienen sitios de unión a Sp1, como la secuencia CACCC o GC-boxes. Sp1 es un reconocido factor de transcripción involucrado en la

activación constitutiva o inducida de la expresión de todos los genes del clúster. También se han descrito en todos estos genes sitios de unión para NF- κ B, factor de transcripción activado por diversas citoquinas y productos bacterianos (300-302). *MUC2* está regulado constitutivamente por CDX2, un factor de transcripción que está implicado en la diferenciación y que activa específicamente genes intestinales (298,303).

En el estómago, las células inflamatorias que se infiltran en la mucosa en respuesta a la infección por *H.pylori* pueden modular la presencia de citoquinas específicas (ver apartado 1.5.1.3). Estas citoquinas pueden activar vías de señalización que regulan genes implicados en la reparación y mantenimiento de la mucosa gástrica como pueden ser enzimas de glicosilación, mucinas y *trefoil factors*. Aunque algunos estudios han abordado el efecto de citoquinas en MUCs en distintos órganos y líneas celulares, existen pocos estudios en células gastrointestinales, especialmente en células gástricas. Las primeras evidencias indicaron que TNF- α estimulaba la expresión del mRNA de *MUC1*, *MUC2* y *MUC5AC* en células de cáncer gástrico (304). Posteriormente se demostró que TNF- α estimulaba significativamente la expresión de *MUC1*, también a nivel de proteína, en la superficie celular de ciertas células de carcinoma gástrico (305). *MUC2* se activa también por la citoquina proinflamatoria IL-1 β por TNF- α , a través de la vía de señalización de NF- κ B en células de cáncer gástrico. INF- γ es la citoquina tipo Th1 más predominante que se produce en el estómago infectado por *H.pylori* y desempeña un papel importante en la inmunidad de la mucosa durante la gastritis. Esta citoquina es capaz de inducir genes específicos de células mucosas del cuello, como *MUC6* (306).

Además del efecto en la expresión de genes *MUC*, las citoquinas también pueden regular la actividad de glicosiltransferasas, enzimas que catalizan la unión de oligosacáridos al núcleo proteico de mucinas, y modular la presencia de determinados antígenos carbohidratos (305).

La expresión de *MUC1* también puede ser regulada por factores hormonales. Diversos estudios han demostrado que *MUC1* responde positivamente a la estimulación por hormonas esteroideas en determinados contextos celulares (307,308). El gen *MUC1* contiene secuencias consenso para la unión de los complejos receptores de estrógeno y progesterona (297,309), y se ha demostrado la unión del receptor de estrógeno a su elemento de respuesta en el promotor *in vitro*. Es posible también que el receptor de estrógeno regule la transcripción de *MUC1* indirectamente, a través de la interacción con otros factores de transcripción como STAT, AP-1, o NF- κ B (310).

Otro mecanismo de regulación de la expresión génica son las modificaciones epigenéticas por metilación del DNA y modificación de histonas. Estos mecanismos epigenéticos son usados comúnmente por las células cancerosas para inhibir determinados genes como los supresores de tumores (311). De hecho, los cuatro genes de mucinas localizados en el clúster 11p15.5 se encuentran en una región de metilación anormal en cáncer (312) y sus promotores son ricos en G y C. Las primeras evidencias de control epigenético de estos genes provienen de diversos estudios que mostraron que la represión de la expresión de *MUC2* en células de carcinoma de colon no mucinoso estaba asociada a la metilación de su promotor proximal, tanto *in vitro* como *in vivo* (313,314). Algunos estudios también han demostrado que la expresión *de novo* de *MUC2*

en células de cáncer gástrico mucinoso está inducida por la demetilación o hipometilación, respectivamente, del promotor (315). Otro estudio, que abarcaba los cuatro genes en 11p15 (316), indica que *MUC2* y *MUC5B* están altamente sometidos a metilación del DNA y modificación de histonas, en diversos tipos celulares incluidas células de cáncer gástrico, mientras que *MUC5AC* raramente está influenciado por regulación epigenética y *MUC6* no está controlado por estos mecanismos. Finalmente, un estudio realizado en diversas líneas celulares de cáncer, aunque ninguna de ellas de cáncer gástrico, ha descrito también la regulación de *MUC1* por una combinación de mecanismos epigenéticos, metilación del DNA y modificación de la lisina 9 de la histona H3 (317). Aunque la mayoría de los estudios se han realizado en líneas celulares, recientemente se ha confirmado que el promotor de *MUC2* está también altamente metilado en tejidos de cáncer colorrectal no mucinoso y este estatus de metilación correlaciona de manera general con el nivel de expresión del gen (318).

Además de todos los mecanismos de regulación anteriormente descritos, existen cada vez más evidencias de la implicación de pequeños RNAs no codificantes, como microRNAs, en la regulación post-transcripcional de genes. Estos microRNAs, que regulan negativamente sus genes diana, están implicados en procesos relacionados con la carcinogénesis y podrían actuar como supresores de tumores u oncogenes (319). Recientemente, diversos grupos han descrito la interacción directa de los microRNAs miR-1226 (320), miR-125b (321) y miR-145 (322) con la región 3'UTR de *MUC1* provocando la represión del gen en líneas celulares de cáncer de mama, principalmente. La regulación negativa de *MUC1* por parte de estos miRNAs se ha relacionado con procesos carcinogénicos, como muerte celular, invasión y metástasis, en estos estudios.

3.4. Funciones de las mucinas

Las mucinas forman parte de la capa de gel viscosa y continua, o mucus, que recubre el epitelio formando una barrera altamente hidratada y lubricada que también sirve de matriz para otros componentes. El volumen y composición de esta capa gelatinosa está determinada por los tipos de mucinas que se expresan en el epitelio y la naturaleza de sus modificaciones post-traduccionales, el grado de uniones intra e intermoleculares y el tipo de moléculas que captura la matriz (274). La capa de mucus protege al epitelio subyacente contra agresiones químicas, enzimáticas, microbianas o mecánicas, y actúa como barrera física para inhibir y atrapar patógenos invasores y ayudar a su eliminación (323). Como se ha comentado anteriormente, el mucus que recubre el epitelio puede separarse en dos capas, la capa externa más laxa, que se renueva constantemente, y la capa interna de mucus firmemente adherente. El grosor relativo de las dos capas componentes del revestimiento mucoso varía a lo largo del tracto gastrointestinal. En el estómago, el mucus firmemente adherente y la capa laxa más superficial presentan un grosor similar, de 80 y 154 μm en el cuerpo y antro, respectivamente (324). Por debajo de la capa de mucus, la segunda línea de defensa está constituida por el glicocáliz epitelial, que está parcialmente integrado en el gel viscoso superficial. El glicocáliz representa una matriz de glicoproteínas y glicolípidos muy diversos expresados en la membrana celular, entre los que se incluyen las mucinas. De la misma manera que el gel mucoso, el glicocáliz se renueva constantemente pudiéndose ajustar rápidamente a los cambios en el ambiente (265).

La capa externa de mucinas secretadas proporciona una **barrera física** contra microorganismos y material insoluble y configura y **mantiene el ambiente molecular** local con respecto a la hidratación, composición y concentración iónica y accesibilidad de macromoléculas. La matriz de mucinas interviene en la discriminación molecular regulando el flujo de moléculas específicas. Los dominios de repeticiones en tándem permiten la creación de elevadas concentraciones locales de estructuras de oligosacáridos neutras o cargadas que podrían actuar como un gel de intercambio de iones o como una resina de filtración para facilitar o inhibir la difusión molecular (274). Un ejemplo de este tipo de función es el que se observa en el mucus gástrico, formado principalmente por MUC5AC y MUC6 (325), que actúa como una barrera selectiva de difusión para el HCl. Los iones bicarbonato que son secretados por el epitelio gástrico se retienen en la capa mucosa, posiblemente mediante interacciones iónicas o de otro tipo con las cadenas de oligosacáridos de mucinas, contrarrestando los protones que se difunden desde el lumen y creando un gradiente de pH (desde 2 en el lumen del estómago hasta 6-7 en la superficie celular epitelial). El HCl interacciona con la capa mucosa gástrica de diversas maneras dependiendo del pH. Por encima de un pH de 4, la inyección de HCl causa la formación de canales en la capa mucosa a través de los cuales el HCl pasa, mientras que por debajo de ese pH, el HCl no penetra la capa mucosa (326). Esta función puede ser alterada por la presencia de otras proteínas y componentes en el mucus que modifiquen sus propiedades bioquímicas, como los TFFs (208,232).

Las mucinas secretadas con propiedades de formación de gel proporcionan la **primera línea de defensa contra la mayoría de las infecciones**. Este tipo de mucinas forman una densa malla que permite atrapar a microorganismos patógenos y proporciona una matriz para retener moléculas antimicrobianas. En el estómago, *H.pylori* se une a la mucina MUC5AC a través de la adhesina bacteriana BabA que reconoce al antígeno carbohidrato Lewis B en esta mucina (327). Aunque por una parte esta unión asistiría en la colonización de la capa mucosa por parte de la bacteria, por el otro lado limitaría la interacción de *H.pylori* con la superficie epitelial bloqueando las estructuras de reconocimiento de BabA.

Existen cada vez más evidencias de la implicación de las mucinas asociadas a membrana, que forman parte del glicocáliz, en limitar la patología que resulta de las infecciones bacterianas del tracto gastrointestinal (328). MUC1 proporciona una barrera protectora contra la colonización tanto aguda como crónica del patógeno *H.pylori* y desempeña un importante papel en atenuar la respuesta inflamatoria epitelial inducida por la infección de la bacteria, tal como se ha demostrado en los estudios con ratones deficientes en MUC1 (329,330). MUC1 puede limitar la colonización de *H.pylori* de diversas maneras. El dominio extracelular filamentoso, de una longitud de 200-500 nm, es capaz de distanciar físicamente la bacteria de la superficie epitelial y, de esa manera, inhibir estéricamente la adhesión de la bacteria a otros ligandos potenciales de la superficie celular. La adhesión de *H.pylori* a MUC1, mediante la unión de las adhesinas SabA y BabA a los carbohidratos de la mucina, también puede provocar el *shedding* del dominio extracelular. En este caso, la forma soluble liberada de MUC1 actuaría como señuelo evitando la adherencia prolongada de la bacteria en la superficie celular (270).

Los complejos geles de mucinas son capaces de **capturar y retener moléculas biológicamente activas**. Estas moléculas pueden actuar como indicadores de una alteración física o molecular en la capa de mucinas y, al ser liberadas, podrían iniciar procesos inflamatorios o de reparación. Uno de los ejemplos mejor caracterizados es el que incluye los TFFs, una familia de péptidos que se expresa en la mayoría de epitelios productores de mucinas, incluido el gastrointestinal, y que contribuyen en el mantenimiento y reparación de la mucosa gástrica (ver APARTADO 2) (173,233,331). Estos péptidos muestran una expresión coordinada y colocalizan con las mucinas en el epitelio (177,332).

El mucus también proporciona una **matriz donde se retienen moléculas antimicrobianas** y otros productos producidos por el tracto gastrointestinal relacionados con la inmunidad. La mayoría de los ejemplos se encuentran en el intestino como, por ejemplo, el componente secretor de la inmunoglobulina A (333) o la proteína de unión a la porción Fc de la inmunoglobulina G (IgGFcγBP) (334), que además se une al dominio C-terminal de MUC2 en el mucus intestinal de ratas (335). Además, las mucinas pueden tener propiedades antimicrobianas que limiten el crecimiento de microorganismos en el mucus. Un monosacárido específico de MUC6 inhibe la síntesis de componentes de la pared celular de *H.pylori* y se ha sugerido que podría limitar el crecimiento de esta bacteria en el mucus gástrico (336). También se ha planteado que las mucinas pueden unir citoquinas, factores de crecimiento y de diferenciación y mediadores de la inflamación, aunque no se han demostrado las interacciones moleculares que expliquen la presencia de de estos factores en el mucus (274).

Las mucinas asociadas a membrana también pueden estar involucradas en la **transducción de señales**. Se ha postulado que estas mucinas sirven como sensores del ambiente externo a través de la unión de ligandos en sus dominios extracelulares o como consecuencia de la alteración de la conformación causada por cambios en las condiciones bioquímicas externas como el pH o la composición iónica. Las señales se transmiten al interior de la célula a través de las modificaciones post-traduccionales del dominio citoplasmático que incluyen procesos de fosforilación y proteolíticos, entre otros (264). MUC1 es una de las mucinas mejor caracterizadas en relación a la transducción de señales. Diversos estudios identificaron la molécula de adhesión intercelular 1 (ICAM-1) como uno de los ligandos directos de MUC1 (337). La unión de este ligando al dominio extracelular de MUC1 inicia la señalización de calcio, previamente implicada en remodelación del citoesqueleto y motilidad (338). Las señales se transmiten al núcleo por interacción del dominio citoplasmático con agentes implicados en la señalización celular como β-catenina, p53 o el receptor del estrógeno α (339-341). MUC1 también está implicado en la activación de la vía de las MAPK, mediante fosforilación de su dominio citoplasmático e interacciones con los receptores ERBB (264). Además, la señalización de MUC1, activada por la interacción del dominio extracelular con bacterias y/o por liberación de este dominio, puede modular la activación de la familia de factores de transcripción NF-κB que son importantes reguladores de la inflamación (330,342,343).

3.5. Papel y expresión de mucinas en la carcinogénesis gástrica

Clásicamente, las mucinas son conocidas por su papel central en la lubricación, protección y mantenimiento de la mucosa que recubre las superficies epiteliales. No obstante, existen también estudios que sugieren su implicación en otros procesos biológicos más complejos como diferenciación, señalización celular o adhesión celular, como se ha comentado anteriormente. Debido a las propiedades multifuncionales de las mucinas, una alteración crónica o permanente en su producción o en sus características bioquímicas puede tener un profundo impacto en el comportamiento celular. De hecho, diversas líneas de evidencia han implicado las mucinas en la patogénesis del cáncer. Específicamente, la mayoría de neoplasmas epiteliales muestran frecuentemente una desregulación de la expresión y/o una glicosilación aberrante de determinadas mucinas. La expresión *de novo* o incremento de la expresión de mucinas asociadas a membrana por las células tumorales se ha asociado generalmente con un incremento del potencial maligno y peor pronóstico para los pacientes. En contraste, la expresión de las mucinas secretadas en ciertos tipos de tumores parece correlacionar con un mejor pronóstico (262,274).

3.5.1. Mucinas en lesiones precursoras

3.5.1.1 Gastritis inducida por *H.pylori*

H.pylori daña la capa mucosa y las células superficiales del epitelio gástrico durante la infección. Las proteasas de *H.pylori* pueden desintegrar la estructura polimérica de las mucinas mientras que las lipasas y la fosfolipasa A2 producidas por la bacteria pueden degradar los lípidos del mucus provocando la pérdida de la hidrofobicidad de la superficie mucosa. La infección produce una alteración de la glicosilación de las mucinas, hecho que puede facilitar la unión de la bacteria. La ureasa producida por la bacteria genera amoníaco, que es necesario para establecer el pH neutro del entorno local y también es capaz de desestabilizar la capa mucosa. Todos estos procesos causan el colapso de la barrera mucosa y favorecen la unión de la bacteria al mucus facilitando la supervivencia de *H.pylori* en la mucosa gástrica (277).

Durante la gastritis crónica causada por la infección por *H.pylori*, se produce una alteración del patrón de expresión de las mucinas secretadas típicamente gástricas, MUC5AC y MUC6. Mientras que la expresión de MUC6 en la superficie mucosa tiende a incrementarse, la de MUC5AC disminuye. En la gastritis atrófica también se observa una expresión alterada de MUC5AC y MUC6 en células columnares superficiales. Si se elimina la infección, el patrón gástrico de expresión normal se recupera (344-346). *H.pylori* está estrechamente asociado con la mucina secretada MUC5AC y con las células que la producen, indicando que MUC5AC desempeña un papel importante en la adhesión de *H.pylori* a la mucosa gástrica (327,347).

3.5.1.2 Metaplasia intestinal

Como se ha comentado anteriormente (ver APARTADO 1.4.3), el tipo de mucinas (sialomucinas, sulfomucinas o neutrales) detectadas en metaplasia intestinal permite la identificación de los tres subtipos principales de esta lesión ya que cada subtipo presenta un

patrón específico de expresión de estas glicoproteínas. La metaplasia intestinal tipo I o completa muestra poca o ausencia de expresión de las mucinas típicamente gástricas MUC1, MUC5AC y MUC6. En este tipo de metaplasia se hace evidente la expresión de novo de la mucina intestinal MUC2 en las células caliciformes. Los tipos II y III, correspondientes a la metaplasia intestinal incompleta, comparten un patrón idéntico de expresión de mucinas que se caracteriza por la recuperación de la expresión de las mucinas típicamente gástricas. MUC1 y MUC5AC se expresan tanto en células caliciformes como columnares y la expresión de MUC6 tiene lugar en el epitelio glandular pero no en la misma medida que en la mucosa gástrica normal. En estos tipos de metaplasia, MUC2 continúa expresándose en las células caliciformes (42,346,348,349).

3.5.2. Mucinas en cáncer gástrico

El epitelio gástrico normal tiene un patrón de expresión de mucinas específico de célula y órgano que se caracteriza por altos niveles de MUC1, MUC5AC y MUC6. El proceso de transformación neoplásica del estómago se asocia con una disminución de estas mucinas y la expresión de novo de MUC2. En las fases avanzadas de carcinoma gástrico y en los carcinomas poco diferenciados se expresan múltiples mucinas comparado con las fases tempranas del cáncer y los carcinomas bien diferenciados. El entorno del órgano, agentes etiológicos y la interacción del tumor con el huésped influyen la producción de mucinas y la morfología del carcinoma gástrico aunque todavía no se conocen los mecanismos moleculares que contribuyen a la expresión alterada de estas glicoproteínas en cáncer gástrico.

Diversos estudios han confirmado que la alteración del patrón de expresión de las mucinas puede contribuir a la alteración del crecimiento de las células epiteliales, la respuesta inmune, la adhesión celular y la interacción con la matriz extracelular, influyendo posiblemente en la tumorigenicidad y propiedades metastáticas de las células cancerosas (274,350). El patrón de expresión de mucinas en el carcinoma gástrico es heterogéneo, sobretudo en las fases más avanzadas del cáncer. El proceso de transformación neoplásica en el estómago está asociado con una reducción de la expresión de las mucinas que se expresan normalmente en la mucosa gástrica, incluyendo a MUC1, MUC5AC, MUC6, y la expresión *de novo* de MUC2 y, en menor medida, MUC3 y MUC4. El cáncer gástrico de tipo intestinal se cree que desarrolla un fenotipo predominantemente intestinal, positivo para MUC2, porque está precedido por una fase precancerosa caracterizada por una secuencia de lesiones específicas: gastritis atrófica, metaplasia intestinal y displasia. El tipo difuso muestra un fenotipo gástrico, positivo para MUC5AC y MUC6 y, en general, carece de la expresión de MUC2 (277).

La expresión de mucinas correlaciona con la localización del tumor gástrico. MUC5AC se asocia con carcinoma antral mientras que MUC2 se asocia con carcinomas del cardias (351). También se han relacionado los patrones de expresión de MUC1 y MUC2 con las características clínico-patológicas. El tamaño del tumor, la profundidad de la invasión, la presencia de metástasis en los ganglios linfáticos y la fase clínica correlacionan positivamente con la expresión de MUC1. MUC1 cuando se expresa en tumores puede actuar inhibiendo la adhesión célula-célula y, por tanto, promoviendo la metástasis del tumor. La detección

immunohistoquímica de MUC1 en los tumores gástricos correlaciona con el desarrollo del cáncer, progresión, invasividad celular y mal pronóstico. Por el contrario, la expresión del antígeno MUC2 en carcinomas gástricos es un factor pronóstico asociado a una evolución clínica favorable en los pacientes (352-354).

Las células cancerosas podrían utilizar las propiedades protectoras que las mucinas ejercen en la mucosa gástrica normal para controlar el entorno molecular durante el crecimiento invasivo y metastático, permitiéndolas sobrevivir en condiciones adversas (274). Por ejemplo, los dominios extracelulares de las mucinas asociadas a membrana podrían unir proteínas o receptores de la misma célula bloqueando el acceso de estos receptores. Estas interacciones autocrinas ocurren, por ejemplo, entre el dominio de repeticiones en tándem de MUC1 y ICAM-1. La unión de esta molécula puede activar una señalización promigratoria que facilitaría la metástasis (338).

Existen cada vez más evidencias que las mucinas asociadas a membrana pueden contribuir a la regulación de la diferenciación y proliferación de células tumorales (355). Un aumento de expresión de estas mucinas puede promover el fenotipo de transición epitelio-mesenquimal a través de la pérdida de la polaridad y de las interacciones célula-célula. Un incremento de la expresión de MUC1 en células cancerosas puede provocar un secuestro de β -catenina, alterando la función de la E-cadherina en las uniones adherentes. Además la interacción con MUC1 estabiliza la β -catenina y promueve la activación de genes diana Wnt mediada por β -catenina. Por tanto, MUC1 puede alterar las uniones adherentes y activar la vía de señalización de Wnt, ambos procesos relacionados con la proliferación y la tumorigénesis (350).

MUC2 es esencial para la protección del epitelio intestinal contra bacterias comensales y otros factores ambientales en el lumen que puedan promover la respuesta inflamatoria (328). La pérdida de expresión en ratones *Muc2*^{-/-} está asociada con el incremento de la proliferación y la supervivencia de las células epiteliales intestinales en respuesta a una exposición creciente a contenidos luminales y a la inducción de la inflamación. Los ratones deficientes en MUC2 desarrollan adenomas en el intestino delgado y grueso que progresan a adenocarcinomas invasivos (356). El papel de MUC2 como supresor de tumores puede parecer paradójico al hecho de que MUC2 es aumentada de expresión en determinadas patologías, incluyendo los carcinomas del tracto gastrointestinal. La expresión de MUC2 en estos carcinomas, como en cáncer gástrico, podría reflejar el origen de estos tumores a partir de células que normalmente expresan MUC2, más que el papel de esta mucina en el proceso maligno. Es posible también que el incremento de expresión de esta mucina en determinados tumores, a través de la generación de una barrera mucosa, proteja contra el reconocimiento por efectores antitumorales y, por tanto, contribuir al fenotipo maligno (350). En este contexto MUC2 podría tener una función de supresor de tumores en inflamación y contribuir a la oncogénesis en otras condiciones. De hecho, la hipersecreción de mucus también se ha asociado a una mayor proliferación bacteriana y a la inducción de respuestas inflamatorias que podría promover el desarrollo tumoral.

3.5.3. Variabilidad genética de mucinas y carcinogénesis gástrica

Los primeros estudios de asociación reportados se centraron en los polimorfismos tipo VNTR de los genes *MUCs*. Los VNTRs de mucinas son un dominio común en toda la familia, muy variable en el número de repeticiones y, por tanto, en el tamaño de estas. La región de VNTRs es esencial para la funcionalidad de las proteínas, ya que en ella se unen los oligosacáridos que formarán la malla densa de protección que recubre el epitelio gástrico. Los alelos cortos para el VNTR de *MUC1* se asociaron con un mayor riesgo de cáncer gástrico, gastritis atrófica y metaplasia intestinal especialmente en población portuguesa, con una de las mayores incidencias de cáncer gástrico en Europa (357-359). La longitud de la región VNTR de *MUC1* también se ha asociado específicamente a la gastritis causada por la infección por *H.pylori* (360), mostrando un mayor riesgo los alelos más cortos. Otro estudio obtuvo resultados similares para los alelos cortos de *MUC6* en relación con el riesgo de cáncer gástrico (361) y, posteriormente, se asociaron a la infección por *H.pylori* (362). No obstante, estos estudios no se han replicado en otras poblaciones debido a la dificultad de análisis de los VNTRs de mucinas que se localizan en una región genéticamente compleja y requieren la aplicación de técnicas como el Southern Blot, poco adecuadas para estudios epidemiológicos con un tamaño de muestra grande. Más recientemente se han reportado otros estudios de asociación, en este caso con polimorfismos tipo SNP. El genotipo homocigoto para la variante A de un SNP sinónimo muy frecuente de *MUC1*, rs4072037 (A/G), se ha asociado significativamente con mayor riesgo de cáncer gástrico en diversos estudios en población asiática (ver APARTADO 1.5.3.2). Este polimorfismo, localizado en el exón 2 de *MUC1*, controla el procesamiento alternativo del gen dando lugar a dos isoformas diferentes, las más importantes expresadas en el epitelio gástrico. También se ha sugerido que este polimorfismo modifica la actividad transcripcional del promotor (166,167,363). En población caucásica se ha confirmado la asociación de este polimorfismo de *MUC1* en dos estudios (168,364), en uno de los cuales que también se encontró una asociación significativa entre el alelo menos frecuente de un SNP de *MUC5AC*, rs868903 y mayor riesgo de cáncer gástrico (364). Sólo existe un único estudio reciente que aborde la posible relación entre SNPs de genes *MUC* y lesiones precursoras. En el estudio de Frank *et al.* se analizó la asociación entre polimorfismos de *MUC1*, *MUC2*, *MUC5B* y *MUC6* y el riesgo de gastritis crónica atrófica en población caucásica sin obtener resultados significativos (365).

Objetivos

Con la finalidad de analizar el papel de los genes de protección de la mucosa en la carcinogénesis gástrica, los objetivos propuestos para esta tesis fueron:

- 1) Identificar y caracterizar la variabilidad genética de los genes *TFF* y *MUC*

- 2) Analizar el papel de los genes *TFF* y *MUC* en la carcinogénesis gástrica mediante estudios de asociación entre polimorfismos en estos genes y la evolución de lesiones gástricas y el cáncer gástrico.

Materiales y Métodos

1. PACIENTES Y CONTROLES

Las muestras de DNA que se han utilizado para la puesta a punto de las técnicas, para validar los SNPs seleccionados en las bases de datos, para realizar la búsqueda de nuevos SNPs son muestras anonimizadas de individuos de la población general disponibles en nuestro laboratorio.

Para los estudios de asociación entre polimorfismos y evolución de lesiones gástricas preneoplásicas y cáncer gástrico se dispuso del DNA y la información de los dos grupos de pacientes y controles descritos en este apartado. En todos los casos los procedimientos fueron aprobados por los comités éticos de los diferentes centros de donde se obtuvieron las muestras y se obtuvo por escrito el consentimiento informado de todos los individuos que participaron en el estudio.

1.1. Muestras procedentes de un estudio español de seguimiento de pacientes con lesiones gástricas preneoplásicas

Para el estudio de asociación de polimorfismos en genes de protección de la mucosa gástrica y genes involucrados en la respuesta a la infección por *H.pylori* con la evolución de lesiones preneoplásicas gástricas, se dispuso de un grupo de 478 muestras procedentes del Hospital general de Soria, área de alto riesgo de cáncer gástrico en España, seleccionadas para un estudio observacional longitudinal con tres puntos finales: cáncer gástrico, úlcera duodenal y enfermedad de Barret (45) (FIGURA M-1).

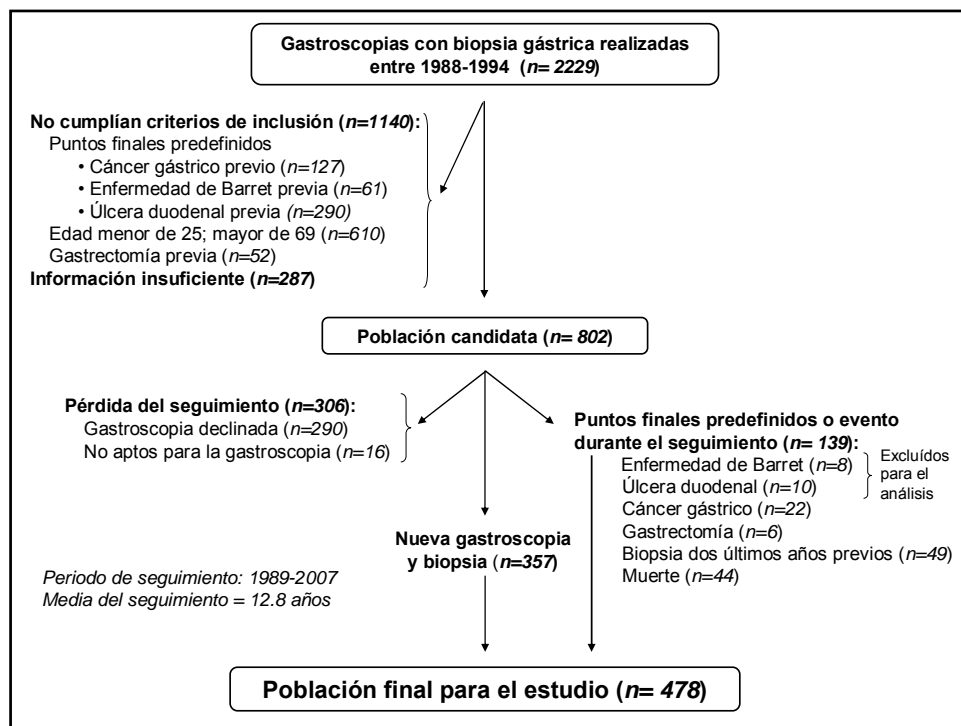


Figura M-1. Diagrama de flujo de selección de pacientes participantes en el estudio observacional longitudinal de lesiones preneoplásicas gástricas en la población de Soria.

Los 478 individuos se seleccionaron de un total de 2229 pacientes a los que, en el período comprendido entre 1988 y 1994, se les practicó una gastroscopia con biopsia. De todos los pacientes identificados inicialmente para el estudio, se excluyeron 1,751 debido a que no cumplían los criterios de inclusión o otras causas que se especifican en la FIGURA M-1. Para realizar la endoscopia y biopsia gástrica de final de seguimiento se siguieron las recomendaciones de Sydney (34), y se obtuvieron 5 muestras (una del ángulo, dos del antro y dos del cuerpo del estómago).

La información disponible de estas muestras, además de edad y sexo, fue: estudio histológico para diagnóstico anatomopatológico, detección de la presencia e infección por *H.pylori* al inicio y final del seguimiento, consumo de tabaco, consumo de AINES y historial familiar de cáncer gástrico. La infección por *H.pylori* se detectó mediante un examen histopatológico de las biopsias con tinción Giemsa. Los individuos fueron entrevistados sobre previas infecciones por *H.pylori*, tratamiento y resultados. Se revisaron, además, los resultados del test rápido de la ureasa de los informes médicos. La detección y caracterización de *H.pylori* también se realizó mediante PCR a partir del DNA extraído del tejido gástrico incluido en parafina (366). Un resultado positivo para la infección por *H.pylori* en alguna de estas fuentes de información, fue considerado como el diagnóstico global para cada individuo.

Las lesiones gástricas diagnosticadas, tanto al inicio como al final del seguimiento, fueron clasificadas, según el índice de Correa (15), en: normales, gastritis superficial, gastritis crónica atrófica, metaplasia intestinal, displasia y cáncer gástrico. El índice de Correa (FIGURA M-2) asigna, a cada una de estas etapas histológicamente diferenciadas, un valor numérico como puntuación global que va desde el 1, que corresponde a un diagnóstico normal, a 6, que corresponde al carcinoma.

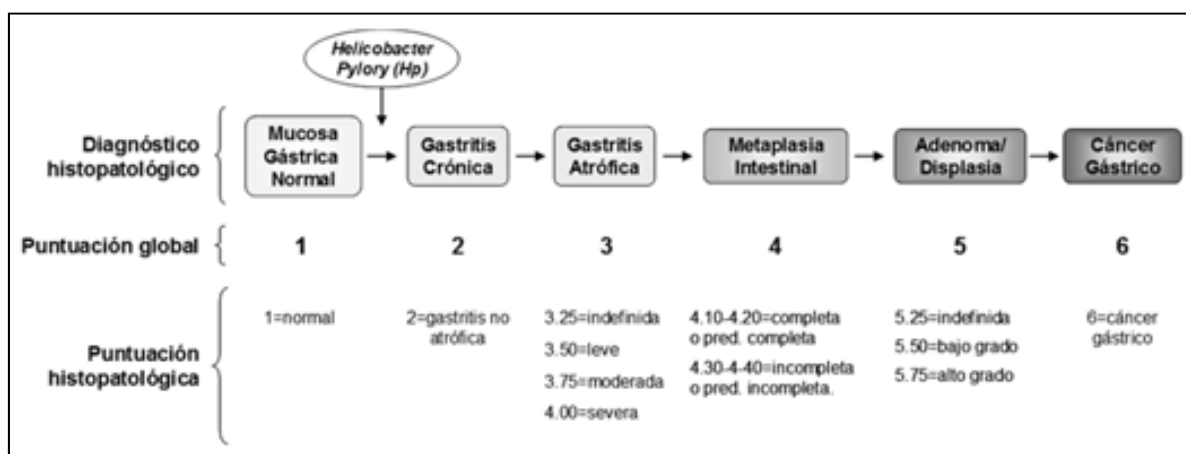


Figura M-2. Índice de Correa para la clasificación de lesiones preneoplásicas gástricas

Además se asigna una puntuación histopatológica en función de la severidad de la lesión. Esta puntuación, generalmente, es más difícil de asignar y su exactitud depende de la categoría de la lesión que se evalúa. A partir del diagnóstico en el momento de reclutamiento y final de seguimiento la evolución de las lesiones se evaluó de acuerdo con los siguientes criterios: se

consideró progresión/regresión un punto de evolución en la puntuación global. En el caso de la metaplasia intestinal se diferenció entre metaplasia intestinal incompleta y completa.

De los 478 individuos finalmente seleccionados para el estudio, no se dispuso de muestra de sangre ni de tejido de tres individuos. De los 475 restantes, se disponía de muestra de sangre periférica para la extracción de DNA de 420 individuos, y de tejido incluido en bloques de parafina para el resto de participantes. En cuanto al tejido incluido en parafina (n=55), 19 bloques no contenían suficiente tejido para realizar la extracción de DNA, por lo que estos individuos tuvieron que excluirse del estudio genético. Además tres muestras adicionales fueron excluidas en el proceso final debido a errores en el control de calidad del genotipado por lo que el número total de individuos para los cuales se pudo realizar el análisis genético fue de 453.

1.2. Muestras procedentes de un estudio caso-control de cáncer gástrico anidado en la cohorte europea EPIC

La cohorte “*European Prospective Study into Cancer and Nutrition (EPIC)*” está constituida por 521.457 participantes, procedentes de 23 centros en 10 países europeos, entre ellos España (367). Los individuos, de ambos sexos y con edades comprendidas entre 35 y 70 años, fueron reclutados entre 1992 y 1998 y seguidos durante una media superior a 10 años. El estudio caso-control de cáncer gástrico anidado en de la cohorte EPIC, comprende todos los casos de adenocarcinoma gástrico identificados en la cohorte durante el seguimiento y con muestras biológicas disponibles para los estudios de genotipado y detección de la infección por *H.pylori*. Por cada nuevo caso incidente, se seleccionaron al azar hasta cuatro individuos controles entre los miembros de la cohorte vivos y no diagnosticados de cáncer en el momento del diagnóstico del caso, apareados por centro, sexo, edad ($\pm 2,5$ años) y fecha de extracción de la sangre (± 45 días). El seguimiento se inició en la fecha de reclutamiento EPIC y acabó en la fecha de diagnóstico de cáncer gástrico, fecha de fallecimiento o fecha final de seguimiento completo (31 de diciembre de 2006). Para los estudios de asociación con variantes genéticas el estudio caso-control incluyó 365 casos (334 de cáncer gástrico y 31 de la unión gastroesofágica) y 1284 controles.

Los casos diagnosticados durante el seguimiento se clasificaron según el código C16 de la 10ª Revisión de la Clasificación Internacional de Enfermedades (*International Classification of Disease, ICD-10*). Un panel de patólogos independientes revisaron los cortes histológicos originales así como los informes proporcionados por cada centro EPIC, con el fin de confirmar y validar el diagnóstico y clasificar los casos según su localización anatómica (cardias y no cardias) y su tipo histológico (cáncer gástrico intestinal o difuso) (368).

De todos los individuos, en el momento del reclutamiento se recogió información sobre la dieta habitual en el año anterior mediante cuestionarios alimentarios previamente validados. Simultáneamente se recogió información mediante otro cuestionario sobre estilo de vida, incluyendo nivel educacional y social familiar, actividad física, historia de consumo de tabaco y alcohol, historia reproductiva y de consumo hormonal, historia de enfermedades previas e

intervenciones quirúrgicas. Así mismo se tomaron medidas antropométricas (peso, altura, circunferencia de cintura y cadera) en condiciones estándares y muestras de sangre (367).

El DNA genómico de los pacientes y controles se extrajo del *buffy coat* (precipitado de linfocitos) preparado de la muestra de sangre y mantenido congelado en nitrógeno líquido hasta su uso. (ver Materiales y Métodos, APARTADO 3.2).

2. BASES DE DATOS Y HERRAMIENTAS BIOINFORMÁTICAS

2.1. Selección de polimorfismos

Para los estudios de asociación entre polimorfismos en genes de protección de la mucosa (*TFF1*, *TFF2*, *TFF3*, *MUC1*, *MUC2* y *MUC6*) y evolución de lesiones preneoplásicas y cáncer gástrico, se seleccionaron de **polimorfismos potencialmente funcionales**, seleccionados de las bases de datos o identificados de nuevo en nuestra población control, y **polimorfismos marcadores (tagSNPs)** de los principales bloques de haplotipos de estos genes. El estudio en los pacientes con lesiones gástricas preneoplásicas se complementó con la selección de polimorfismos en genes involucrados en la respuesta inflamatoria (*IL10*, *IL1A*, *IL1B*, *IL1RN*, *IL8*, *IL4*, *IL12B*, *LTA*, *TNF*, *IL6*, *IL4R*), la reparación del DNA (*XRCC1*, *ERCC2*) y la infección por *H.pylori* (*TLR2*, *CD14*, *LY96*, *PSCA*, *TLR4*, *IFNG*, *IFNGR1*, *PTPN11*, *NOD1*, *NOD2*, *CDH1*, *NFkB*, *NOS2A*). De estos genes se incluyeron polimorfismos para los cuales existieran antecedentes publicados de su relación con la carcinogénesis gástrica y que estuvieran descritos y validados en las bases de datos públicas. También se seleccionaron tagSNPs de las principales regiones de interés de alguno de los genes *ILs* (*IL1A*, *IL1RN*, *IL1B*, *IL4*, *IL4R*, *IL12B*).

2.1.1. Selección de polimorfismos potencialmente funcionales

Principalmente se consultó la **dbSNP** (*National Center for Biotechnology Information*, NCBI; <http://www.ncbi.nlm.nih.gov/projects/SNP>) que es una base de datos de acceso libre donde se catalogan todos los polimorfismos descritos en el genoma humano y ofrece información sobre su localización en las distintas regiones del gen. Todos los polimorfismos seleccionados para su posterior validación se encuentran en esta base de datos. Se utilizaron otras bases de datos y herramientas (Ensembl, SNP500Cancer y PupaSuite) para complementar y comparar la información obtenida en dbSNP. **Ensembl** (<http://www.ensembl.org>) mantiene una anotación automática de los genomas e integra cualquier información biológica relacionada con características derivadas de la secuencia genómica, como es el caso de las variantes polimórficas. **SNP500Cancer** (<http://snp500cancer.nci.nih.gov/home.cfm>) es una base de datos que proporciona información de SNPs en genes relacionados con cáncer e información sobre ensayos de genotipado para los mismos. Finalmente la aplicación web **PupaSuite** (<http://PupaSuite.pupasuite.bioinfo.cipf.es>) permite la selección de polimorfismos con efectos funcionales. Su interfaz interactiva permite utilizar como filtros información funcional, tanto a nivel transcripcional como traduccional, así como frecuencias alélicas y parámetros de desequilibrio de ligamiento. Esta aplicación utiliza el mapado de SNPs en el genoma

proporcionado por Ensembl. Los criterios para escoger polimorfismos potencialmente funcionales de las bases de datos públicas anteriormente descritas, fueron los siguientes:

- SNPs en regiones codificantes de cambio de aminoácido.
- SNPs en regiones promotoras localizados en posibles sitios de unión a factores de transcripción o a otros elementos reguladores.
- SNPs en regiones intrónicas que pudieran afectar al procesamiento alternativo o *splicing*.
- SNPs en regiones transcritas pero no traducidas, regiones 5'UTR o 3'UTR.

Para aquellos polimorfismos que estaban previamente validados en las bases de datos públicas, se tuvo en cuenta que el alelo menos común tuviera una frecuencia superior al 1% (*Minimum Allele Frequency* o MAF > 0.01).

Además, se utilizaron otras herramientas bioinformáticas para el análisis más detallado del posible efecto funcional de algunas de las variantes. Para el análisis de variantes reguladoras que podrían afectar sitios de unión a factores de transcripción se utilizó regSNP (369). Esta herramienta utiliza las bases de datos TRANSFAC y JASPAR para la predicción de los sitios reguladores y proporciona una puntuación para cada uno de los alelos del SNP. Además, se consultó la herramienta is-rSNP (370) que predice, para un SNP determinado, un grupo de factores de transcripción cuya unión podría verse afectada por la variante, asignando un valor p a cada una de las predicciones. Para el análisis variantes que podían afectar el *splicing* se utilizaron los programas de predicción de sitios donadores o aceptores *Human Splicing Finder* (371) y FSPLICE (<http://linux1.softberry.com>). Las variantes de cambio de aminoácido se analizaron, inicialmente, con las herramientas PolyPhen (372) y SIFT (373). PolyPhen-2 asigna una puntuación (de 0 a 1) a la probabilidad de que la variante pueda ser perjudicial y clasifica las variantes como: “seguramente perjudiciales”, “posiblemente perjudiciales” y “benignas”. Además, realiza un alineamiento múltiple de la secuencia proteica entre diferentes especies. SIFT predice el efecto del cambio de aminoácido en la proteína en función del grado de conservación del residuo de interés. Finalmente el programa ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) se utilizó para realizar alineamientos múltiples de las secuencias de la región 3'UTR.

2.1.2. Selección de TagSNPs a partir de los datos de HapMap

Con el fin de obtener la máxima información sobre la variabilidad genética de los genes TFF, *MUC*, con un menor genotipado de SNPs, se ha utilizado la información de la base de datos HapMap (<http://www.hapmap.org>) y el programa haploview (374) para identificar SNPs marcadores (tagSNPs) de los principales bloques de haplotipos. El proyecto internacional HapMap tiene como objetivo mapear los patrones de diversidad genética en el genoma en distintas poblaciones humanas y, de esta manera, proporcionar a la comunidad científica una descripción del desequilibrio de ligamiento (*linkage disequilibrium*, LD) en las distintas regiones y poblaciones, facilitando la optimización del proceso de selección de SNPs para los estudios de asociación. En el momento de la selección de polimorfismos para *TFFs* y *MUCs* el proyecto se

encontraba en su segunda fase (123), por lo que fueron los datos de esta fase (HapMap Data Rel 24 phase II, Nov08) los que se analizaron en nuestro estudio.

Para descargar los datos de genotipado de un determinado gen, se aumentó la región analizada a 50 kb a 5' y 30 kb a 3' del gen y, en función de la información obtenida (presencia de elementos reguladores, que puedan encontrarse en secuencias genómicas a 5' o 3' del gen candidato o presencia de otros genes en el mismo bloque LD), el área finalmente analizada para la selección de tagSNPs se redujo a la considerada relevante para cada gen: todo el cluster de genes *TFF*, y región adyacente, localizado en la región cromosómica 21q22.3; el gen *MUC1* en 1q21, y el cluster de mucinas en la región 11p15, donde se localizan *MUC2*, *MUC6* y *MUC5AC*. Los datos de genotipado disponibles para población de origen europeo (individuos de Utah en EEUU, CEU) se grabaron en un archivo de texto, que fue el que posteriormente se utilizó para el análisis de LD y selección de tagSNPs utilizando el programa **Haploview** (<http://www.broad.mit.edu/mpg/haploview>). Posteriormente, este programa también se utilizó para el análisis de LD e identificación de haplotipos a partir de los datos obtenidos en el genotipado de los polimorfismos seleccionados en este trabajo. La selección de polimorfismos para su análisis en Haploview se realizó según los siguientes criterios:

- Un valor p para el test del equilibrio de Hardy-Weinberg mayor de 0.001.
- Un mínimo del 75% de individuos genotipados para cada polimorfismo.
- Como máximo un error de herencia mendeliana. Este tipo de errores son generalmente debidos a errores de genotipado o asignación errónea del parentesco de los individuos.
- Una MAF igual o mayor a 0.05.

Una vez seleccionados los polimorfismos a incluir en el análisis, Haploview crea una representación gráfica del LD a partir de la cual se definieron los bloques de ligamiento. Se utilizó el método de intervalos de confianza descrito por Gabriel et al. (375). En algunos casos, estos bloques se modificaron manualmente para que abarcaran una región más amplia de la zona de interés. Seguidamente se determinaron los parámetros para seleccionar los polimorfismos (tagSNPs) que identificaran los bloques anteriormente definidos. El método escogido para realizar la selección de tagSNPs fue a pares (*pairwise*) y el coeficiente de correlación, r^2 , de medida de equilibrio de ligamiento, se fijó en ≥ 0.8 . En el caso de tener especial interés en alguno de los polimorfismos de la región analizada, por ejemplo en el caso de SNPs potencialmente funcionales anteriormente seleccionados y que se encontraban en la base de datos de HapMap, se seleccionaron para forzar al programa a incluirlos como tagSNPs, si cumplían con los requisitos. El programa también permite seleccionar la frecuencia mínima de los haplotipos que se van a incluir en el análisis y que, por tanto, influirá en el número de tagSNPs a seleccionar. En todos los casos testamos diferentes frecuencias: haplotipos con una frecuencia mayor del 1, 5 y 10%, lo que nos proporcionó mayor flexibilidad en el momento de la selección de tagSNPs. Para la selección definitiva de tags se tuvo en cuenta la relación entre el número de tags seleccionados y la cobertura de la variabilidad genética en cada bloque. Una vez realizada la selección de tagSNPs, algunos de los tags inicialmente seleccionados tuvieron que

ser sustituidos por otros de las mismas características (preferiblemente en total LD con los inicialmente seleccionados) para poder ser correctamente genotipados.

2.2. Selección de miRNAs

Para seleccionar microRNAs (miRNAs) que tuvieran como posible diana secuencias de la región 3'UTR de genes *TFF* donde se localizara un polimorfismo, se utilizaron diversos programas de predicción de dianas de miRNAs. Los miRNA seleccionados se utilizaron posteriormente para comprobar experimentalmente, mediante ensayos luciferasa, si realmente se daba esta interacción y si la presencia de polimorfismos en la región 3'UTR podía alterarla.

Todas las secuencias de los miRNAs seleccionados para su posterior validación experimental, fueron extraídas de la base de datos **miRBase** (<http://microrna.sanger.ac.uk/>, actualmente <http://www.mirbase.org/>), centrada básicamente en la distribución de la información genómica de miRNAs, recopilando y anotando todas las secuencias publicadas de genes de miRNAs (Griffiths-Jones 2006, 2008). La base de datos de secuencias de miRBase contiene tanto las secuencias de miRNAs maduras como las secuencias de los miRNAs precursores (pre-miRNAs). La base de datos también contiene información relacionada con la estructura i la función de los miRNAs, y proporciona las coordenadas de su posición genómica. La base de datos incluida en miRBase, **miRBase:Targets**, utiliza el algoritmo de miRanda para predecir sitios diana en regiones 3'UTR. El algoritmo miRanda (Enright 2003) es un método predictivo basado en las estructuras heterodúplex formadas entre miRNA y región 3'UTR del mRNA, que asigna pesos a los alineamientos en función de la complementariedad y cálculos termodinámicos. Actualmente la base de datos de dianas de miRBase se ha renombrado como microCosm y está alojada en el *European Bioinformatics Institute* (EBI).

Otros programas de predicción utilizados fueron MicroRNA.org, TargetScan, MicroInspector, PicTar, Diana-Micro T y TarBase, que utilizan distintos parámetros para su predicción en relación al apareamiento de las secuencias, la termodinámica de la interacción y la conservación entre secuencias. **MicroRNA.org** (<http://www.microrna.org>) es un recurso para la predcción de dianas de miRNAs y perfiles de expresión en diferentes tejidos. **TargetScan** (<http://www.targetscan.org/>) predice las dianas de miRNAs buscando en éstas la presencia de sitios de 7 y 8 bases conservados que sean complementarios a la región *seed* de los miRNAs. El programa **MicroInspector** (<http://bioinfo.uni-plovdiv.bg/microinspector/>) genera una lista de posibles sitios diana en la secuencia de interés, ordenándola por valores de energía libre y permite visualización de estructuras secundarias. **Pic Tar** (<http://pictar.mdc-berlin.de/>) Los datos de entrada de PicTar son un alineamiento múltiple de secuencias ortólogas 3'UTR y un grupo predefinido de miRNAs maduros. Seguidamente localiza todos los posibles sitios de unión a miRNAs y los filtra por su energía libre predicha. La aplicación **DIANA-micro-T** (<http://diana.cslab.ece.ntua.gr/>) permite la búsqueda de dianas de miRNAs anotados o de secuencias de miRNAs definidas por el usuario. El algoritmo combina elementos de reconocimiento de miRNAs conservados y no conservados en una puntuación de predicción final.

Finalmente, **TarBase** (<http://diana.cslab.ece.ntua.gr/tarbase/>) es una base de datos que engloba una serie de dianas de miRNAs confirmadas experimentalmente.

Una vez realizadas las predicciones de unión a miRNAs para las regiones 3'UTR de *TFF*, se evaluó la interacción de los miRNAs seleccionados y la región 3'UTR con el programa miRanda v1.0b, desarrollado por el *Computational Center of Memorial Skan-Kettering Cancer Center* y basado en el algoritmo miRanda descrito anteriormente. Esta aplicación nos permitió testar cada uno de los alelos de los polimorfismos de regiones 3'UTR y comprobar si estos cambiaban la energía libre de la interacción miRNA-mRNA.

Los criterios para la selección de miRNAs fueron los siguientes:

- Fueron seleccionados miRNAs en que el sitio de unión predicho en el gen diana contuviese alguno de los polimorfismos identificados en *TFF*.
- Se priorizaron aquellos miRNAs que aparecieran en, como mínimo, dos de los programas consultados.
- Únicamente se escogieron miRNAs para los que el polimorfismo presente en la región 3'UTR tuviera efecto en la interacción según la predicción realizada por el programa miRanda v1.0b.

2.3. Análisis de secuencias

Todas las secuencias de los genes *TFFs* y *MUCs* se obtuvieron de la base de datos de disponibilidad pública **GenBank** (<http://www.ncbi.nlm.nih.gov/genbank/>) del NCBI. Esta base de datos de secuencias genéticas del NIH (*National Institutes of Health*) de Estados Unidos forma parte del *International Nucleotide Sequence Database Collaboration*, donde también se integran otras bases de datos internacionales, como el *DNA DataBank of Japan* (DDBJ) y el *European Molecular Biology Laboratory* (EMBL).

La localización exacta de todos los polimorfismos en la secuencia de referencia de DNA genómico en GenBank se determinó mediante la herramienta bioinformática *Basic Local Alignment Search Tool* del NCBI (**BLAST**, <http://www.ncbi.nlm.nih.gov/blast>), que incorpora una serie de algoritmos que permiten comparar una secuencia de DNA o proteína problema contra todas las secuencias de una base de datos específica.. Esta herramienta también se utilizó para comprobar la idoneidad de los cebadores diseñados para amplificar fragmentos de interés y verificar los resultados obtenidos de la secuenciación directa de dichos fragmentos. El programa presenta una lista con todos los alineamientos ordenados de mayor a menor identidad y asigna a cada uno de ellos una significancia estadística calculada. Principalmente se ha utilizado el programa **blastn**, que es el que compara una secuencia problema de nucleótidos contra la base de datos de DNA genómico y mRNA, y el **BL2seq** que permite comparar dos secuencias problema entre ellas. Las secuencias de referencia de cada gen en estudio fueron editadas con la aplicación **Discovery Studio® (DS) Gene** (Accelrys). La utilización de este programa permitió mapear en cada secuencia, de una manera gráfica, todas las regiones de interés y sus características, como los polimorfismos identificados en la región, así como la localización de los cebadores diseñados para PCR y clonación. Esta aplicación también se utilizó para diseñar los

experimentos de restricción, ya que nos permitió buscar y visualizar los sitios de corte en las secuencias de interés, tanto de fragmentos amplificados por PCR como, en algunos casos, de los plásmidos que se utilizaron para la clonación.

3. EXTRACCIÓN Y CUANTIFICACIÓN DE DNA.

3.1. Extracción de DNA genómico a partir de sangre periférica

En el caso de los pacientes procedentes del estudio de seguimiento de lesiones preneoplásicas para los que se pudo obtener 10 - 12 ml de sangre periférica recogida con ácido etilendiaminotetracético (EDTA) (n=420), la extracción del DNA de los linfocitos se llevó a cabo con el *Flexigene DNA kit de Qiagen* (ver PROTOCOLO I).

El DNA genómico de las muestras procedentes del estudio caso-control de cáncer gástrico fue extraído a partir de alícuotas de 0,5 ml de *buffy coat* (precipitado de linfocitos), conservado en nitrógeno líquido desde el momento de la recogida y procesamiento de la sangre. La mayoría de las muestras fueron extraídas en la IARC (*International Agency for Research on Cancer*), utilizando el *Puregene DNA Purification System* adaptado para la plataforma robotizada de manipulación del DNA *Gentra Autopure LS* (Gentra Systems). Las muestras de DNA se pipetearon en placas de 96 pocillos para su cuantificación mediante el sistema PicoGreen, liofilización y distribución a los centros. Las muestras procedentes de Malmö fueron extraídas por el método fenol-cloroformo, también a partir de pajuelas de 0,5 ml de *buffy coat* y también se distribuyeron liofilizadas en placas de 96 pocillos. Las muestras liofilizadas fueron reconstituidas con agua bidestilada en nuestro laboratorio antes de su uso y se conservaron congeladas a -20°C.

3.2. Extracción de DNA genómico a partir de tejido incluido en parafina

Se extrajo DNA a partir de tejido incluido en 37 bloques de parafina de procedentes del estudio de seguimiento de lesiones preneoplásicas. En primer lugar, con un micrótopo se realizaron 10 cortes de 10 micras de cada bloque y se siguió un proceso, para cada corte, de desparafinación y digestión enzimática (ver PROTOCOLO II). El tejido así tratado se digiere con una mezcla de proteinasa K y un tampón de lisis de tejidos. Seguidamente se procede a la purificación del DNA genómico mediante la utilización del *QIAamp DNA Mini Kit* de Qiagen. Este kit se basa en un sistema de columnas con membranas de sílica, que adsorben el DNA para permitir los lavados con etanol y final elución con agua bidestilada.

3.3. Cuantificación del DNA

La concentración del DNA extraído de muestras de sangre y de tejido incluido en parafina, así como de las preparaciones plasmídicas, se determinó mediante la lectura de la absorbancia a 260nm (A_{260}) en un espectrofotómetro *NanoDrop ND-1000* (Thermo Scientifics). La calidad del DNA se estableció mediante los ratios $A_{260/230}$ y $A_{260/280}$, que indican la pureza de los ácidos nucleicos con respecto a sales, alcoholes y otros reactivos y a proteínas, respectivamente. Algunas muestras de DNA anteriores, así como todas las procedentes del estudio caso-control

europeo, también fueron cuantificadas utilizando *PicoGreen*[®] (Molecular Probes), para precisar mejor su concentración antes de ser genotipadas en plataformas de alto rendimiento (ver Materiales y Métodos, APARTADO 7.4). El sistema *PicoGreen*[®] permite determinar la concentración del DNA mediante el uso de un colorante fluorescente que se une al DNA de doble cadena y su cuantificación en un fluorímetro. Esta técnica es mucha más sensible que la medida de la absorbancia a 260 nm y, además, permite detectar cantidades más pequeñas de DNA.

4. REACCIÓN EN CADENA DE LA DNA POLIMERASA (PCR)

La amplificación de DNA mediante la reacción de la cadena polimerasa (PCR) es actualmente un procedimiento básico en un laboratorio de biología molecular, y sobre el cuál se basan gran número de metodologías. En este trabajo, la PCR se utilizó, principalmente, para amplificar fragmentos correspondientes a las regiones funcionales de los genes *TFFs* y *MUC1*, concretamente los exones y regiones adyacentes, para posteriormente realizar la validación y búsqueda de SNPs mediante RFLP, SSCPs y/o HRM. Mediante esta técnica también se amplificaron los miRNA para su posterior clonaje. Finalmente, la PCR, y posterior visualización en geles de agarosa, sirvió como método de genotipación del polimorfismo tipo VNTR (*Variable Number Tandem Repeat*) de *TFF2*.

El diseño de los cebadores es crítico para obtener un buen resultado en la amplificación de los fragmentos. Seguir los criterios más abajo especificados para el diseño facilita el proceso de optimización de la PCR y la utilización de condiciones estándar en el proceso de amplificación:

- Evitar la autocomplementariedad de la secuencia o secuencias invertidas repetidas en el cebador ya que pueden formar bucles estables impidiendo el anillamiento de los cebadores a su molde.
- Evitar la complementariedad entre los dos oligonucleótidos que van a utilizarse como cebadores, ya que puede causar la formación de dímeros. Este factor es especialmente importante en los extremos 3'.
- Intentar que la última base del oligonucleótido sea una C o G, ya que aumenta la estabilidad de la unión, y las bases A o T pueden favorecer los emparejamientos erróneos.
- La temperatura de anillamiento de los dos oligonucleótidos utilizados en la misma PCR debe ser, aproximadamente, la misma.
- El tamaño de ambos oligonucleótidos ha de ser similar y de una longitud aproximada de 18 a 21 bases.
- El contenido en C y G debe ser entre un 30 y 70%, y la distribución de las diferentes bases en la secuencia ha de ser equilibrada.

Los oligonucleótidos cebadores utilizados en la reacción de PCR fueron diseñados manualmente y comprobados en el programa informático *Oligo6* (Molecular Biology Insights, Inc.) para identificar intra o intercomplementariedad entre cebadores. Se diseñaron, siguiendo los

criterios anteriores, cebadores específicos para la región 5' de *TFF2* y los exones y región adyacente de los genes *TFF* y *MUC1* y cebadores para amplificar la región que incluía los polimorfismos tipo VNTR del intrón 2 de *TFF2* y de *IL1RN*. Estos cebadores se describen en las TABLAS M-1 y M-2 y su localización en los genes *TFF* y *MUC1* en la FIGURA M-3. Los cebadores destinados a la amplificación de exones de *MUC1*, se diseñaron teniendo en cuenta que el tamaño de los fragmentos amplificados debía ser lo más pequeño posible, 100-200 pb, no superando nunca los 250 pb, ya que es uno de los requisitos para la amplificación y genotipado mediante PCR a tiempo real con la metodología *High Resolution Melting* (ver APARTADO 7.3.3). Los cebadores utilizados en las diversas estrategias de clonación, así como las condiciones de amplificación, se describen en el APARTADO 8.2 de Materiales y Métodos, ya que estos oligonucleótidos requieren características adicionales.

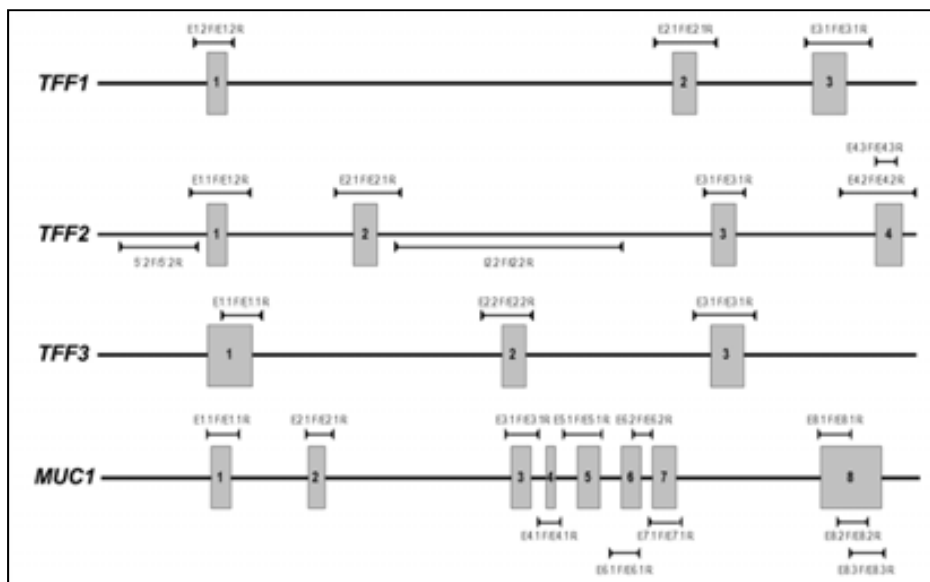


Figura M-3. Distribución a lo largo de los genes *TFF1*, *TFF2*, *TFF3* y *MUC1* de los cebadores diseñados

En la reacción de amplificación se tuvieron en cuenta diferentes parámetros como la concentración de reactivos, el número de ciclos y la temperatura de anillamiento. Se tuvo en cuenta, especialmente, la concentración de $MgCl_2$, factor crucial en el funcionamiento de la polimerasa. El procedimiento general de PCR se describe en el PROTOCOLO III y las condiciones específicas de amplificación para cada fragmento se detallan en la TABLA M-3. En todas las tandas de amplificación por PCR se incluía un control negativo, con dH_2O en vez de DNA.

Las reacciones de PCR se realizaron en los termocicladores *GeneAmp 2400* (Perkin Elmer), *GeneAmp 9700 dual* (Applied Biosystems) y *PTC-240 Tetrad 2* (MJ Research). La aparición de termocicladores que permiten programar condiciones de PCR más variables, como la tetrada *PTC-240* que incorpora la opción de autoextensión y de gradiente, facilitó el proceso de optimización de la PCR.

Tabla M-1. Cebadores utilizados para amplificar el DNA genómico de los genes TFF.

Nombre	Secuencia	Nº bases	Posición ^a	Localización	Región amplificada	pb ^b	Tm ^c	%GC
<i>TFF1</i> E1.2 F	5' TAGGCAAATGTTATCTAACG 3'	20	883-902	Genómica a 5'	Exón 1	228	54°	35,0%
<i>TFF1</i> E1.2 R	5' CCCACAGAGCAGGAAGAAG 3'	19	1092-1110	Intrón 1			60°	57,9%
<i>TFF1</i> E2.1 F	5' GCTGATGTCATTGGAGAAGG 3'	20	3986-4005	Intrón 1	Exón 2	362	60°	50,0%
<i>TFF1</i> E2.1 R	5' GACTCTGTGTAAGGCATAG 3'	20	4328-4347	Intrón 2			58°	45,0%
<i>TFF1</i> E3.1 F	5' CATCTCTTTCTCCCTCCCC 3'	19	4965-4983	Intrón 2	Exón 3	378	60°	57,9%
<i>TFF1</i> E3.1 R	5' TGCCCACTTGTATCCCAGC 3'	19	5324-5342	Genómica a 3'			60°	57,9%
<i>TFF2</i> 5'.2 F	5' ACCCTGTGAGCCTCATCTC 3'	19	15941-15969	Genómica a 5'	Genómica 5'	453	60°	57,9%
<i>TFF2</i> 5'.1 R	5' CGTGGAGGCGAAGAGGAGGC 3'	20	16377-16396	Genómica a 5'			68°	70,0%
<i>TFF2</i> E1.1 F	5' CCTCCTCTTCGCCTCCAC 3'	18	16378-16395	Genómica a 5'	Exón 1	357	60°	66,7%
<i>TFF2</i> E1.2 R	5' AACACACATTGCCAGATGCC 3'	20	16716-16735	Intrón 1			60°	50,0%
<i>TFF2</i> E2.1 F	5' CAACAAGACAGTAGAGGCATC 3'	21	17367-17387	Intrón 1	Exón 2	371	62°	47,6%
<i>TFF2</i> E2.1 R	5' ACTAACCACCCAGGCTAAC 3'	20	17719-17738	Intrón 2			62°	55,0%
<i>TFF2</i> I2.2 F	5' GTTAGCCTGGGTGGGTTAGTC 3'	21	17719-17739	Intrón 2	Intrón 2	1152 o 1277	66°	57,1%
<i>TFF2</i> I2.2 R	5' CCTTCCTTTGAGCACTCGTCG 3'	21	19250-19270	Intrón 2			66°	57,1%
<i>TFF2</i> E3.1 F	5' ACTGGGTCTCGGTGCTGTG 3'	19	19826-19844	Intrón 2	Exón 3	244	62°	63,2%
<i>TFF2</i> E3.1 R	5' CTACGGACGGAGGAGGAATC 3'	20	20050-20069	Intrón 3			64°	60,0%
<i>TFF2</i> E4.2 F	5' GCCCCACAGAGGAAACTC 3'	18	20758-20774	Intrón 3	Exón 4	427	60°	61,1%
<i>TFF2</i> E4.2 R	5' ACACGCACACCACCATTAC 3'	19	21166-21184	Genómica a 3'			58°	52,6%
<i>TFF2</i> E4.3 F	5' TACTAAGAGAGGTCTGTTC 3'	19	20954-20972	Intrón 3	Exón 4	127	54°	42,1%
<i>TFF2</i> E4.3 R	5' GAGGAAAAGATGGTTAAGAAA 3'	21	21060-21080	Exón 4			56°	127
<i>TFF3</i> E1.1 F	5' AGCAAACAATCCAGAGCAGC 3'	20	52017-52036	Exón 1	Exón 1	235	60°	50,0%
<i>TFF3</i> E1.1 R	5' ATCCCTTCCCTTCACGCC 3'	19	52233-52251	Intrón 1			62°	63,2%
<i>TFF3</i> E2.2 F	5' GTGGTTGAACTCGGCGGG 3'	18	53741-54768	Intrón 1	Exón 2	295	60°	66,7%
<i>TFF3</i> E2.2 R	5' CCCAGACCACGATGCCAC 3'	18	54018-54035	Intrón 2			60°	63,2%
<i>TFF3</i> E3.1 F	5' GAGCCTGCCCTGACCTTG 3'	18	55144-55161	Intrón 2	Exón 3	364	60°	66,7%
<i>TFF3</i> E3.1 R	5' GACTCTCCCCTGACACC 3'	18	55490-55507	Genómica a 3'			60°	66,7%

a, posición en la secuencia de referencia del DNA genómico del cluster de *TFFs* (AB038162); b, tamaño del fragmento amplificado en pares de bases; c, Tm calculada por el método $2(A+T)+4(G+C)$.

Tabla M-2. Cebadores utilizados para amplificar el DNA genómico del gen *MUC1*.

Nombre	Secuencia	Nº bases	Posición ^a	Localización	Región amplificada	pb ^b	Tm ^c	%GC
<i>MUC1</i> E1.1 F	5'-CGGTATAAAGCGGTAGGCG-3'	19	6651357-6651375	Genómica a 5'			60°C	57,9%
<i>MUC1</i> E1.3 R	5'-GAGCAAAACCCACAGAAAGAC-3'	21	6651147-6651167	Intrón 1	Exón 1	216	62°C	47,6%
<i>MUC1</i> E2.1 F	5'-CAGCCCCTAAACCCGCAAC-3'	19	6650719-6650737	Intrón 1			62°C	63,2%
<i>MUC1</i> E2.1 R	5'-GAGTGGTGGAGGAGCCTG-3'	18	6650555-6650572	Intrón 2	Exón 2	184	60°C	66,7%
<i>MUC1</i> E3.1 F	5'-TACCTCCTCTCACCTCCTC-3'	19	6649426-6649246	Intrón 2			60°C	57,9%
<i>MUC1</i> E3.1 R	5'-ACAGTTCTGATGGCTGCTTC-3'	20	6649227-6649246	Intrón 3	Exón 3	218	60°C	50,0%
<i>MUC1</i> E4.1 F	5'-CATCAGAAGTGTCCACACCC-3'	20	6649219-6649238	Intrón 3			62°C	55,0%
<i>MUC1</i> E4.1 R	5'-CCACCCTCCAACCACCAC-3'	18	6649083-6649100	Intrón 4	Exón 4	156	60°C	66,7%
<i>MUC1</i> E5.1 F	5'-GACCGTAGGGAGGGACTG-3'	18	6649054-6649071	Intrón 4			60°C	66,7%
<i>MUC1</i> E5.1 R	5'-GCAGCCAGGGAAGTAGCC-3'	18	6648818-6648835	Intrón 5	Exón 5	247	60°C	66,7%
<i>MUC1</i> E6.1 F	5'-TCCCCGCTCTTTCTTAGTG-3'	20	6648755-6648774	Intrón 5			62°C	55,0%
<i>MUC1</i> E6.1 R	5'-GGCAATGAGATAGACAATGGC-3'	21	6648576-6648596	Exón 6	Exón 6	199	62°C	47,6%
<i>MUC1</i> E6.2 F	5'-GGTGCTGGTCTGTGTTCTG-3'	19	6648606-6648624	Exón 6			60°C	57,9%
<i>MUC1</i> E6.2 R	5'-TGGGGAGATAGGAGGTTATG-3'	20	6648494-6648513	Intrón 6	Exón 6	131	60°C	50,0%
<i>MUC1</i> E7.1 F	5'-CTGCCATAACCTCCTATCTC-3'	20	6648498-6648517	Intrón 6			60°C	50,0%
<i>MUC1</i> E7.1 R	5'-AAACCCTCTGCTTCCCCTG-3'	19	6648303-6648321	Intrón 7	Exón 7	205	60°C	57,9%
<i>MUC1</i> E8.1 F	5'-ATCTCTTTTCTTCCACCCAG-3'	21	6647328-6647348	Intrón 7			62°C	47,6%
<i>MUC1</i> E8.1 R	5'-CTGAGCAGCCACCTGAAC-3'	19	6647130-6647147	Exón 8	Exón 8	219	62°C	63,2%
<i>MUC1</i> E8.2 F	5'-GCCAGCCAGTGCCATTCC-3'	18	6647209-6647226	Exón 8			60°C	61,1%
<i>MUC1</i> E8.2 R	5'-GAGCCCCCACCACAACAC-3'	18	6647024-6647041	Exón 8	Exón 8, 3'UTR	203	60°C	66,7%
<i>MUC1</i> E8.3 F	5'-TCAGGTGGGCTGCTCACAG-3'	19	6647127-6647145	Exón 8			62°C	63,2%
<i>MUC1</i> E8.3 R	5'-CACGGGTCACAGATGAAAGAT-3'	21	6646915-6646935	Genómica a 3'	Exón 8, 3'UTR	231	62°C	47,6%

a, posición en la secuencia de referencia del DNA genómico de *MUC1* (NT_004487.19); b, tamaño del fragmento amplificado en pares de bases; c, Tm calculada por el método $2(A+T)*4 (G+C)$

Tabla M-3. Condiciones específicas para la amplificación de diferentes fragmentos de gDNA de los genes *TFF1*, *TFF2*, *TFF3*, *MUC1*

Gen	Cebadores	pb	Condiciones de la reacción (volumen final 25 µl) ^a					Condiciones de la amplificación ^b					
			Tampón 10X	MgCl2 25mM	dNTPs 2mM	Cebadores 10µM	Taq 5U/µl	Ciclos	Desnatar.	Anillam.	Extensión	Aplicación	
TFF1	TFF1 E1.2F/TFF1 E1.2R	228	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	58°C 30s	72°C 30 s	SSCP, RFLP	
TFF1	TFF1 E2.1F/TFF1 E2.1R	362	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	60°C 30s	72°C 30 s	SSCP	
TFF1	TFF1 E3.1F/TFF1 E3.1R	378	2,5	1	3	1,25+1,25	0,1	40	94°C 30s	59°C 30s	72°C 30 s	SSCP	
TFF2	TFF2 5'.2F/TFF2 5'.1R	453	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	63°C 30s	72°C 30 s	SSCP	
TFF2	TFF2 E1.1F/TFF2 E1.2R	357	2,5	0	3	1,25+1,25	0,1	40	94°C 30s	66°C 30s	72°C 30 s	SSCP	
TFF2	TFF2 E2.1F/TFF2 E2.1R	371	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	63°C 30s	72°C 30 s	SSCP	
TFF2	TFF2 I2.2F/TFF2 I2.2R	1152	2,5	0	3	1,25+1,25	0,15	15	94°C 30s	65°C 30s	72°C 2min	Electroforesi	
								25	94°C 30s	65°C 30s	72°C		
TFF2	TFF2 E3.1F/TFF2 E3.1R	244	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	63°C 30s	72°C 30 s	SSCP	
TFF2	TFF2 E4.2F/TFF2 E4.2R	427	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	58°C 30s	72°C 30 s	SSCP	
TFF3	TFF3 E1.1F/TFF3 E1.1R	235	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	66°C 30s	72°C 30 s	SSCP, RFLP	
TFF3	TFF3 E2.2F/TFF3 E2.2R	295	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	66°C 30s	72°C 30 s	SSCP	
TFF3	TFF3 E3.1F/TFF3 E3.1R	364	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	63°C 30s	72°C 30 s	SSCP	
MUC1	MUC1 E1.1F/MUC1 E1.3R	229	2,5	1	3	1,25+1,25	0,1	35	94°C 30s	59°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E2.1F/MUC1 E2.1R	184	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E3.1F/MUC1 E3.1R	218	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E4.1F/MUC1 E4.1R	156	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E5.1F/MUC1 E5.1R	231	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E6.1F/MUC1 E6.1R	199	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E6.2F/MUC1 E6.2R	131	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	60°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E7.1F/MUC1 E7.1R	205	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E8.1F/MUC1 E8.1R	219	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E8.2F/MUC1 E8.2R	203	2,5	0	3	1,25+1,25	0,1	35	94°C 30s	60°C 30s	72°C 30 s	SSCP, HRM	
MUC1	MUC1 E8.3F/MUC1 E8.3R	231	2,5	1	3	1,25+1,25	0,1	35	94°C 30s	61°C 30s	72°C 30 s	SSCP, HRM	

a, Las unidades que se indican son µl; b, Todas las reacciones de amplificación se inician con una desnaturalización a 94°C de 10 minutos y finalizan con una extensión a 72°C 7 minutos.

4.1. Purificación de productos de PCR

En este trabajo se ha realizado la purificación de fragmentos amplificados por PCR cuando éstos requerían ser secuenciados o clonados. Este proceso es muy importante ya que los reactivos residuales que no se hayan eliminado pueden interferir en la eficiencia de las subsiguientes aplicaciones. Para purificar los productos de PCR de los diferentes reactivos de la mezcla de amplificación, como cebadores, dNTPs, sales y enzimas, se utilizaron kits comerciales específicos para este fin: *JETQUICK PCR Product Purification Spin Kit* (Genomed) o *QIAquick PCR purification kit* (Qiagen). Ambos kits permiten la purificación de productos de PCR de un tamaño de entre 100 pb y 10 kb. Estos kits comerciales se basan en la utilización de columnas que contienen membranas de sílica, con propiedades específicas de unión DNA. El DNA se adsorbe en la membrana, en presencia de una alta concentración de sales, y las impurezas pasan a través de la columna durante las diferentes fases del lavado. Finalmente se eluye el DNA con agua bidestilada, obteniendo el producto de PCR purificado. En ambos casos se siguieron las instrucciones del fabricante.

5. SECUENCIACIÓN DEL DNA

En este trabajo la secuenciación se utilizó, principalmente, para confirmar y caracterizar nuevas variantes, previamente observadas mediante el análisis de SSCP o HRM. Esta técnica también permitió confirmar la identidad de los fragmentos amplificados para su posterior clonaje y, en una fase posterior, para comprobar que los clonajes definitivos fueran correctos.

La determinación de la secuencia de DNA de los productos de PCR se realizó mediante secuenciación automática, basada en el método de Sanger.

Para las reacciones de secuenciación se usó el reactivo comercial *Big Dye Terminator Mix® v3.1* (Applied Biosystems) que incluye ddNTPs, cada uno de ellos marcado con diferente fluorescencia, permitiendo automatizar el proceso en una sola reacción (ver PROTOCOLO IV). El producto de PCR a secuenciar debe estar previamente purificado, por columnas o por extracción de banda a partir de un gel de agarosa (ver Materiales y Métodos, APARTADOS 4.1 y 6.1.1). Para cada reacción sólo se utiliza uno de los cebadores (directo o reverso) y de todos los fragmentos se secuenciaron las dos cadenas (directa y reversa).

Una vez finalizada la reacción de secuenciación, el producto se purificó por filtración en gel, en columnas de exclusión molecular *Sephadex G-50* (Amersham) o, alternativamente, mediante precipitación con etanol, tal y como se describe en el PROTOCOLO V. Las muestras eluidas o precipitadas se liofilizaron y se guardaron a -20°C hasta su análisis en el secuenciador.

Los productos de la PCR de secuenciación se analizaron con el secuenciador automático *ABI PRISM 3130 XL Genetic Analyzer* o *ABI PRISM 3730* (Applied Biosystems), utilizando el programa de análisis de secuencias *DNA Sequencing Analysis Software v5.1*. Las secuencias resultantes fueron comparadas con la secuencia de referencia utilizando el programa *Sequencher v4.5* o *DNA Star*.

6. ELECTROFORESIS

Los productos de PCR y de digestiones con enzimas de restricción fueron separados electroforéticamente en geles de agarosa o en geles de poliacrilamida. También se sometió a electroforesis el DNA plasmídico para comprobar su calidad.

La electroforesis en gel es una técnica que permite separar las moléculas en solución según su movilidad en un campo eléctrico y, generalmente, a través de una matriz porosa. En estas condiciones, las moléculas de ácidos nucleicos son atraídas hacia el polo opuesto a su carga neta, que es negativa. La migración de las moléculas a través de la matriz, por tanto, se realiza en base a su tamaño y carga eléctrica.

6.1. Electroforesis en geles de agarosa

Este tipo de electroforesis se utilizó como técnica analítica para separar y visualizar fragmentos de DNA genómico, amplificados mediante PCR o generados mediante digestión con enzimas de restricción, y DNA plasmídico. La electroforesis en geles de agarosa ofrece un método rápido y sencillo para la separación, identificación y purificación de fragmentos de ácidos nucleicos de entre 0,1-0,2 y 25 kb. El porcentaje de agarosa a utilizar, que determina el tamaño del poro por donde las moléculas van a migrar, varía en función del tamaño del ácido nucleico. Para separar fragmentos grandes, de aproximadamente 2 kb en adelante, se utilizan geles poco concentrados, entre el 0,7% y el 1%. En cambio, fragmentos más pequeños se discriminan mejor a concentraciones de agarosa algo mayores, de 1% a 2%. Para fragmentos muy pequeños, menores de 100-200 pb, se utiliza un porcentaje del 3% aproximadamente.

El procedimiento general, PROTOCOLO VI, puede dividirse en tres etapas: preparación del gel de agarosa a la concentración adecuada para el tamaño de los fragmentos a ser separados, carga de las muestras en los pocillos del gel y aplicación de un voltaje eléctrico y, finalmente, tinción y visualización del gel. La tinción se realizó con bromuro de etidio, un agente intercalante de ácidos nucleicos que permite la visualización del DNA al emitir fluorescencia cuando es estimulado con luz ultravioleta. En la mayoría de casos la visualización se llevó a cabo en un transiluminador GelDoc (BioRad). Alternativamente, se utilizó un transiluminador UV Transilluminator 2000 (BioRad) para facilitar, principalmente, la visualización y corte de banda en procesos de purificación del DNA a partir de geles de agarosa.

6.1.1. Purificación del DNA de geles de agarosa

Los fragmentos de DNA separados por electroforesis en geles de agarosa pueden recuperarse intactos para utilizarlos en aplicaciones posteriores. Una vez realizada la electroforesis, preferentemente en un gel de agarosa de bajo punto de fusión y a una concentración comprendida entre el 1-2%, se recortó la banda de interés del gel y se purificó el DNA, utilizando los kits comerciales *DNA QIAquick Gel Extraction Kit* (Qiagen) o *Wizard® DNA Clean-Up System* (Promega) y siguiendo el protocolo recomendado por el fabricante con alguna pequeña modificación. La agarosa se disuelve mediante tampones que contienen sales

caotrópicas (hidrocloruro de guanidina), liberando el DNA para permitir su unión a la membrana de sílica dispuesta en la columna.

6.2. Electroforesis en geles de poliacrilamida

Los geles de poliacrilamida, ver PROTOCOLO VII, se utilizaron para separar fragmentos de restricción muy pequeños (inferiores a 200 pb) o similares entre sí y para el análisis de polimorfismos de conformación de la cadena sencilla y heterodúplex (SSCP/HD) de fragmentos de hasta 450 pb, amplificados a partir de DNA genómico. Ésta última aplicación presenta algunas diferencias respecto a una electroforesis convencional que se describen en el APARTADO 7.2 de Materiales y Métodos.

Los geles de poliacrilamida se forman por la copolimerización del monómero de acrilamida con el agente entrecruzador bis-acrilamida. La polimerización requiere la presencia de un iniciador N,N,N',N'-tetrametiletilendiamina (TEMED) y un catalizador, el persulfato de amonio (APS). La porosidad de los geles de poliacrilamida está determinada por la proporción relativa acrilamida - bis-acrilamida y por la concentración global de acrilamida en el gel. Estos geles permiten una mayor resolución que los geles de agarosa, ya que estos últimos presentan unas dimensiones de poro más grande. Los geles se prepararon con una proporción acrilamida-bis-acrilamida 29:1, para SSCPs o 19:1, para visualizar fragmentos de restricción, y con un porcentaje global del 10-12% dependiendo del tamaño de las moléculas a separar.

Los fragmentos se visualizaron mediante tinción con bromuro de etidio y posterior exposición a luz ultravioleta en un transiluminador, o mediante revelado con nitrato de plata (ver PROTOCOLO VIII). El revelado con nitrato de plata es más laborioso pero mucho más sensible que la tinción con bromuro de etidio y permite conservar el gel.

7. ANÁLISIS DE POLIMORFISMOS

Los SNPs potencialmente funcionales seleccionados de las bases de datos públicas se validaron en una población control de 50 individuos (el equivalente a 100 cromosomas) mediante análisis con enzimas de restricción o SSCPs de fragmentos previamente amplificados por PCR. La técnica de SSCPs nos permitió también identificar nuevos polimorfismos a lo largo de las regiones funcionales de los genes de *TFFs* y *MUC1*. En el gen *MUC1*, la búsqueda de polimorfismos se realizó por duplicado utilizando la técnica *High Resolution Meeting* (HRM) de Roche, además del análisis de los SSCPs.

El genotipado de los polimorfismos seleccionados en genes de protección de la mucosa (n=58) y genes involucrados en reparación (n=4), inflamación (n=31) e interacción con Hp (n=18) se realizó con diversas metodologías y plataformas descritas en este apartado. En las TABLA R-1 de resultados se indican los polimorfismos analizados y la metodología utilizada en las muestras del estudio de seguimiento de pacientes con lesiones gástricas y en las del estudio de cáncer gástrico, respectivamente.

7.1. Análisis de las variantes con enzimas de restricción

El análisis de restricción o *Restriction Fragment Length Polymorphisms* (RFLPs) se utilizó para validar los polimorfismos seleccionados en las bases de datos de genes *TFFs*, en los que una de las variantes destruía o generaba una diana de restricción. La búsqueda de las posibles dianas de restricción específicas de cada secuencia se realizó con el programa DS Gene. Las condiciones específicas del análisis de restricción para cada polimorfismo, así como las enzimas y tampones correspondientes utilizados se detallan en la TABLA M-4. En algunos casos, se añadió albúmina bovina sérica (BSA), una proteína que estabiliza la enzima y puede unirse a contaminantes inhibidores, permitiendo obtener la máxima actividad de la enzima y aumentar la eficiencia de la digestión.

En el PROTOCOLO IX se describe el procedimiento general para la digestión con enzimas de restricción que puede utilizarse tanto para el análisis de variantes polimórficas como en diferentes fases de la clonación. Los fragmentos obtenidos se visualizaron mediante electroforesis en gel de agarosa al 3%, o en geles de poliacrilamida 19:1 al 12% cuando el tamaño de los fragmentos era muy pequeño o similar, y se tiñeron con bromuro de etidio.

Tabla M-4. Polimorfismos analizados utilizando enzimas de restricción

SNP	Cebadores	Enzima	Secuencia diana	Condiciones de digestión (V _r =20μl)				
				Ta ^a	Enz U	Tampón ^b	BSA 10ng/μl	PCR
rs11701143	TFF3_E1 fw1/rv1	Avall	5'...GGWCC...3' 3'...CCWGG...5'	37°	5 U	4 (2 μl)	-	5 μl
rs215630	TFF1_E1 fw2/rv2	BsrDI	5'...GCAATGNN...3' 3'...CGTTAÇNN...5'	65°	1 U	2 (2 μl)	0,2 μl	6 μl
rs11546960	TFF3_E1 fw1/rv1	HaeII	5'...GGCC...3' 3'...CCGG...5'	37°	5 U	2 (2 μl)	-	5 μl

a, Temperatura de incubación durante toda la noche; b, N° de Tampón correspondiente a cada enzima (New England Biolabs®)

7.2. Polimorfismo de conformación de cadena sencilla y heterodúplex (SSCP/HD)

El análisis de los polimorfismos de conformación de cadena sencilla (*Single Strand Conformation Polymorphism*, SSCP) se fundamenta en la tendencia que tienen las cadenas sencillas de DNA de formar apareamientos intracatenarios que resultan en estructuras secundarias únicas. Estas estructuras tienen una movilidad electroforética específica en geles no desnaturizantes. Un único cambio de nucleótido puede alterar el apareamiento de bases del DNA de cadena sencilla y, consecuentemente, su conformación al plegarse. Estas alteraciones se pueden observar como cambios en la movilidad electroforética que generarán nuevos patrones de bandas. En el análisis de SSCPs, algunas de las moléculas de DNA de cadena sencilla consiguen renaturalizar y formar homodúplex, si la unión entre las cadenas es totalmente complementaria, o heterodúplex, debido a la unión de una cadena nativa con la cadena que presenta la variante. Estos heterodúplex también presentarán un patrón de movilidad electroforética diferente de la de los homodúplex.

La técnica de SSCs, PROTOCOLO X, se utilizó para validar y calcular las frecuencias alélicas de los polimorfismos potencialmente funcionales seleccionados de *TFFs* y *MUC1* para los que no se pudo realizar el análisis de restricción, a la vez que permitió la identificación de nuevos polimorfismos en los fragmentos analizados. También se realizaron SSCs para la búsqueda de nuevos polimorfismos en el resto de fragmentos amplificados por PCR correspondientes a zonas funcionales, mayoritariamente exones y región adyacente, de estos genes. De esta manera, mediante la técnica SSCs se analizaron todos los amplicones de *TFF1*, *TFF2*, *TFF3* y *MUC1* especificados en la TABLA M-1 y M-2, exceptuando los correspondientes al intrón 2 y al fragmento de 127 pb del exón 4, de *TFF2* y al intrón 2 de *IL1RN*.

7.3. Genotipado de polimorfismos mediante PCR a tiempo real

El genotipado mediante PCR a tiempo real, en presencia de sondas marcadas o colorantes con fluorescencia, permite realizar todo el proceso, amplificación y discriminación alélica, en un sólo paso, simplificando el proceso y evitando posibles contaminaciones. Las reacciones que incorporaban sondas de hibridación FRET (del inglés *Fluorescence Resonance Energy Transfer*, FRET) o HRM se llevaron a cabo en el termociclador para PCR a tiempo real, LightCycler® (LC) 480 (Roche Applied Science). El genotipado mediante sondas de hidrólisis TaqMan se realizó en el RealTime ABI PRISM® 7900HT Sequence Detector (Applied Biosystems).

7.3.1. Genotipado con sondas de hibridación FRET

Las sondas de hibridación FRET constan de una sonda *sensor*, complementaria a la secuencia polimórfica y una sonda *anchor*, complementaria a una secuencia adyacente (a 2-3 nucleótidos) a la anterior.

Una de las dos sondas marcada con un fluoróforo a 3' actuará como donadora y la otra, marcada a 5', como receptora. Cuando ambas sondas hibriden específicamente, la proximidad permitirá el fenómeno de transferencia de energía (FIGURA M-4A) y, como resultado, se detectará la emisión de fluorescencia. La incorporación de sondas fluorescentes en la mezcla de reacción hace posible un seguimiento de la reacción de amplificación a lo largo del tiempo (PCR a tiempo real) y, posteriormente, el análisis del genotipo mediante la determinación de la curva de fusión de la sonda *sensor*, específica para el SNP a analizar. Una vez finalizada la PCR, para determinar la curva de fusión, se realizó una desnaturalización total a 95°C, una renaturalización inespecífica a baja temperatura y una lenta desnaturalización, con 1- 5 adquisiciones de fluorescencia por grado, en la que las sondas deshibridan a una determinada temperatura y dejan de emitir fluorescencia. El software determina la variación de fluorescencia en función de la temperatura, y obtiene la temperatura de fusión (*melting temperature*, T_m) de la sonda *sensor*, que es aquella en la que un 50% de las moléculas han deshibridado. Cuando la sonda es totalmente complementaria a la cadena molde su T_m es mayor, ya que hibrida completamente. Si en la secuencia existe una variante, el apareamiento es incompleto (*mismatch*) y la T_m menor (FIGURA M-4B).

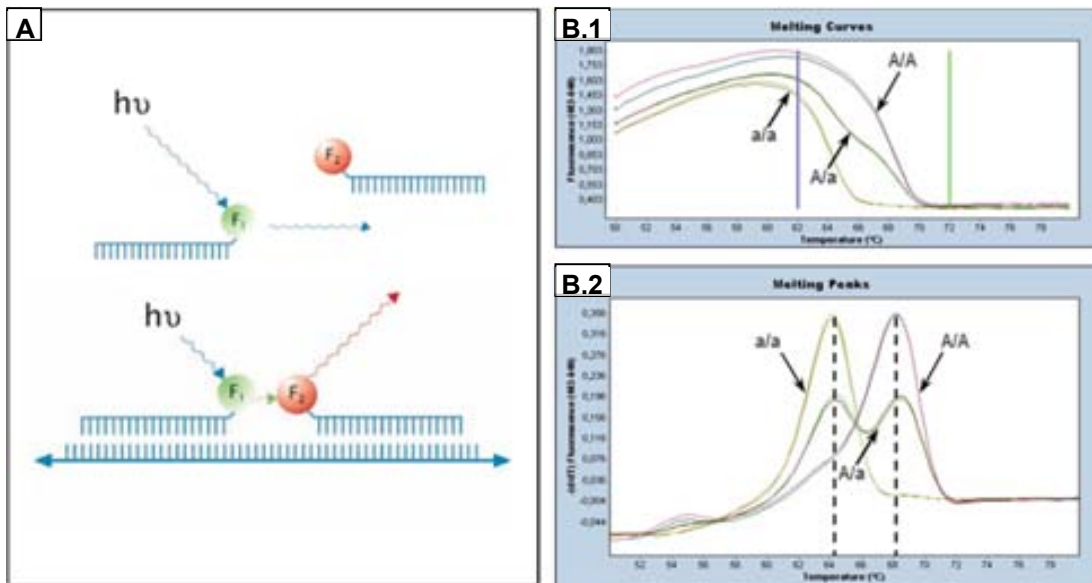


Figura M-4. A. Transmisión de Energía Fluorescente, FRET. B. Análisis de un polimorfismo mediante sondas de hibridación FRET en el LC480. B1. Curva de fusión de la sonda sensor específica para el alelo A. B2. Segunda derivada de la curva de fusión indicada en B1. B.2. Tm del fragmento con el alelo específico (A) para la sonda sensor: 68°C. Tm del fragmento que presenta un mismatch para la sonda sensor (a) 64°C.

Todas las sondas fueron sintetizadas por TIB MOLBIOL (<http://www.TIB-MOLBIOL.de>). En algunos casos los diseños de las sondas y cebadores utilizados han sido realizados por TIBMOLBIOL y, en otros, han sido diseños propios con la ayuda de los programas *Oligo 6* y el *LightCycler Probe Design Software 2* (Roche Applied Science), teniendo en cuenta las siguientes consideraciones:

- Seguir los criterios descritos en el APARTADO 4 para el diseño de oligonucleótidos cebadores.
- Que no exista otra variación en la secuencia circundante al SNP que queremos analizar, sobre todo en la secuencia donde finalmente hibridarán las dos sondas.
- Que la temperatura de fusión de la sonda sensor fuera 4-6°C menor que la de la sonda anchor, de modo que esta última se desprenda más tarde y se pueda asegurar que la temperatura de fusión medida es la de la sonda sensor.

El procedimiento general se describe en el PROTOCOLO XI y en la TABLA M-5 se indican los SNPs genotipados en este trabajo mediante el uso de sondas FRET en un el LC 480, así como los cebadores y sondas utilizadas para cada polimorfismo analizado y las condiciones de amplificación y análisis de la curva de fusión.

Tabla M-5. Cebadores, sondas FRET y condiciones de amplificación y determinación de la curva de fusión en el genotipado de SNPs con el LC480

ID SNP	Oligon.	Secuencia	Amplificación	Fusión
TFF1 rs2156310	Anchor	5'-LC640-CCAAAGCGACCCCCGAGTC-PH-3'		
	Sensor A	5'-GGCCAATTACCTCCTCTCTGC-Fluoresceína-3'	95°C 10 s	95°C 30 s
	Directo	5' TAGGCAAATGTTATCTAACG 3'	60°C 30 s	45°C 30 s
	Reverso	5' CCCACAGAGCAGGAAGAAG 3'	72°C 15 s	75°C 3adq/°C
TFF3 rs11701143	Anchor	5'-CTCCGGGACGCAGCTCAGGAC-Fluoresceína-3'		
	Sensor C	5'-LC640-CGCTTCATGGCCAGGAG-PH-3'	95°C 10 s	95°C 30 s
	Directo	5'-TGCAAACAACGGTGCATAA-3'	56°C 30 s	45°C 30 s
	Reverso	5'-CTCACACAGGCCACGTA-3'	72°C 18 s	75°C 5adq/°C
TFF2 rs13052596	Anchor	5'-LC640-GGCCTGTACATTTTCCCCAGTAAACTTGGC-PH-3'		
	Sensor T	5'-GGGAGTGGACTTTTCCATGCTATCTGC-Fluoresceína-3'	95°C 10 s	95°C 30 s
	Directo	5'-TCCTGCTGAAGATGGGACA-3'	63°C 30 s	45°C 30 s
	Reverso	5'-AGGGACCACAGGAATTTAGAA-3'	72°C 18 s	80°C 3adq/°C
TFF2 rs4920084	Anchor	5'-LC640-TCTTTTGGTTTCATTAACACAGAGGTGGT-PH-3'		
	Sensor C	5'-TCTTTAAATAAAACCCTTAAATCTGCTAA-Fluoresceína-3'	95°C 10 s	95°C 30 s
	Directo	5'-CTTCTAGACTGCCATTACTAAGAGA-3'	64°C 30 s	40°C 30 s
	Reverso	5'-ATACACTCACACATATGCACATTC-3'	72°C 18 s	70°C 1adq/°C
TFF2 rs225334	Anchor	5'-LC640-TCTTAACCATCTTTTCTCATTTCATGGTT-PH-3'		
	Sensor C	5'-CTTCATATTCATGAAATCCTGGGTT-Fluoresceína-3'	95°C 10 s	95°C 30 s
	Directo	5'-CTTCTAGACTGCCATTACTAAGAGA-3'	64°C 30 s	40°C 30 s
	Reverso	5'-ATACACTCACACATATGCACATTC-3'	72°C 18 s	70°C 1adq/°C
ERCC2 rs13181	Anchor	5'-GCCTGGAGCTAGAATCAGAGGAGA-Fluoresceína-3'		
	Sensor C	5'-LC640-CTGCAGAGGATAGAGCAGATTGC 3'	95°C 10 s	95°C 30 s
	Directo	5'-TCTGGATTATACGGACAT-3'	60°C 30 s	45°C 30 s
	Reverso	5'-GTCACCTGACTTCATAAGA-3'	72°C 30 s	80°C 3adq/°C
ERCC2 rs1799793	Anchor	5'-LC640-TGCTGCAGTGAGCCCCGACC-PH-3'		
	Sensor A	5'-AACCCCGTGCTGCCCAACGAA-Fluoresceína-3'	95°C 10 s	95°C 30 s
	Directo	5'-GATCAAAGAGACAGACGAGC-3'	58°C 30 s	40°C 30 s
	Reverso	5'-ACTTCACGTACTIONCCAGCAG-3'	72°C 18 s	80°C 3adq/°C
XRCC1 rs25487	Anchor	5'-LC640-CCTCACACGCCAACCCCTGCTCCTT-PH-3'		
	Sensor A	5'-GCTGCCCTCCCAGAGGTAAG-Fluoresceína-3'	95°C 10 s	95°C 30 s
	Directo	5'-TTGCCAACACCCCCAGTA-3'	60°C 30 s	45°C 30 s
	Reverso	5'-GGCAGGCCCCAGTCTGA-3'	72°C 15 s	72°C 3adq/°C
XRCC1 rs1799782	Anchor	5'-AGGCCGGGGCTCTCT-Fluoresceína-3'		
	Sensor G	5'-LC640-TTCAGCCGGATCAACAAGACAT-PH-3'	95°C 10 s	95°C 30 s
	Directo	5'-TGTACGGTTTCATAGCC-3'	58°C 30 s	45°C 30 s
	Reverso	5'-CACGAGTCTAGGTCTCA-3'	72°C 18 s	72°C 3adq/°C

a, Oligonucleótidos (sonda: anclor y sensor; cebador: directo y reverso); b, Desnaturalización inicial (600s a 95°C) seguido de 45 ciclos de amplificación.

7.3.2. Genotipado con sondas de hidrólisis TaqMan®

Las sondas de hidrólisis TaqMan® son oligonucleótidos de 18-22 pb, aproximadamente, marcados con un fluorocromo donador (o *reporter*) en el extremo 5', que emite fluorescencia al ser excitado, y con una molécula inhibidora (o *quencher*) en el extremo 3', que absorbe la fluorescencia liberada por el donador. La proximidad del *quencher* a la molécula donadora, en una sonda intacta, provoca la inhibición de la señal fluorescente.

Cuando se usa este tipo de química para el genotipado de polimorfismos, se utilizan dos sondas TaqMan. A diferencia de las sondas FRET, cada una de las sondas TaqMan es específica para un alelo del polimorfismo y cada una está marcada con fluorocromos distintos, generalmente VIC y FAM. Durante la amplificación del DNA diana, si la sonda hibrida completamente a su cadena complementaria en el DNA molde, ésta es hidrolizada por la actividad 5' exonucleasa de la polimerasa. De esta manera se libera la molécula reportera, que ya no se encontrará próxima al *quencher* y, consecuentemente, la intensidad de la fluorescencia emitida por el fluorocromo donador incrementa y puede ser detectada (FIGURA M-5).

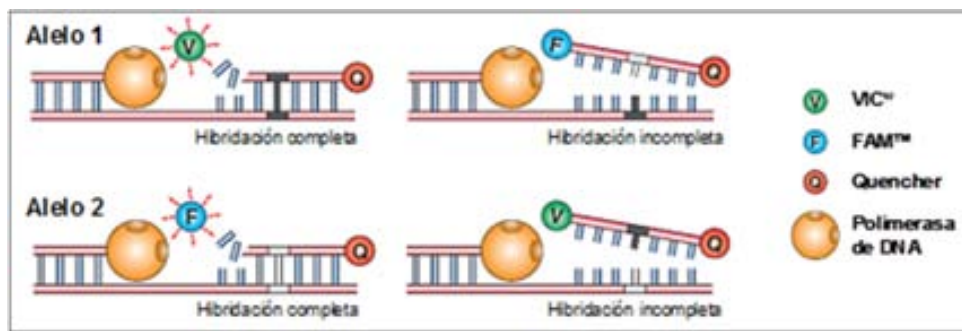


Figura M-5. Funcionamiento de las sondas de hidrólisis TaqMan

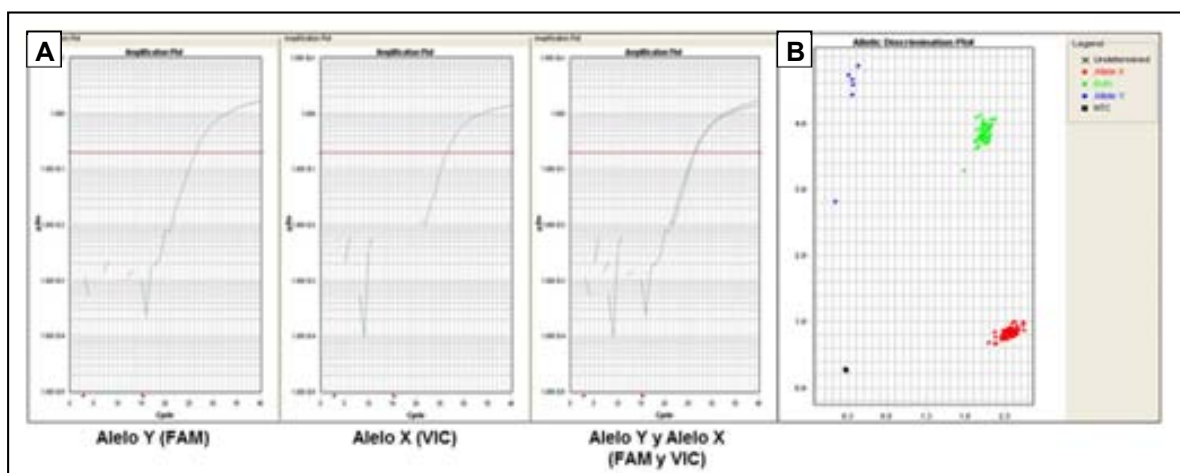


Figura M-6. A. Ejemplo de gráficos de amplificación. B. Diagrama de puntos de un experimento de genotipado mediante sondas TaqMan®. En rojo, fluorescencia obtenida por el fluorocromo VIC, que en este caso es el que marca la sonda diseñada para el alelo X. En azul, fluorescencia obtenida con FAM, que marca el alelo Y y en verde, ambas fluorescencias, que representarían a los heterocigotos.

Así, la cantidad de fluorescencia emitida por cada fluorocromo donador liberado es proporcional a la amplificación de cada uno de los alelos y la genotipación se puede realizar a través de dos tipos de análisis:

- Mediante la visualización de las curvas de amplificación, donde se puede observar si sólo captamos fluorescencia de la hidrólisis de una sonda (homocigotos) o de las dos (heterocigotos) (FIGURA M-7.A)
- Mediante el análisis del diagrama de puntos o los *dot-plots*. Un gráfico ilustra el ratio de la intensidad de las fluorescencias (VIC y FAM) captadas al final de la reacción, representando cada una de ellas en los ejes X e Y. De esta manera, los homocigotos se representan más o menos paralelos a uno de los ejes y los heterocigotos aparecen en la diagonal del gráfico (FIGURA M-7.B).

Tabla M-6. Polimorfismos en genes de respuesta inflamatoria genotipados mediante el uso de sondas TaqMan en el termociclador a tiempo real, ABI 7900HT.

Gen	ID SNP	Tipo SNP	Nº referencia	Mix	Cebadores	Sondas μ M	VIC	FAM	Cadena diseño
<i>IL10</i>	rs1800896	5' IL10	C_1747360_10	20X	18	4	T	C	Directa
<i>IL10</i>	rs180872	5' IL10	C_1747363_10	20X	18	4	T	G	Reversa
<i>IL12B</i>	rs1368439	3' UTR	C_7537839_10	20X	18	4	G	T	Directa
<i>IL12B</i>	rs3212227	3' UTR	C_2084293_10	40X	36	8	G	T	Directa
<i>IL12B</i>	rs3181224	Genómico	C_27467944_1	40X	36	8	A	G	Reversa
<i>IL12B</i>	rs3181226	Genómico	C_27467946_1	40X	36	8	C	G	Directa
<i>IL1A</i>	rs2856836	3' UTR	C_1839916_10	20X	18	4	A	G	Directa
<i>IL1A</i>	rs17561	A114S	C_9546471_10	20X	18	4	A	C	Reversa
<i>IL1A</i>	rs3783546	Intrónico	C_7628676_10	40X	36	8	C	G	Reversa
<i>IL1A</i>	rs1800587	5' UTR	C_9546481_20	40X	36	8	A	G	Reversa
<i>IL1B</i>	rs1143627	5' IL1B	C_1839944_10	20X	18	4	G	A	Directa
<i>IL1B</i>	rs1143623	5' IL1B	C_1839941_10	20X	18	4	C	G	Reversa
<i>IL1B</i>	rs1143634	F105F	C_9546517_10	20X	18	4	G	A	Directa
<i>IL1B</i>	rs16944	5' IL1B	C_1839943_10	20X	18	4	G	A	Directa
<i>IL1RN</i>	rs2637988	Intrónico	C_3133508_10	40X	36	8	A	G	Directa
<i>IL1RN</i>	rs4251985	Intrónico	C_3133512_10	40X	36	8	G	T	Reversa
<i>IL1RN</i>	rs4251961	5' IL1RN	C_32060323_1	40X	36	8	C	T	Reversa
<i>IL4</i>	rs2227284	Intrónico	C_11818513_1	20X	18	4	G	T	Directa
<i>IL4</i>	rs2070874	5' UTR	C_16176215_1	20X	18	4	C	T	Reversa
<i>IL4</i>	rs2243250	5' IL4	C_16176216_1	40X	36	8	C	T	Directa
<i>IL4R</i>	rs1805013	S436L	C_8903093_10	20X	18	4	C	T	Reversa
<i>IL4R</i>	rs2057768	Genómico	C_2769607_10	40X	36	8	C	T	Reversa
<i>IL4R</i>	rs1801275	Q576R	C_2351160_20	40X	36	8	A	G	Directa
<i>IL4R</i>	rs1808011	E400A	C_8903098_20	40X	36	8	A	C	Reversa
<i>IL4R</i>	rs1805010	I75L	C_2769554_10	40X	36	8	A	G	Directa
<i>IL6</i>	rs1800797	5' IL6	C_1839695_20	40X	36	8	A	G	Reversa
<i>IL8</i>	rs4073	Genómico	C_11748116_1	40X	36	8	A	T	Directa
<i>LTA</i>	rs909253	Intrónico	001_0238	40X	36	8	A	G	Reversa
<i>TNF</i>	rs1800629	5' TNF	C_7514879_10	20X	18	4	A	G	Reversa
<i>TNF</i>	rs361525	5' TNF	C_2215707_10	20X	18	4	A	G	Reversa

Todas las sondas TaqMan utilizadas en este trabajo fueron adquiridas en Applied Biosystems, que también proporciona los cebadores adecuados a utilizar con cada sonda en el análisis específico de los diversos polimorfismos. En la TABLA M-6 están indicados los SNPs analizados con este tipo de sondas y el número de referencia, en Applied Biosystems, del diseño de los ensayos correspondientes. El uso de estos ensayos TaqMan, que en su mayoría ya han sido validados por la casa comercial, permiten estandarizar el procedimiento de amplificación que es común para todos, tal y como se especifica en el PROTOCOLO XII.

7.3.3. Análisis de las variantes polimórficas con *High Resolution Melting* (HRM)

Esta técnica se utilizó, paralelamente a los SCCPs, para la validación y búsqueda de nuevas variantes en los exones de *MUC1*. También se utilizó para genotipar simultáneamente el SNP (rs225334) y una inserción de 3 pb (c.*68_69InsCTT) en el exón 4 de *TFE2*, en las muestras procedentes del estudio de seguimiento de pacientes con lesiones gástricas. Todos los análisis se llevaron a cabo en el instrumento de PCR a tiempo real LightCycler 480.

Previo análisis con HRM, se optimizó la amplificación de todos los fragmentos de interés mediante PCR a tiempo real en presencia de SYBR Green I, para comprobar que la reacción era específica. El SYBR Green I es un colorante fluorescente que durante la PCR se une al surco menor del DNA de doble cadena y emite fluorescencia a una intensidad proporcional al número de copias amplificadas, lo que permite la detección de los productos sintetizados. El análisis de la curva de fusión, una vez finalizada la amplificación, permite caracterizar el producto amplificado. Este análisis consiste en incubar el producto amplificado durante unos segundos a la temperatura en la que se mantiene la doble cadena, a 65°. Posteriormente se aumenta lentamente la temperatura, de manera que la doble cadena se desnaturaliza y se libera el SYBR Green I que dejará de emitir fluorescencia. El software determina la variación de fluorescencia en función de la temperatura y obtiene una T_m específica para cada secuencia. Una diferencia pequeña entre dos secuencias, cambios de un nucleótido o pequeñas inserciones o deleciones, varían muy ligeramente la T_m de los fragmentos. El SYBR Green I no tiene la sensibilidad suficiente para detectar estos pequeños cambios de T_m en el análisis de la curva de fusión.

El HRM es un método para la detección de mutaciones basado en la PCR a tiempo real y el análisis de alta resolución de la curva de fusión del DNA de doble cadena. Este método no utiliza SYBR Green I, sino un agente intercalador saturante del DNA que no interfiere con la polimerasa (376). De esta manera, se obtiene una mayor sensibilidad a los cambios de T_m producidos por pequeñas variaciones de la secuencia permitiendo detectar cambios de un sólo nucleótido, pequeñas deleciones o inserciones, sin necesidad del diseño de cebadores y/o sondas específicas marcadas. Con esta metodología, para identificar las variantes, no se compara la T_m de las diferentes muestras sino el perfil de la curva de fusión. Las variantes de secuencia en heterocigosis aparecerán durante el análisis de la curva de fusión como heterodúplex del DNA. Estas muestras mostrarán una curva de fusión con diferente perfil que aquellas que sean homocigotos salvajes u homocigotos mutantes. Los homodúplex salvajes o mutados formados por las variantes en homocigosis también pueden presentar un perfil diferente de curva de fusión

(FIGURA M-7). La sensibilidad de la técnica para la detección de mutaciones o variantes es mayor cuando los amplicones son pequeños. Por esta razón, los cebadores utilizados en la amplificación de los distintos exones de *MUC1*, que también se utilizaron para el análisis de SSCPs, se diseñaron específicamente para poder ser utilizados de manera óptima con esta técnica y los fragmentos amplificados no superaron las 250 pb (ver TABLA M-2). En el diseño de cebadores también se puso especial atención en la temperatura de anillamiento, para que ésta fuera similar para todos los fragmentos amplificados y poder realizar análisis simultáneos en placas de 384 pocillos. En el caso del exón 4 de *TFF2* se rediseñaron los cebadores de la reacción de PCR (TFF2 E4.3 F/E4.3 R, TABLA M-1) para que permitieran amplificar un fragmento más pequeño, 127 pb, y mejorar la detección de los alelos. Los fragmentos se amplificaron en presencia del agente intercalador, *ResoLight*, incluido en la *LightCycler® 480 High Resolution Melting Master Mix 2X* (Roche Applied Science), que también contiene la FastStart Taq DNA polimerasa, el tampón de reacción y los dNTPs. Una vez finalizada la amplificación, se determinó la curva de fusión de alta resolución mediante una desnaturalización muy lenta de 15-25 adquisiciones por grado. Para el análisis de la curva de fusión se utilizó el *LightCycler® 480 HRM Gene Scanning Software*. Los pasos a seguir en este análisis son:

- Normalización de las curvas de fusión: Se definen 2 regiones de la curva, una en que la fluorescencia sea del 100% (rango de temperatura pre-fusión) y otra en que la fluorescencia sea del 0% (rango de temperatura post-fusión), tal como se muestra en la FIGURA M-8A.
- Unificación de los puntos de fusión: El eje de la temperatura de cada curva se desplaza en la región de baja fluorescencia, para que las curvas formen grupos más compactos obteniendo la máxima distancia entre grupos (FIGURA M8-B).
- Representación gráfica de los perfiles de curva de fusión (FIGURA M8-C): El programa selecciona el grupo que contiene más muestras. El perfil de curva de este grupo se designa como el genotipo de referencia (curva base). La diferencia de fluorescencia entre esta referencia y cada curva restante (eje Y) se representa gráficamente en función de la temperatura (eje X).

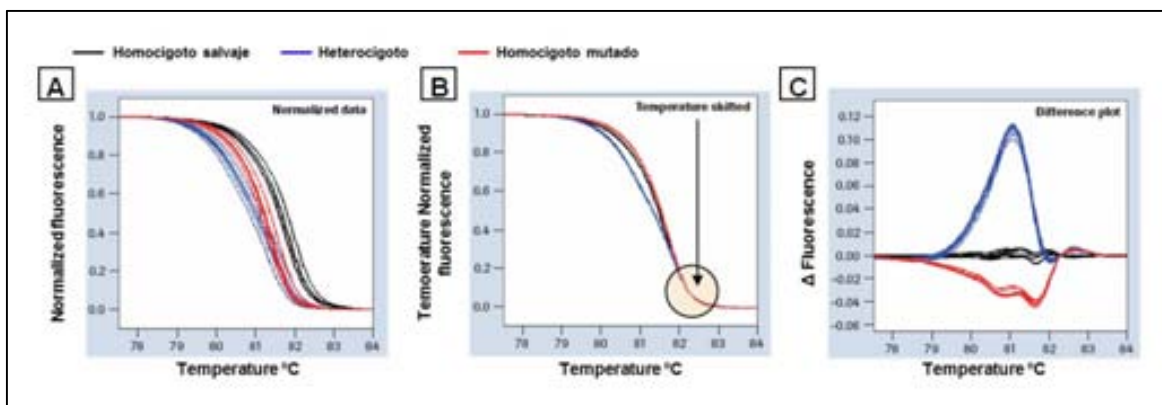


Figura M-7. Análisis de una variante mediante HRM. A. Representación gráfica de la normalización de la señal de fluorescencia de la curva de fusión. B. Representación gráfica de los datos mostrados en A después de la unificación de los puntos de fusión. C. Diferencia de los perfiles de curva de fusión. Se ha tomado como muestra de referencia un homocigoto salvaje.

En el caso de búsqueda de nuevos polimorfismos, como en *MUC1*, cuando se detectó más de un perfil de curva de fusión, correspondiente a una variación en la secuencia, se caracterizó el cambio mediante secuenciación directa de un nuevo producto de PCR. Para el genotipado de la inserción de 3 pb (c.*68_69InsCTT) y el SNP (rs225334) en la región 3'UTR de *TFF2*, se compararon los perfiles de la curva de fusión de las muestras procedentes del estudio de seguimiento de lesiones gástricas con muestras control para cada uno de los haplotipos posibles.

El procedimiento general de análisis de las variantes polimórficas con HRM se describe en el PROTOCOLO XIII y a continuación se detallan las condiciones específicas de amplificación, en base al protocolo general, para cada uno de los fragmentos de interés:

- Las condiciones de amplificación fueron las mismas para todos los fragmentos de *MUC1*: 0,2 μ M de cada cebador, 3 mM $MgCl_2$, 20 ng de DNA inicial, temperatura de anillamiento de 62°C y 15 adquisiciones/°C.
- Para la amplificación del fragmento de 127 pb del exón 4 de *TFF2* se utilizó 0,3 μ M de cada cebador, 2,5 mM $MgCl_2$, 25 ng de DNA inicial y temperatura de anillamiento de 54°C y 15 adquisiciones/°C.

7.4. Genotipado de polimorfismos en plataformas de alto rendimiento

La posibilidad de utilizar las plataformas y servicios del Centro Nacional de Genotipado (CeGen) para el análisis a media/gran escala de polimorfismos a un coste muy inferior a su análisis individual nos permitió ampliar el número de genes y polimorfismos a analizar en nuestros estudios. Por ello, en una segunda fase del estudio se optó por incrementar el número de SNPs seleccionados (ver TABLA R-6 del apartado 1.5 RESULTADOS), que incluían polimorfismos potencialmente funcionales y tagSNPs, y utilizar los servicios que ofrece CeGen para su genotipado.

El genotipado de los polimorfismos para las muestras procedentes del estudio de seguimiento de pacientes con lesiones gástricas preneoplásicas, exceptuando los polimorfismos de *TFFs* y en genes de reparación e inflamación ya genotipados mediante PCR a tiempo real en el LC480 y ABI PRISM 7900HT, se realizó con la tecnología MassArray™ (iPlex-Gold) de Sequenom. Para el genotipado de todos los polimorfismos seleccionados en las muestras procedentes del estudio caso-control de cáncer gástrico se escogió la plataforma Veracode™ de Illumina, ya que el número de muestras a analizar fue considerablemente mayor así como el número de polimorfismos, 33 en *TFFs* y 25 en *MUCs* que formaban parte de un total de 1536 polimorfismos en genes candidatos incluidos en el proyecto “*Susceptibilidad genética, exposición ambiental y riesgo de cáncer gástrico en la población europea*”.

7.4.1. MassArray™ (iPlex-Gold) de Sequenom.

La tecnología de MassArray (Sequenom, <http://www.sequenom.com/ipleX>) se basa en la detección de los productos de la reacción de discriminación alélica mediante espectrometría de masas (MALDI-TOF MS). Existen dos estrategias de genotipado: hME y iplex Gold, siendo esta última la utilizada en nuestro proyecto para genotipar 67 SNPs en el grupo de pacientes con

lesiones gástricas (tagSNPs seleccionados de *TFFs* y *MUCs* y los potencialmente funcionales que no se genotiparon mediante PCR a tiempo real, además de los polimorfismos en genes involucrados en la interacción con *H.pylori*, (ver TABLA R-6 del APARTADO 1.5 de resultados).

Uno de los factores más críticos a tener en cuenta para el genotipado de SNPs en la plataforma de Sequenom es la calidad y concentración de las muestras de DNA. De los 456 pacientes incluidos en el estudio de seguimiento de pacientes con lesiones gástricas, preneoplásicas con muestras de DNA extraídas de sangre periférica o de tejido incluido en parafina, se excluyeron 69 debido a la baja calidad o cantidad del DNA correspondiente. Las 387 muestras restantes se enviaron a CeGen para su genotipado. Las muestras fueron cuantificadas previamente, mediante la lectura de densidad óptica con el espectrofotómetro y diluidas en agua hasta una concentración final de 20 ng/μl. Algunas de las muestras fueron, posteriormente, recuantificadas mediante el sistema de PicoGreen®. Se distribuyeron 50 μl de cada muestra, la cantidad mínima requerida es de 30 μl más 4 μl por ensayo realizado, en placas de 96 pocillos. En todas las placas se dejan 6 pocillos vacíos, donde se incluyen como controles de genotipado dos tríos de muestras analizadas en HapMap y procedentes del Instituto Coriell.

Se incluyeron en las placas un total de 17 muestras duplicadas (4,2% del total) escogidas al azar. Estas muestras de DNA, identificadas con el mismo nombre, se reparten en pocillos escogidos aleatoriamente y actúan como control de calidad adicional de los resultados de genotipado.

7.4.2. Illumina GoldenGate

La tecnología GoldenGate de Illumina son unos ensayos flexibles y pre-optimizados que utilizan una polimerasa y ligasa de DNA para discriminar 96 o desde 384 a 3072 SNPs simultáneamente (http://www.illumina.com/technology/goldengate_genotyping_assay.ilmn) y, por tanto, este proceso que requiere una mayor cantidad de DNA genómico, un mínimo de 250 ng.

Illumina asigna una puntuación (0-1) correspondiente a la probabilidad teórica del éxito del ensayo diseñado para cada polimorfismo. A partir de esta puntuación, se llevó a cabo la selección final de SNPs, intentando sustituir a los polimorfismos con una puntuación menor de 0.8, ya que una probabilidad alta teórica de éxito requiere una puntuación de entre 0.6 y 1, por otros en elevado o completo LD ($r^2=0.8-1$) con tuvieran una puntuación de 0.8 o mayor.

En esta plataforma se analizaron 1536 SNPs, entre los cuales los de *TFFs* y *MUCs* objeto de este trabajo, en un total de 2005 muestras. Se envió 0.75-1 μg de cada muestra, a una concentración aproximada de 50 ng/ μl, distribuidas en placas de 96 pocillos. La concentración de todas las muestras se cuantificó previamente mediante el sistema de PicoGreen®. La primera placa y los pocillos A1, B1, C1 y D1 del resto de placas no contenían muestra ya que estaban destinadas a los controles internos de genotipado incluidos por Cegen. Además de estos controles internos, el 5% de las muestras (n=100) se enviaron por duplicado para evaluar la tasa de error del genotipado. Aunque la calidad del DNA extraído no se evaluó específicamente para este

estudio se realizaron medidas espectrofotométricas de un subgrupo de 75 muestras que mostraron ratios OD260/OD280 de 1.8, indicativo de una buena calidad de DNA.

8. CLONAJE DE DNA DENTRO DE UN PLÁSMIDO

Paralelamente al análisis de asociación entre polimorfismos de *TFFs* y la carcinogénesis gástrica se realizó un estudio funcional, mediante ensayos luciferasa, para dilucidar el efecto de algunos de los polimorfismos potencialmente funcionales identificados. Para llevar a cabo este estudio se seleccionaron los polimorfismos de la región 3'UTR de *TFFs* (*TFF1* c.*41G>A, *TFF2* c.*68_69InsCTT, *TFF2* c.*82T>C (rs225334) y *TFF3* c.*41G>A) y un SNP del promotor de *TFF2* c.-409G>T (rs13052596), ya que estos polimorfismos fueron los primeros en validarse o identificarse en una población control y podían ser testados utilizando la misma estrategia experimental.

El primer paso fue clonar las regiones que contenían cada uno de los polimorfismos en los vectores reporteros apropiados para el ensayo luciferasa que iba a realizarse. Además, los miRNAs previamente seleccionados por tener como posibles dianas a las secuencias polimórficas en las regiones 3'-UTR de los genes *TFF* se clonaron en un vector de expresión específico para este tipo de RNAs.

Para decidir la estrategia de clonaje, en cada caso, se tuvo en cuenta los siguientes factores:

- Elección del vector plasmídico adecuado para la aplicación en la que se va a utilizar.
- Elección de los sitios de restricción en el vector.
- Preparación del inserto y el vector.
- Condiciones óptimas para la ligación.
- Metodología a utilizar para validar los clones de interés.

En este trabajo se realizaron 3 tipos de clonaje:

- Clonación mediante extremos cohesivos simétricos de la región 3'UTR de los 3 genes *TFF* en un vector que contenía el gen de la luciferasa regulado por un promotor constitutivo.
- Clonación direccional en un vector de expresión de los genes que codificaban para los miRNA seleccionados previamente.

8.1. Vectores utilizados en la clonación

Las regiones 3'UTR de *TFFs* se clonaron en un vector **pGL4.13** (Promega). Los vectores reporteros pGL4 son equivalentes a los pGL3, pero se trata de una nueva generación, también optimizada para su expresión en células de mamífero. Se escogió el vector pGL4.13, ya que contiene el promotor constitutivo SV40 (*simian virus-40*) y, en este caso, la finalidad era testar el efecto modulador de las regiones 3'UTR de *TFFs* en la expresión del gen. Para clonar los fragmentos correspondientes a la región 3'UTR se escogió la diana de restricción de la enzima XbaI, localizada en la posición 2156 del vector.

Los miRNAs seleccionados se clonaron en el vector de expresión **pSilencer 4.1-CMV puro** (pSIL) de Ambion. Estos vectores contienen un promotor modificado CMV (*Cytomegalovirus*), del tipo polimerasa de RNA II, para dirigir la expresión en una gran variedad de tipos celulares de los miRNAs clonados y una señal de poliadenilación SV40 modificada para finalizar la transcripción de los miRNAs. También incluyen el gen de resistencia a puromicina que ofrece un mecanismo para seleccionar las células transfectadas que expresan el DNA introducido. Para la clonación se escogieron las dianas para las enzimas BamHI y HindIII, ya que son las únicas presentes en el vector que pueden utilizarse para liberar, mediante una reacción de digestión, la secuencia de control negativo que contiene el vector pSIL utilizado en este trabajo.

8.2. Obtención del inserto y del vector

Diseño de oligonucleótidos y amplificación por PCR: Todos los fragmentos de DNA a clonar en los vectores se obtuvieron mediante amplificación por PCR de la secuencia de interés. Como se ha descrito anteriormente, uno de los factores críticos en la amplificación por PCR es el diseño de los cebadores a utilizar en la reacción, especialmente cuando el producto de la amplificación va a ser clonado. Para diseñar los cebadores se siguieron los criterios especificados anteriormente en el APARTADO 4 de Materiales y Métodos, pero además se añadió a 5' del oligonucleótido la secuencia diana para la enzima de restricción que se iba a utilizar en la preparación del inserto mediante digestión. Debido a que, de esta manera, la secuencia diana de la enzima se localizaría muy próxima a los extremos del DNA amplificado y dificultaría la unión de la enzima con su secuencia impidiendo el corte, se añadió también una secuencia de anclaje en el extremo 5' de la secuencia diana. Esta secuencia consiste en 3-6 bases no apareadas que permiten la unión más estable de la enzima con su secuencia. La utilización de este tipo de cebadores permitió crear sitios de restricción en los extremos del inserto durante la amplificación, para poder ser digeridos con las enzimas de restricción apropiadas y generar extremos cohesivos. La secuencia completa de todos los cebadores utilizados y la región que amplifican en cada caso se describen en la TABLA M-7. Las muestras de DNA utilizadas fueron muestras control, previamente genotipadas para cada uno de los polimorfismos en la fase previa de validación y búsqueda de nuevos polimorfismos, cuyo genotipo fuera homocigoto para el alelo más común. Los cebadores, tanto directo como reverso, para amplificar la región 3'UTR de los genes *TFF* se adaptaron añadiendo una diana para la enzima de restricción XbaI, que también está contenida en el vector receptor pGL4, a 3' del gen reportero *luc2* y a 5' de la señal de poliadenilación SV40. Los fragmentos amplificados con estos cebadores abarcan desde el nucleótido 1 de la región 3'UTR hasta el último nucleótido de la secuencia del transcrito de los genes *TFF1*, *TFF2* y *TFF3*. Para la amplificación de miRNAs se utilizaron cebadores diseñados específicamente para introducir las dianas de restricción para las enzimas BamHI, en el cebador directo, y HindIII, en el cebador reverso. Las dianas se encuentran en las posiciones 516 y 463, respectivamente, del vector receptor pSIL. La secuencia específica de cada miRNA seleccionado se buscó en el miRBase y se confirmó en el NCBI *nucleotide database*. El fragmento amplificado corresponde a la secuencia del pre-miRNA (o *hairpin*), más 50 pb, aproximadamente, a 5' y 3'.

Tabla M-7. Cebadores utilizados en la clonación de la región 3'UTR de TFFs y de miRNAs y del promotor de TFF2

Gen	ID (NCBI)	Cr.	nº acceso (GenBank)	Posición	Cebador	Enzima	Secuencia ^a	pb ^b
<i>TFF1</i>	7031	21	NM_003225	298-316	TFF1 3'UTR F	XbaI	5'-ACACACT <u>TCTAGA</u> ACTTCTGCAGGGATCTGCC-3'	193+24
<i>TFF1</i>	7031	21	NM_003225	466-490	TFF1 3'UTR R	XbaI	5'-ACACACT <u>TCTAGA</u> AATATCGATCTCTTTAATTTTAG-3'	
<i>TFF2</i>	7032	21	NM_005423	533-550	TFF2 3'UTR F	XbaI	5'-ACACACT <u>TCTAGA</u> GAGAGGCTGGTTCCAGAG-3'	176+24
<i>TFF2</i>	7032	21	NM_005423	688-708	TFF2 3'UTR R	XbaI	5'-ACACACT <u>TCTAGA</u> TTTAGCAGATTTTAAGGGTTT-3'	
<i>TFF3</i>	7033	21	NM_003226	466-481	TFF3 3'UTR F	XbaI	5'-ACACACT <u>TCTAGA</u> GGCACCTCCAGCTGCC-3'	226+24
<i>TFF3</i>	7033	21	NM_003226	650-667	TFF3 3'UTR R	XbaI	5'-ACACACT <u>TCTAGA</u> CAGGCACGAAGAACTGTC-3'	
<i>hsa-miR-127</i>	406914	14	NT_026437.12	82349265-82349283	miR-127 F	BglII	5'-AAAAA <u>AGATCT</u> CACCCACAGGTTCCCAAGG-3'	203+22
<i>hsa-miR-127</i>	406914	14	NT_026437.12	82349449-82349467	miR-127 R	HindIII	5'-AAAAA <u>AAGCTT</u> CGGACTGCCACACCCATAC-3'	
<i>hsa-miR-92b</i>	693235	1	NT_004487.19	6653538-6653556	miR-92b F	BglII	5'-AAAAA <u>AGATCT</u> GGTGAACCCAACTCCCCAG-3'	224+22
<i>hsa-miR-92b</i>	693235	1	NT_004487.19	6653761-6653743	miR-92b R	HindIII	5'-AAAAA <u>AAGCTT</u> GGATGGAGAGCCAGGAGGG-3'	
<i>hsa-miR-615</i>	693200	12	NT_029419.12	16570974-16570991	miR-615 F	BamHI	5'-AAAAA <u>GGATCC</u> CTCCCTCCCTCCCTCCC-3'	220+22
<i>hsa-miR-615</i>	693200	12	NT_029419.12	16571176-16571193	miR-615 R	HindIII	5'-AAAAA <u>AAGCTT</u> TCCGCCAGGCAGGAAAC-3'	
<i>hsa-miR-671</i>	768213	7	NT_007914.15	11531071-11531089	miR-671 F	BamHI	5'-AAAAA <u>GGATCC</u> AGGCTTCTGCGGAGGGCTG -3'	232+22
<i>hsa-miR-671</i>	768213	7	NT_007914.15	11531282-11531302	miR-671 R	HindIII	5'-AAAAA <u>AAGCTT</u> CGGTGGTAGAGTTCTTCACTG -3'	
<i>hsa-miR-760</i>	100126348	1	NT_032977.9	64284247-64284265	miR-760 F	BamHI	5'-AAAAA <u>GGATCC</u> GCGTCCTTGAGGCAGGTG-3'	197+22
<i>hsa-miR-760</i>	100126348	1	NT_032977.9	64284424-64284443	miR-760 R	HindIII	5'-AAAAA <u>AAGCTT</u> CGTCTGTGTGCGTGCTTGTG-3'	
<i>hsa-miR-524</i>	574478	19	NT_011109.16	26482420-26482440	miR-524 F	BamHI	5'-AAAAA <u>GGATCC</u> TAACCTGGGCAAGGAAAATTC-3'	189+22
<i>hsa-miR-524</i>	574478	19	NT_011109.16	26482587-26482608	miR-524 R	HindIII	5'-AAAAA <u>AAGCTT</u> GCTCTACAAATGTATTACCGAG -3'	
<i>hsa-miR-652</i>	724022	X	NT_011651.17	32594806-32594827	miR-652 F	BamHI	5'-AAAAA <u>GGATCC</u> TTGAGTGTCTTTTTGGCTTTG -3'	240+22
<i>hsa-miR-652</i>	724022	X	NT_011651.17	32595026-32595045	miR-652 R	HindIII	5'-AAAAA <u>AAGCTT</u> CTTACCCTGAGGACAATCCC -3'	

a, En negrita y subrayado la secuencia del sitio de restricción introducida en el cebador; b, Tamaño del fragmento amplificado más el número de bases inespecíficas introducidas con el cebador.

Debido a que una de las dianas escogidas, BamHI, se encontraba ya presente en la secuencia de dos de los miRNAs seleccionados se diseñó una estrategia alternativa para su clonación. En este caso el cebador directo se diseñó para incorporar la diana de restricción de una de las enzimas ya disponibles en nuestro laboratorio, BglII. Paralelamente se diseñaron dos oligonucleótidos complementarios que pudieran adaptarse e hibridarse con el fragmento digerido por BglII y que, además, crearan un extremo cohesivo equivalente al generado por la digestión con la enzima BamHI, ya que esta enzima es la que se utilizaría para digerir el vector pSIL. La secuencia de estos adaptadores es la siguiente:

Adapt BamHI-BglII F: 5'-GATCCTCTTGTGACCTTGTGTA-3

Adapt BamHI-BglII R: 5'-GATCTACACAAGGTCACAAGAG-3'

Purificación del producto de PCR: El producto amplificado por PCR se purificó mediante kits específicos para este fin (APARTADO 4.1 y 6.1.1 de Materiales y Métodos). La presencia de polimerasa y reactivos residuales del proceso de amplificación en el producto de PCR, frecuentemente interfiere en los métodos subsiguientes de adaptación de los extremos del DNA amplificado para su clonación. Por esta razón la purificación del producto de PCR es un paso imprescindible previo a la digestión con enzimas de restricción.

Se comprobó la identidad de todos los fragmentos amplificados y purificados mediante **secuenciación directa** con cebadores específicos de para cada fragmento.

8.3. Clonaje mediante extremos cohesivos

Los productos amplificados por PCR y posteriormente purificados correspondientes a la región 3'UTR de *TFFs* y miRNAs, se sometieron a digestión con enzimas de restricción, sencilla o doble (PROTOCOLO IX), para convertir los extremos de estos fragmentos en cohesivos para su posterior clonación. Los vectores receptores, pGL4 y pSIL, también se digirieron con las mismas enzimas.

8.3.1. Extremos cohesivos generados por digestión con una sola enzima de restricción

En el caso de las regiones 3'UTR, los extremos cohesivos se crearon digiriendo tanto vector como inserto con una única enzima de restricción que cortaba asimétricamente en la secuencia diana. De esta manera, los extremos protuberantes de vector e inserto son compatibles y forman una molécula híbrida que, por mediación de una ligasa, se recirculará produciendo el plásmido recombinante. Los plásmidos circulares recombinantes que se obtienen con esta estrategia pueden contener el inserto de DNA en cualquier orientación. Tanto el vector pGL4.13 como el producto amplificado por PCR y purificado, correspondientes a la región 3'UTR, se digirieron con la enzima de restricción XbaI (Takara). Los fragmentos digeridos se visualizaron en un gel de agarosa al 2%, en presencia de un control no digerido, y se recuperaron por extracción de banda.

Cuando se utilizaron plásmidos con extremos cohesivos generados por digestión con una única enzima se eliminaron los grupos fosfato del extremo 5'. La defosforilación del vector

receptor, PROTOCOLO XIV, tuvo como objetivo evitar la autoligación y recircularización del plásmido y disminuir el número de colonias bacterianas transformadas que contuvieran plásmido vacío. Los vectores desfosforilados se purificaron por extracción de banda a partir de gel de agarosa para su posterior utilización en la reacción de ligación.

8.3.2. Clonaje direccional

Una manera de aumentar la cantidad de plásmidos recombinantes monoméricos, es decir, que contengan una única copia del inserto, es utilizar una estrategia de clonación en la que los extremos de las moléculas de DNA en la reacción de ligación no sean equivalentes. Esto puede conseguirse digiriendo el inserto con dos enzimas de restricción que reconozcan secuencias diana diferentes, evitando que se ligen diferentes insertos en tándem, y digiriendo el vector receptor con las mismas enzimas. Una de las ventajas de esta estrategia es que la orientación del inserto en el plásmido está predefinida. Al preparar el plásmido para el clonaje direccional mediante esta estrategia se debe evitar utilizar enzimas de restricción que corten a menos de 12 pb entre ellas, ya que cuando uno de los sitios de restricción haya sido cortado, el segundo sitio quedará localizado demasiado próximo al extremo del DNA ya linearizado para permitir el corte por la segunda enzima. Se debe comprobar previamente que las dos enzimas tengan una actividad óptima en el mismo tampón de digestión. Si es así, la digestión del vector y el fragmento de DNA puede llevarse a cabo simultáneamente con ambas enzimas. Si las dos enzimas de restricción requieren tampones o condiciones diferentes, como fue nuestro caso, es mejor llevar a cabo las digestiones secuencialmente (ver PROTOCOLO IX, digestión doble). Tanto los fragmentos correspondientes a miRNAs amplificados por PCR mediante cebadores adaptados específicamente, como el vector pSIL, se digirieron con dos enzimas de restricción, BamHI y HindIII, que cortaban en secuencias diferentes y, consecuentemente, generaban extremos cohesivos diferentes. En el diseño de las secuencias dianas añadidas al cebador se tuvo en cuenta la orientación que se le quería dar al inserto.

Una vez digeridos, tanto inserto como vector, se separaron mediante electroforesis en gel de agarosa y se recuperó el DNA mediante purificación por extracción de banda. En la electroforesis se incluyeron siempre alícuotas del plásmido e inserto no digerido y alícuotas de plásmido e inserto digeridos con una única enzima.

8.4. Ligación

Cada inserto y vector receptor correspondiente, previamente digeridos y purificados, se incubaron en presencia de una enzima ligasa en una reacción de ligación para generar los plásmidos recombinantes. La ligasa de DNA cataliza la unión de dos hebras de DNA entre los grupos 5'-fosfato y 3'-hidroxilo de nucleótidos adyacentes. La enzima utilizada para llevar a cabo la reacción fue una ligasa derivada del bacteriófago T4.

La concentración del producto de PCR y del vector digerido debe calcularse previamente a la reacción de ligación. La cantidad de DNA presente se estima mediante electroforesis analítica en un gel de agarosa comparando la intensidad de las bandas del fragmento de interés

con las bandas de un marcador estándar de concentración conocida. En el caso del vector pGEM-T comercial, éste viene suministrado a una concentración de 50 ng/ μ l.

Todas las reacciones de ligación fueron optimizadas para el ratio inserto:vector molar. El rango general de este ratio está entre 2:1 y 6:1. En el caso del vector pGEM se utilizó un ratio inserto:vector de 3:1, y en el resto de ligaciones se utilizó un ratio 6:1. Para el cálculo de la cantidad apropiada de producto de PCR se aplicó la ecuación detallada en el PROTOCOLO XV, empleando siempre 50 ng del vector. En este protocolo se describe el procedimiento general para realizar las reacciones de ligación.

9. TRANSFORMACIÓN DE CÉLULAS COMPETENTES

La transformación es el proceso por el cual se introduce DNA plasmídico foráneo en células bacterianas competentes. El estado de competencia en bacterias que no lo presentan de forma natural, como es el caso de *E. coli*, puede ser inducido en el laboratorio mediante técnicas que se basan en diversos tratamientos químicos o físicos. Estos tratamientos producen microporos en la célula, que permiten la introducción del DNA exógeno (transformación) de modo eficiente. En este trabajo se utilizó el método del cloruro cálcico, inicialmente descrito por S.N Cohen (Cohen et al. 1972) para preparar células competentes de *E. coli* (PROTOCOLO XVI). La transformación con DNA procedente de una ligación o de una preparación plásmidica se realizó mediante choque térmico utilizando la cepa comercial JM109 (Promega) o las células competentes de *E. coli* ($\geq 1 \times 10^8$ cfu/ μ g DNA) preparadas manualmente.

El PROTOCOLO XVII describe el proceso de transformación con DNA obtenido en una reacción de ligación, que incluye la preparación de placas para la siembra de los controles de la reacción de ligación. Para la transformación con plásmidos procedentes de preparaciones plasmídicas, el protocolo es básicamente el mismo pero, en este caso, se transforman 50 μ l de células competentes con 1-20 ng de plásmido en y sólo se preparan 2 placas por transformación, donde se siembra en cada una de ellas 1/10 de las bacterias transformadas.

10. COMPROBACIÓN DE LOS PLÁSMIDOS RECOMBINANTES

La caracterización de las colonias que habían incorporado el DNA plasmídico de interés se realizó mediante PCR de cribaje para determinar que colonias contenían el inserto y las minipreparaciones plasmídicas (ver APARTADO 12 de Materiales y Métodos) de las colonias que resultaron positivas se digirieron para liberarlo. En los casos en que se determinó la presencia del inserto, se realizó la secuenciación del fragmento de interés para confirmar que el clonaje se había realizado correctamente y que la secuencia no presentase cambios respecto a la original. Se realizaron midi y maxipreparaciones de las construcciones correctas para realizar los ensayos funcionales posteriores.

La PCR de cribaje (PROTOCOLO XVIII) se basa en realizar una reacción de PCR, utilizando las colonias transformadas, en la que se usan una pareja de cebadores que correspondan a la región del vector flanqueante al inserto. Si queremos saber si el fragmento de DNA que hemos

introducido dentro del vector se ha insertado en una dirección concreta, se utiliza una pareja de cebadores en la que uno de ellos corresponda al vector y el otro, al fragmento que hemos clonado. En el caso de la PCR de cribaje de los insertos clonados a 3' del vector pGL4 los cebadores utilizados fueron los mismos que en la clonación. En la TABLA M-8 se detallan los cebadores utilizados para la comprobación de los plásmidos recombinantes y en la TABLA M-9 se describe la secuencia de los cebadores específicos de los vectores utilizados en este trabajo.

Tabla M-8. Cebadores utilizados para la comprobación de los plásmidos recombinantes en la clonación de la región 3'UTR de *TFFs*, miRNAs y el promotor de *TFF2*.

Plásmidos	Características	Tamaño	Cebadores secuenciación ^a		Enzima
			Directo	Reverso	
pGL4.13	SV40/luc2	4641	RVprimer3		
pSIL puro	CMV	4781	pSIL F	pSIL R	
pGL4-TFF141G	SV40/luc2-TFF1 3'UTR	4840	RVprimer3	TFF1 3'UTR R	XbaI
pGL4-TFF141A	SV40/luc2-TFF1 3'UTR	4840	RVprimer3	TFF1 3'UTR R	XbaI
pGL4-TFF282C	SV40/luc2-TFF2 3'UTR	4823	RVprimer3	TFF2 3'UTR R	XbaI
pGL4-TFF282T	SV40/luc2-TFF2 3'UTR	4823	RVprimer3	TFF2 3'UTR R	XbaI
pGL4-TFF282C-CTT	SV40/luc2-TFF2 3'UTR	4823	RVprimer3	TFF2 3'UTR R	XbaI
pGL4-TFF282T-CTT	SV40/luc2-TFF2 3'UTR	4823	RVprimer3	TFF2 3'UTR R	XbaI
pGL4-TFF341G	SV40/luc2-TFF3 3'UTR	4873	RVprimer3	TFF3 3'UTR R	XbaI
pGL4-TFF341A	SV40/luc2-TFF3 3'UTR	4873	RVprimer3	TFF3 3'UTR R	XbaI
pSIL-miR615	CMV/ miR615	4954	pSIL F	pSIL R	BamHI/HindIII
pSIL-miR671	CMV/ miR671	4966	pSIL F	pSIL R	BamHI/HindIII
pSIL-miR760	CMV/ miR760	4931	pSIL F	pSIL R	BamHI/HindIII
pSIL-miR652	CMV/miR652	4974	pSIL F	pSIL R	BamHI/HindIII
pSIL-miR127	CMV/miR127	4937	pSIL F	pSIL R	BglII/HindIII
pSIL-miR92b	CMV/miR92b	4958	pSIL F	pSIL R	BglII/HindIII
pSIL-miR524	CMV/miR524	4923	pSIL F	pSIL R	BamHI/HindIII

a, La secuencias de estos cebadores se encuentran en las tablas M-7 (secuencias 3'UTR de *TFFs*) y M-9 (secuencias en los vectores usados).

Tabla M-9. Cebadores específicos para los vectores utilizados en este trabajo

Cebadores	Vector	Secuencia
RVprimer3	pGL4	5'-CTAGCAAAATAGGCTGTCCC-3'
pSIL F	pSilencer 4.1	5'-AGGCGATTAAGTTGGGTA-3'
pSIL R	pSilencer 4.1	5'-CGGTAGGCGGTACGGTG-3'

11. MUTAGÉNESIS DIRIGIDA

Todos los vectores reporteros en los que se había clonado fragmentos de *TFFs* (regiones 3'UTR y promotor de *TFF2*), se utilizaron como molde para introducir los cambios deseados, que correspondían a diferentes polimorfismos identificados en estas regiones, mediante mutagénesis dirigida. Este proceso consiste en introducir mutaciones puntuales o pequeñas inserciones o

deleciones en los genes o fragmentos clonados previamente en plásmidos. Para ello se utilizaron dos cebadores, directo y reverso, complementarios a la región a mutagenizar y que contienen el cambio deseado. La utilización de este tipo de cebadores durante la amplificación del plásmido, permite transferir el cambio a las nuevas cadenas amplificadas. En la reacción de amplificación se usó la polimerasa Pfu caracterizada por la alta fidelidad en la replicación del DNA, que junto con el bajo número de ciclos de replicación, favorecen una elevada eficiencia de incorporación de la mutación específica con una baja tasa de mutaciones al azar. Para eliminar los plásmidos parentales, y seleccionar sólo los mutados, se trató el producto de la reacción con la endonucleasa DpnI, que reconoce y digiere el DNA metilado y hemimetilado producido por bacterias como *E.coli*, que ha sido la utilizada en este trabajo para transformar con los plásmidos recombinantes originales.

El producto amplificado se transformó por choque térmico en células competentes y se realizaron minipreparaciones plasmídicas de diversas colonias para comprobar, posteriormente, el resultado de la mutagénesis mediante secuenciación directa del inserto. Con la finalidad de obtener la cantidad suficiente de DNA plasmídico para realizar los ensayos posteriores, se realizaron maxi y midipreparaciones de los plásmidos que contenían el cambio deseado.

En el proceso de mutagénesis dirigida, uno de los factores clave es el diseño de los cebadores utilizados en la amplificación, que presentan características diferentes a los diseñados para PCR convencionales. Se tuvo en cuenta las siguientes consideraciones:

- Cada uno de los cebadores (tanto directo como reverso) debe contener la mutación deseada e hibridar en la misma secuencia del plásmido, tanto en la cadena + como en la cadena.
- Deben tener un tamaño entre 25-45 pb y la T_m debe ser mayor o igual que 78°C. Para calcularla debemos aplicar la siguiente fórmula, donde N es el número de nucleótidos del cebador:
 - ♦ En cambio de nucleótidos: $T_m = 81,5 + 0,41(\%GC) - 675/N - (\%mismatch)$
 - ♦ En deleciones o inserciones: $T_m = 81,5 + 0,41(\%GC) - 675/N$
- La mutación deseada debe localizarse en el centro de cada cebador, con 10-15 bases flanqueantes a cada lado.
- Los cebadores deben tener un mínimo del 40% GC y deberían poseer una o más Cs o Gs en su extremo 3'.
- Los cebadores no deben estar fosforilados en su extremo 5' y deben haber sido purificados por FPLC o PAGE.

En la TABLA M-10 se describen todos los cebadores utilizados en el proceso de mutagénesis, los plásmidos originales que sirvieron como molde y el cambio que se pretendía introducir. Para la mutagénesis se utilizó el kit comercial *QuikChange® II Site-Directed Mutagenesis Kit* (Stratagene) siguiendo las instrucciones recomendadas por el fabricante o, alternativamente, se utilizó el procedimiento indicado en el PROTOCOLO XIX.

Tabla M-10. Cebadores diseñados para los experimentos de mutagénesis dirigida

Plásmido ^a	Secuencia cebadores mutagénesis ^b	Cambio
pGL4-TFF141G	Directo: 5'-GACGCGGTGCCATCCCCAGCACG-3' Reverso: 5'-CGTGCTGGGGATGGCACC GCGTC-3'	c.*41G>A
pGL4-TFF282C	Directo: 5'-GAAGAACTTCGCCTTCTTATCAGCTTCATACTTC-3' Reverso: 5'-GAAGTATGAAGCTGATAAGAAGGCGAAGTTTCTTC-3'	c.*67_68InsCTT
pGL4-TFF282C-CTT	Directo: 5'-CTTCGCCTTCTTATCAGCTTCATATTTTCATGAAATCCTGGG-3' Reverso: 5'-CCCAGGATTTTCATGAAATATGAAGCTGATAAGAAGGCGAAG-3'	c.*82C>T
pGL4-TFF282C	Directo: 5'-CTTCGCCTTATCAGCTTCATATTTTCATGAAATCCTGGG-3' Reverso: 5'-CCCAGGATTTTCATGAAATATGAAGCTGATAAGAAGGCGAAG-3'	c.*82C>T
pGL4-TFF341G	Directo: 5'-GATGCGAGGCTCGAAGCACCTTGCCC-3' Reverso: 5'-GGGCAAGGGTGCTTCGAGCCTCGCATC-3'	c.*41G>A

a, plásmido original que sirvió como molde en la reacción de mutagénesis; b, en negrita está indicada la base mutada.

12. EXTRACCIÓN DEL DNA PLASMÍDICO DE BACTERIAS

La extracción de DNA plasmídico de células bacterianas se realizó con el sistema comercial de Qiagen, en sus diferentes formatos *QIAprep Spin Miniprep kit*, *QIAGEN Plasmid Midi kit* y *QIAGEN Plasmid Maxi kit*. El procedimiento para la purificación de plásmidos que utilizan estos kits se basa un método modificado de lisis alcalina desarrollado por Birnboim y Doly (Birnboim&Doly 1979). Las células son lisadas en presencia de un tampón que contiene NaOH, SDS (*Sodium dodecyl sulfate*) y RNasa A. El detergente SDS solubiliza los fosfolípidos y proteínas que componen la membrana celular, provocando la lisis y liberando el contenido celular, mientras que las condiciones alcalinas desnaturalizan el DNA plasmídico y cromosómico y las proteínas. Posteriormente la solución es neutralizada por adición de acetato de potasio, el DNA cromosómico y las proteínas precipitan debido a que es imposible que renaturalicen correctamente, por su gran tamaño, mientras que los plásmidos, más pequeños, renaturalizan correctamente quedando en solución. Para cuantificar y comprobar la calidad de todas las preparaciones plasmídicas, se visualizaron en geles de agarosa al 1% y se realizó la lectura de la absorbancia (A_{600}) en el espectrofotómetro. Aquellas que contenían el inserto de interés, se digirieron con las enzimas correspondientes (ver TABLA M-7) para liberar el inserto y se secuenciaron.

12.1. Minipreparación

Las Minipreparaciones se realizaron con el *QIAprep Spin Miniprep kit* (Qiagen, PROTOCOLO XX). En este caso, el lisado de células es neutralizado y, simultáneamente, ajustado a una alta concentración de sales para permitir la posterior adsorción del DNA en una membrana de sílica. Esta membrana, en combinación con los tampones utilizados en la lisis, permite que únicamente el DNA sea adsorbido, mientras que el RNA, proteínas celulares y otros contaminantes no son retenidos. Después de realizar los lavados del DNA retenido en la membrana, se eluye en condiciones de concentración baja de sales.

12.2. Midi y Maxipreparación

Las midis y maxipreparaciones se llevaron a cabo con el *QIAGEN Plasmid Midi kit* y *QIAGEN Plasmid Maxi kit*, respectivamente. Este tipo de preparaciones permiten obtener una mayor cantidad de DNA plasmídico, indispensable para los posteriores ensayos funcionales. El protocolo de purificación de plásmidos de estos kits se basa en el método modificado de lisis alcalina y la posterior unión del plásmido a una resina de intercambio de aniones, en condiciones específicas de pH y baja concentración de sal. La resina está dispuesta en columnas diseñadas para funcionar por flujo de gravedad. Las proteínas, el RNA y otras impurezas de bajo peso molecular se eliminan mediante lavados moderadamente salinos. Finalmente el DNA es eluído en un tampón altamente salino y, posteriormente, precipitado con isopropanol para concentrarlo y eliminar las sales.

El protocolo seguido para realizar las midi y maxipreparaciones ha sido el mismo en ambos casos (PROTOCOLO XXI), diferenciándose únicamente en el volumen de los reactivos utilizados en cada tipo de preparación plasmídica.

13. ENSAYOS CON LÍNEAS CELULARES

Se realizaron ensayos luciferasa en líneas celulares para comprobar si los polimorfismos en 3'UTR de *TFFs* modificaban la expresión de la luciferasa o alteraban la interacción con miRNAs. Para ello se transfectaron las líneas HEK-293T y AGS con cada uno de los vectores pGL4 que contenían las regiones 3'UTR con los diferentes alelos de los polimorfismos a testar. También se contranfectaron, en la línea HEK-293T, las construcciones anteriormente descritas con el vector de expresión pSIL que contenía cada uno de los miRNAs predichos a interactuar.

13.1. Cultivo y mantenimiento de líneas celulares

Las líneas celulares escogidas para realizar los ensayos luciferasa fueron las **HEK-293T**, una línea celular altamente transfectable derivada de la línea celular HEK-293 (*Human Embryonic Kidney 293*), que expresa constitutivamente el antígeno T de SV40, y las **AGS**, una línea de células epiteliales adherentes derivada de un adenocarcinoma gástrico. Las células se cultivaron en placas de 100 mm (Iwaki) con *Dulbecco's Modified Eagle Medium* (DMEM) (Gibco BRL) para las células HEK293-T o con RPMI 1640 para las células AGS. En ambos casos, el medio se suplementó con 10% de suero fetal bovino inactivado por calor (FBS, Gibco BRL) y 100 U/ml penicilina y 100 µg/ml estreptomycin. A las 48-72 horas se obtuvieron cultivos subconfluentes y se realizaron los pases de mantenimiento correspondientes, desprendiendo las células de las placas mediante tratamiento con tripsina y subcultivándolas a la dilución apropiada en cada caso (1/5-1/10 para HEK-293T y 1/3-1/8 para AGS), tal y como se describe en el PROTOCOLO XXII. Alternativamente, las células se subcultivaron en placas de 24 pocillos para llevar a cabo transfecciones transitorias. Para estandarizar las condiciones de cultivo, para su mantenimiento o transfección, se determinó previamente el número de células de los cultivos utilizando un hemacitómetro Neubauer (ver PROTOCOLO XXII). Todas las incubaciones de los cultivos se

realizaron en un incubador humificado a 37°C y 5% CO₂, siendo la concentración de CO₂ muy importante para controlar el pH del cultivo.

Las líneas celulares se congelaron esporádicamente (PROTOCOLO XXIII) para su conservación a largo plazo, reducir el riesgo de contaminación y minimizar la acumulación de cambios genéticos. Para la congelación se utilizó células crecidas hasta un 70-80% de confluencia (fase logarítmica del cultivo) y un agente de congelación crioprotector, el dimetilsulfóxido (DMSO), sin el cuál el proceso sería letal para las células. La utilización de un contenedor de congelación (Nalgene) permite la congelación gradual de las células, con un descenso de aproximadamente 1°C/min. Cuando las células alcanzaron un número elevado de pasajes se descartaron y se descongelaron viales de células criopreservadas (PROTOCOLO XXIII). Las células deben descongelarse rápidamente y plaquearse a alta densidad para garantizar su crecimiento.

13.2. Transfecciones transitorias

Para optimizar las condiciones experimentales de las transfecciones, se utilizó un vector de expresión que contenía el gen de la proteína verde (*Green Fluorescent Protein*, GFP) y diversos reactivos de lipofectamina (Invitrogen), un lípido catiónico que media la entrada del DNA en las células. Estos lípidos interactúan con el esqueleto de fosfatos de los ácidos nucleicos y los condensan, formando una estructura liposomal con la superficie cargada positivamente. Estos liposomas también median la interacción del ácido nucleico con la membrana celular, permitiendo la fusión del complejo liposoma/ácido nucleico con la membrana celular cargada negativamente, produciéndose la endocitosis del complejo en la célula. El reactivo que se utilizó para todas las transfecciones, tanto en células HEK-293T como AGS, fue la lipofectamina 2000 de Invitrogen (PROTOCOLO XXIV)

Las transfecciones se realizaron en presencia del medio apropiado para cada tipo celular, sin suero, ya que contiene proteínas que podrían interferir en la formación de los complejos, ni antibióticos, que podrían ser citotóxicos debido a la mayor permeabilidad de la membrana en presencia de los lípidos catiónicos. Dos factores importantes que se tuvieron en cuenta fueron la densidad celular de las placas donde se iba a realizar la transfección y la calidad y cantidad del DNA a transfectar, utilizando siempre plásmidos intactos y libres de impurezas. Para realizar los ensayos de transfección las células HEK-293T y AGS se subcultivaron en placas de 24 pocillos en las condiciones anteriores, utilizando 200000 o 250000 células por pocillo, respectivamente. Cuando las placas alcanzaron la subconfluencia, a las 24 horas, se utilizaron para los experimentos de transfección.

La cantidad total de DNA transfectado fue en todos los experimentos de 810 ng. La cantidad de DNA total se ajustó, cuando fue necesario, con el vector pGL3 sin inserto. Se realizaron diversos bloques de transfecciones para los ensayos luciferasa:

- un bloque en el que se transfectó de manera independiente cada una de las construcciones que contenían las regiones 3'UTR de *TFFs*, a 3' del gen luciferasa, con los distintos alelos para los polimorfismos deseados, tanto en células HEK-293T como AGS. Se transfectaron 200 ng de cada construcción.

- un bloque en el que se cotransfectó en células HEK-293T cada una de las construcciones anteriores con el vector pSIL que expresaba el miRNA predicho a interactuar con cada región. Para cada condición se incluyó un control negativo en el que se cotransfectó cada una de las construcciones que contenían la región 3'UTR de *TFFs* con un vector pSIL vacío. Se transfectaron 200 ng/pocillo de la construcción que contenía la región 3'UTR y 600 ng/pocillo del vector que expresaba el miRNA o pSIL vacío. En el caso de miR-524 también se utilizó el miRIDIAN Mimic C-0300807-005 de Ambion cotransfectando 40 pmoles del mimíc o del control negativo (miRIDIAN microRNA Mimic Negative control #2, Ambion).

Todos los pocillos de las placas utilizadas en estos ensayos se cotransfectaron con 10 ng (1/20 de la cantidad del vector de luciferasa de luciérnaga) del vector que expresa la luciferasa *Renilla* (pGL4.74 [CMV/*hRluc*] de Promega). Este vector, que expresa constitutivamente la *Renilla*, sirve como control interno para normalizar la expresión del gen reportero luciferasa de luciérnaga. En todas las tandas de transfección se añadió como control positivo en pocillos independientes, el vector pGL4 intacto, que expresa constitutivamente la luciferasa de luciérnaga, transfectándolo o cotransfectándolo en las mismas condiciones que el resto de pocillos. Cada condición experimental se realizó por triplicado, para aumentar la precisión y reducir la variabilidad intraensayo, y todos los ensayos se realizaron, como mínimo, tres veces en días diferentes.

13.3. Ensayos luciferasa

Los ensayos luciferasa permiten cuantificar la expresión del gen luciferasa en células transfectadas con este gen. De esta manera, se puede analizar la expresión de los promotores clonados a 5' del gen reportero o el efecto modulador de las regiones 3'UTR, clonadas a 3'. En este trabajo, esta técnica se utilizó para testar si los polimorfismos en 3'UTR de *TFFs* modificaban la expresión de la luciferasa, por si sólo o debido a la interacción con miRNAs. Para llevar a cabo los ensayos luciferasa se utilizaron las placas transfectadas transitoriamente tal y como se describe en el APARTADO 13.2 de Materiales y Métodos.

Cuando la luciferasa *Firefly*, producida por el gen sintético contenido en el vector reportero donde se clonaron las regiones 3'UTR, interacciona con su sustrato luciferina se desencadena una reacción de oxidación dependiente de ATP produciendo luminiscencia. De esta manera, la luciferasa intracelular puede cuantificarse añadiendo una solución tamponada, que contiene un detergente para lisar las células y, posteriormente, un sustrato de luciferasa para iniciar la reacción de luminiscencia.

Para realizar estos ensayos, se escogió un sistema doble, que incluía un segundo gen reportero como control interno para normalizar la eficiencia de transfección y la diferencia de viabilidad de las células entre experimentos. Este segundo gen reportero, regulado por un promotor constitutivo, codifica para la luciferasa *Renilla* que requiere un sustrato diferente a la luciferasa *Firefly* para desencadenar la reacción que producirá luminiscencia. De esta manera, la actividad luminiscente de cada una de las dos luciferasas puede medirse por separado. Para la

cuantificación de la luminiscencia de las placas transfectadas se utilizó el sistema comercial *Dual-Luciferase® Reporter Assay System* (Promega) tal y como se indica en el PROTOCOLO XXV. Los valores obtenidos se expresan como unidades relativas (RLU, *Relative Light Units*) de la actividad de la luciferasa *Firefly/Renilla*.

14. ANÁLISIS ESTADÍSTICOS

Se calcularon las frecuencias genóticas y alélicas de todos los polimorfismos genotipados en este trabajo para:

- El grupo de muestras control utilizadas para validar e identificar polimorfismos en los genes *TFF* y *MUC1*.
- La población procedente del estudio observacional longitudinal de pacientes con lesiones gástricas.
- La población procedente del estudio caso-control de cáncer gástrico.

En estas dos últimas poblaciones se realizaron análisis de asociación de los genotipos y haplotipos de los polimorfismos analizados con la variable respuesta que se evaluaba en cada caso. En la primera población, de seguimiento de lesiones gástricas, la variable respuesta de evolución se determinó tal y como se describe en el APARTADO 1.1 de Materiales y Métodos. Al no ser una variable dicotómica se establecieron los siguientes bloques de análisis de asociación con la progresión/regresión de lesiones:

- Primer bloque: progresión *versus* estables + regresión
- Segundo bloque: y regresión *versus* estables + progresión.

En la segunda población, el estudio caso-control de cáncer gástrico, se evaluó la asociación de los polimorfismos con el riesgo de cáncer gástrico y los subtipos histológicos y anatómicos (intestinal/difuso y cardias/no cardias, respectivamente).

Como requisito indispensable previo al análisis de asociación, se comprobó que todos los polimorfismos estuvieran en equilibrio Hardy-Weinberg en las poblaciones control estudiadas. Para ello, se compararon las frecuencias genóticas observadas con las esperadas según las frecuencias alélicas en la población control, mediante el test de χ^2 implementado en la librería *SNPassoc* (377) en la interfase R. Todos los polimorfismos con un valor p asociado a este test menor de 1×10^{-4} se excluyeron del análisis.

Se realizó, en primer lugar, una estadística descriptiva de las principales características de cada población y se evaluó la asociación de estos factores con la variable respuesta mediante regresión logística univariante. Aquellas que resultaron ser significativas, en el estudio de evolución de lesiones gástricas, se incluyeron en el modelo final como variables de ajuste. En el estudio caso-control se incluyeron como variables de ajuste las utilizadas en el apareamiento de los casos con las muestras control. Finalmente se realizó el análisis de asociación entre los genotipos de los diversos polimorfismos y la variable respuesta mediante regresión logística multivariante (donde se incluyen las covariables) no condicional calculando los valores *Odd Ratio* (OR) y el intervalo de confianza (IC) en un 95%. En el análisis se consideraron 5 modelos de

herencia posibles: codominante, dominante, recesivo, sobredominante y aditivo. Se realizaron análisis estratificados en función del sexo, la infección por *H.pylori* y/o el consumo de AINES. Las posibles interacciones entre el SNP de interés y alguna de las variables citadas se abordó mediante del test de razón de verisimilitud (*likelihood ratio test*) Para los cálculos estadísticos se utilizó el programa estadístico SPSS versión 15 y la librería SNPAssoc en la interfase R. En el estudio caso control se realizó un test de permutaciones para cada gen, para tener en cuenta los múltiples comparaciones en relación a los diferentes modelos de herencia y SNPs considerados. Para cada permutación el valor p mínimo observado para gen fue registrado para construir la distribución nula. Además se consideró la posibilidad de la estratificación de la población. Los SNPs se ajustaron a una distribución uniforme, y no se requirió un control genómico, indicando que ajustando para el país de procedencia de los individuos corregía cualquier supuesta heterogeneidad étnica (165).

El desequilibrio de ligamiento es la asociación no aleatoria entre dos o más loci. En este contexto, los haplotipos se forman por la combinación de los alelos más o menos frecuentemente de lo que se esperaría por azar. Para el cálculo de las principales medidas de LD (D, D'y r²) entre todos polimorfismos genotipados de un mismo gen o región cromosómica, se utilizó el programa Haploview y la librería Genetics en la interfase R. Estas medidas de LD se basan en las diferencias entre las frecuencias alélicas observadas y las esperadas asumiendo una distribución al azar de los distintos alelos de cada polimorfismo. Debido a que el estudio se ha realizado a partir de la genotipado de diversos polimorfismos en individuos no relacionados, y no en familias o genealogías, no se ha podido determinar la fase de los haplotipos para cada individuo. Por esta razón, se determinaron las frecuencias haplotípicas mediante una estimación estadística a partir de las frecuencias alélicas observadas para cada polimorfismo. Para la inferencia de los haplotipos se utilizó el algoritmo E-M (*Expectation-Maximitation*) que proporciona estimaciones de máxima verosimilitud de las frecuencias haplotípicas. Este algoritmo se encuentra implementado en las librerías que se utilizaron para la inferencia en R. También se calculó, mediante regresión logística, la asociación entre los haplotipos estimados para cada región de interés y las variables respuesta especificadas anteriormente. En este caso, se incorpora la incertidumbre en la estimación de los haplotipos a los coeficientes de asociación del modelo de regresión.

Resultados

1. VARIABILIDAD GENÉTICA DE *TFFS* Y *MUCS*

Para analizar la variabilidad genética en genes de protección de la mucosa gástrica se procedió, en primer lugar, a la selección de polimorfismos potencialmente funcionales para validarlos, posteriormente, en una población control, y se identificaron nuevas variantes en los genes *TFF*. En el caso de mucinas, esta primera fase sólo se llevó a cabo para *MUC1* ya que el tamaño y complejidad de los otros genes *MUC*, con múltiples exones y secuencias repetitivas, dificultaban el proceso de validación. En segundo lugar, se realizó un análisis de desequilibrio de ligamiento (LD, *Linkage Disequilibrium*) e identificación de haplotipos en las regiones donde se localizan los genes *TFF* y *MUC*. Este análisis nos permitió seleccionar SNPs marcadores (tagSNPs) de la variabilidad genética de la región que, juntamente con los SNPs potencialmente funcionales, pasarían a formar parte del análisis de asociación en el estudio de seguimiento de pacientes con lesiones gástricas y en el estudio caso control de cáncer gástrico.

1.1. Selección de SNPs potencialmente funcionales descritos en las bases de datos

Se realizó una búsqueda de SNPs potencialmente funcionales en las bases de datos públicas, que serían posteriormente validados en un grupo de 40- 50 muestras controles. Hay que tener en cuenta que, en el momento de la búsqueda, los datos disponibles correspondían a la versión 126 de dbSNP. Debido a la actualización de las bases de datos consultadas, tanto del dbSNP (actualmente versión 137) como de la base de secuencias de referencia del NCBI, desde que se inició este estudio ha aumentado considerablemente el número de SNPs descritos para todos los genes en estudio y, en unos pocos casos, se han producido pequeñas variaciones en la posición relativa de los SNPs en las secuencias de referencia. Estas variaciones no afectan, en esencia, a la tipología y posible funcionalidad de estos polimorfismos (ver TABLA A-1 del apéndice II), por lo que en esta tesis se utilizará la nomenclatura de los polimorfismos correspondiente a la versión dbSNP 137 que se basa en las secuencias de referencia más actuales para cada gen.

En el momento de la búsqueda se encontraron 17 polimorfismos en *TFF1*, 14 en *TFF2* y 7 en *TFF3*, descritos en alguna de las tres bases de datos consultadas: SNP500Cancer, dbSNP (versión 126) y Ensembl (TABLA A-1). En SNP500Cancer sólo se encontraron 6 SNPs descritos, todos pendientes de validación y ninguno en *TFF2*. Gran parte de los polimorfismos descritos en el dbSNP (41.7%) no estaba validado y el 77.8% de los SNPs eran intrónicos. En el caso de *MUC1*, se encontraron un total de 35 polimorfismos descritos en dbSNP (TABLA A-1 del apéndice I). Un elevado porcentaje de los SNPs descritos (34.2%) se localizaban en la región genómica a 5' o 3' del gen. De los 23 SNPs localizados propiamente en *MUC1*, prácticamente la mitad no estaban validados (47.8%) y la mayoría eran intrónicos (60.9%). El único polimorfismo localizado en *MUC1* validado y con una frecuencia superior al 5% en población caucásica fue el rs4072037, un SNP sinónimo (T31T) en el exón 2, relacionado con el procesamiento alternativo y más recientemente asociado con cáncer gástrico (166). De hecho, este SNP también fue el único que estaba descrito

en el gen *MUC1* en SNP500cancer y en HapMap, en el momento de la búsqueda. En el caso de la base de datos Ensembl, se incluían 2 SNPs localizados en *MUC1* y región adyacente, el rs4072037, citado anteriormente, y el rs11465204 en la región 5'UTR del gen).

También se consultó la aplicación web PupaSuite para determinar los posibles efectos funcionales de algunos de los SNPs de *TFFs* y *MUC1* descritos en las bases de datos, especialmente los de la región promotora y SNPs intrónicos próximos a los exones. La utilización de esta herramienta únicamente nos permitió identificar como potencialmente funcional el SNP rs13052596 (c.-450G>T en *TFF2*) al localizarse en una secuencia consenso para el sitio de unión del factor de transcripción NFκB.

Siguiendo los criterios descritos en materiales y métodos (APARTADO 2.1.1) se seleccionaron un total de 10 SNPs en *TFFs*: 2 en *TFF1*, 4 en *TFF2* y 3 en *TFF3*, todos ellos localizados en zonas codificantes, en regiones 5'UTR o 3'UTR y regiones promotoras. Además se incluyó el polimorfismo VNTR del intrón 2 de *TFF2* ya descrito en la literatura (APARTADO 1.5.3 de la introducción). En *MUC1* se seleccionaron también 10 SNPs: 6 SNPs exónicos, 3 en la región 3'UTR y 1 SNP localizado en la región 5'UTR, muy próximo al exón 1. Los SNPs seleccionados para su validación se muestran en negrita en la TABLA A-1 del apéndice

1.2. Validación de SNPs potencialmente funcionales e identificación de nuevas variantes

La validación de los polimorfismos de *TFFs* seleccionados previamente, se realizó mediante digestión o SSCPs en un máximo de 50 muestras control. Debido a la naturaleza de la técnica, los SSCPs realizados para validar las variantes descritas en las bases de datos sirvieron también para identificar nuevos polimorfismos. Adicionalmente, todas las regiones funcionales de los genes en estudio (exones con secuencias intrónicas flanqueantes y regiones 5' y 3' del gen) también se analizaron mediante esta técnica con el fin de ampliar la búsqueda de nuevas variantes. La validación de los SNPs seleccionados en *MUC1* y la búsqueda de nuevos polimorfismos se realizaron paralelamente mediante SSCPs y/o HRM en un grupo control de 40-60 muestras. En la TABLA R-1 se resumen las características de todos los polimorfismos potencialmente funcionales que se seleccionaron para su validación y las nuevas variantes identificadas en genes *TFF*.

1.2.1. *TFF1*

En el gen *TFF1* se seleccionaron dos polimorfismos, el c.-2C>T (rs215630) en la región 5'UTR y el c.*41G>A (rs13276) en la región 3'UTR. El SNP c.-2C>T se validó mediante análisis de restricción con la enzima *BsrDI* de un fragmento de 227 pb correspondiente al exón 1 (FIGURA R-1). Mediante la visualización del patrón de bandas se confirmó la presencia del SNP, obteniéndose una frecuencia alélica del alelo menor (MAF) de 0.12, muy similar a la publicada para población caucásica en el dbSNP (MAF=0.13).

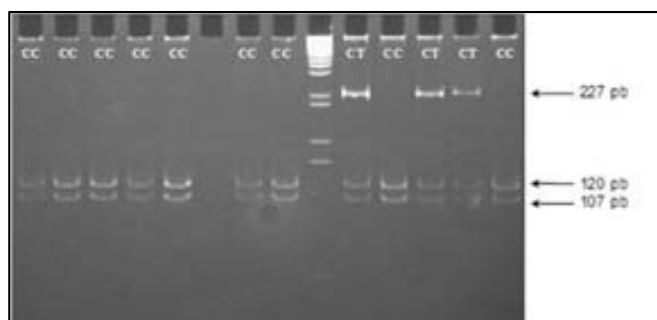


Figura R-1. Digestión enzimática con *BsrDI* de *TFF1* c.-2C>T. Alelo C: 120pb + 107pb. Alelo T: 227pb. En el carril 9 se muestra el marcador de 1 kb.

El mismo fragmento se analizó mediante SSCPs para buscar otros polimorfismos no descritos. La secuenciación de las muestras que presentaban patrones diferentes en el análisis confirmó de nuevo el SNP c.-2C>T, validado anteriormente por digestión, y permitió la identificación de una nueva variante, c.-29G>A, localizada en la región 5'UTR (FIGURA R-2). También se observaron los patrones correspondientes a otro polimorfismo intrónico muy frecuente descrito en dbSNP en esta región, el c.85+8G>A (rs225358).

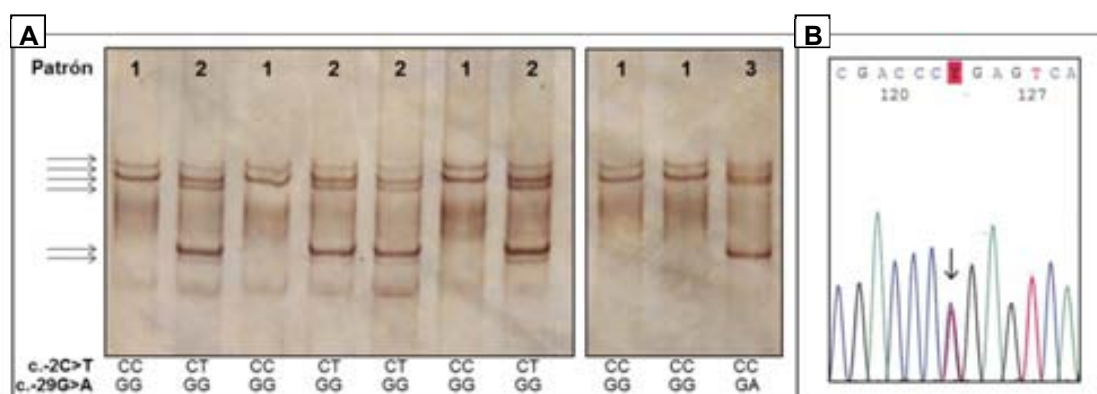


Figura R-2. A. SSCPs del exón 1 de *TFF1*. Se muestra el genotipo de c.-2C>T, ya analizado previamente por digestión con *BsrDI*, y el nuevo polimorfismo identificado c.-29G>T. B. Secuencia correspondiente al heterocigoto del SNP c.-29G>A, cadena reversa

El polimorfismo c.*41G>A (rs13276), en el exón 3, se validó mediante SSCPs de un fragmento de 378pb, detectándose 3 patrones diferentes (FIGURA R-3A). Las secuencias de las muestras correspondientes a cada patrón confirmaron la existencia del polimorfismo a validar y revelaron 2 nuevos polimorfismos c.239-9C>T, en el intrón 2, y c.*62C>T en la región 3'UTR (FIGURA R-3B). Las frecuencias alélicas para el SNP c.*41G>A, que no estaba validado en las bases de datos, fueron de 0.975(G) y 0.025(A). También se analizó un fragmento de 362 pb correspondiente al exón 2 de *TFF1*. Se analizaron 50 muestras controles, de las que 2 mostraron un patrón diferente al resto confirmándose mediante secuencia la presencia del polimorfismo c.238+18G>C, un SNP intrónico descrito en las bases de datos (rs7281220) que no fue seleccionado para este estudio.

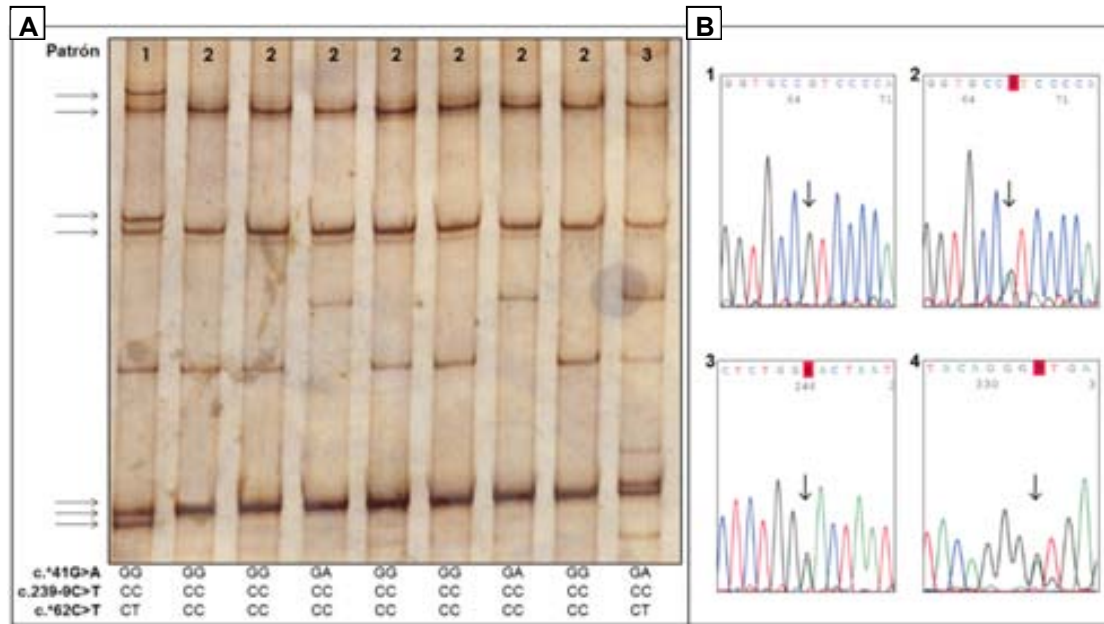


Figura R-3. A. SSCPs del exón 3 de *TFF1*. Se distinguen tres patrones diferentes correspondientes a los polimorfismos c.*41G>A, c.239-9C>T y c.*62C>T. B. Secuencias correspondientes a los polimorfismos detectados en la figura A: 1- c.*41G>A, homocigoto GG. 2- c.*41G>A, heterocigoto GA. 3- c.239-9C>T, heterocigoto GA, cadena reversa. 4- c.*62C>T, heterocigoto GA, cadena reversa.

1.2.2. *TFF2*

Los polimorfismos c.-450G>T (rs13052596), c.7C>T (rs7277409), c.*82T>C (rs225334) y c.*160C>T (rs4920084), seleccionados previamente en *TFF2*, se validaron mediante SSCPs.

En el análisis de un fragmento correspondiente a la región promotora 5' del gen, donde se localizaba el SNP c.-450G>T, se detectaron 3 patrones diferentes (FIGURA R-4A). Las correspondientes secuencias confirmaron que el cambio de movilidad se debía a los 3 genotipos del polimorfismo a validar (FIGURA R-4B) y no se halló en la secuencia ninguna otra variante. Las frecuencias alélicas obtenidas en la muestras control para el SNP c.-450G>T, MAF=0.33, coincidían con las reportadas en dbSNP.

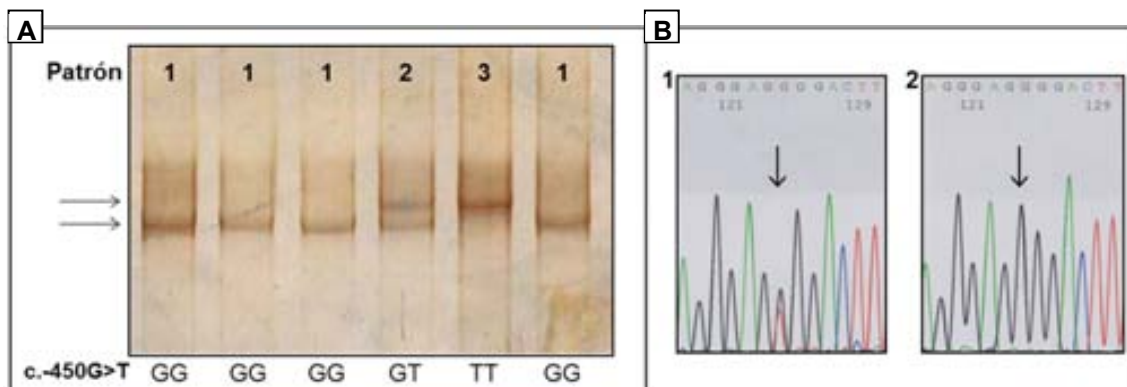


Figura R-4. A. SSCPs de la región 5' del gen *TFF2*. Se distinguen tres patrones diferentes correspondientes a los 3 genotipos del SNP c.-450G>T. B. Secuencias correspondientes al SNP c.-450G>T: 1- heterocigoto GT. 2- homocigoto GG.

El análisis del fragmento de 357 pb correspondiente al exón 1, donde se localizaba otro de los SNPs descritos (c.7C>T, R3W), también se detectaron tres patrones diferentes (FIGURA R-5A). Sin embargo, todos los patrones correspondieron a muestras homocigotas CC para el polimorfismo c.7C>T y los cambios de movilidad electroforética se debían a una nueva variante identificada, c.20A>C (Q7P) (FIGURA R-5B). En este fragmento no se identificó ninguna otra variante.

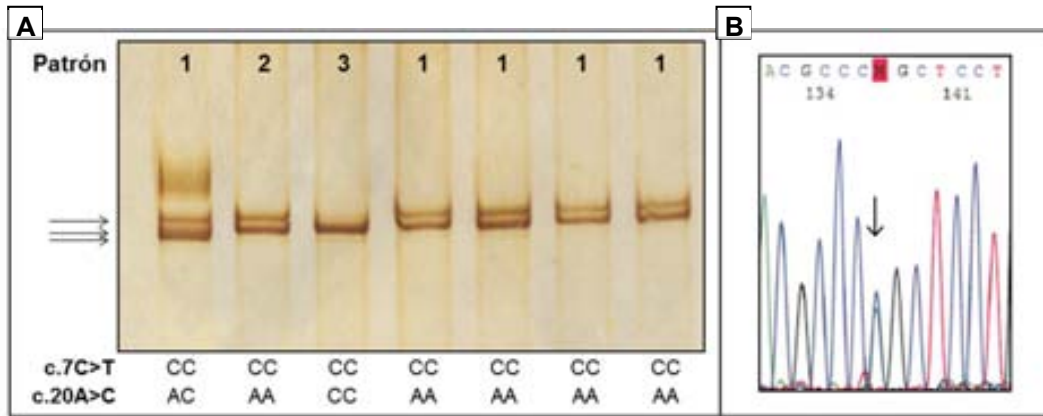


Figura R-5. A. SSCPs del exón 1 de *TFF2*. Se distinguen tres patrones diferentes correspondiente al polimorfismo c.20A>C. B. Secuencia correspondiente al heterocigoto del SNP c.20A>C.

Los polimorfismos de 3'UTR c.*82T>C y c.*160C>T se validaron mediante el análisis por SSCPs de un fragmento de 427 pb correspondiente al exón 4, detectándose cinco patrones diferentes (FIGURA R-6A).

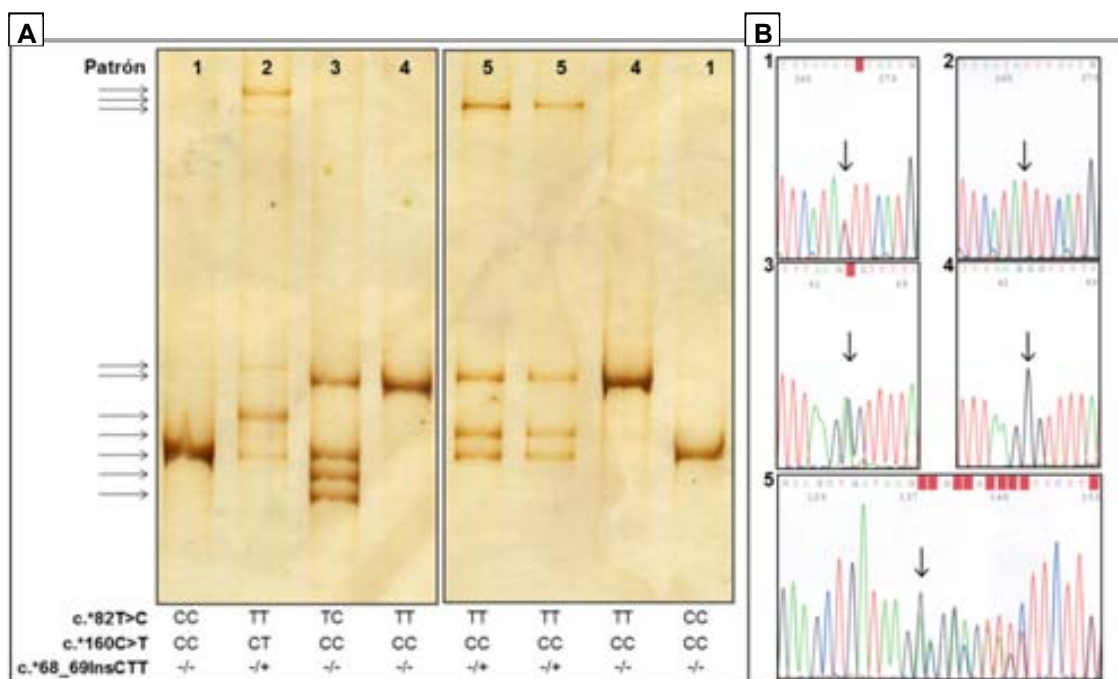


Figura R-6. A. SSCPs del exón 4 de *TFF2*. Patrones correspondientes a los polimorfismos c.*82T>C, c.*160C>T y c.*68_69insCTT. B. Secuencias correspondientes a los polimorfismos detectados en la figura A: 1- c.*82T>C, heterocigoto TC. 2- c.*82T>C, homocigoto TT. 3- c.*160C>T, heterocigoto GA, cadena reversa. 4- c.*160C>T, homocigoto GG, cadena reversa. 5- c.*68_69insCTT, heterocigoto.

Se secuenció una muestra de cada patrón y se confirmaron los dos polimorfismos descritos. Además se identificó en las secuencias un nuevo polimorfismo de inserción de 3pb, c*68_69insCTT, en la región 3'UTR del gen (FIGURA R-6B).

Debido a la dificultad de genotipar una variante tipo inserción mediante técnicas convencionales, como las sondas de hibridación FRET, y al encontrarse este polimorfismo muy próximo al SNP c.*82T>C se puso a punto la metodología *High Resolution Melting* (HRM) para su análisis. Además, esta técnica permitía el genotipado simultáneo de las dos variantes de una manera rápida y con un rendimiento óptimo. En la FIGURA R-7 se muestran los patrones correspondientes a los diferentes haplotipos, formados por los polimorfismos c*68_69insCTT y c.*82T>C.

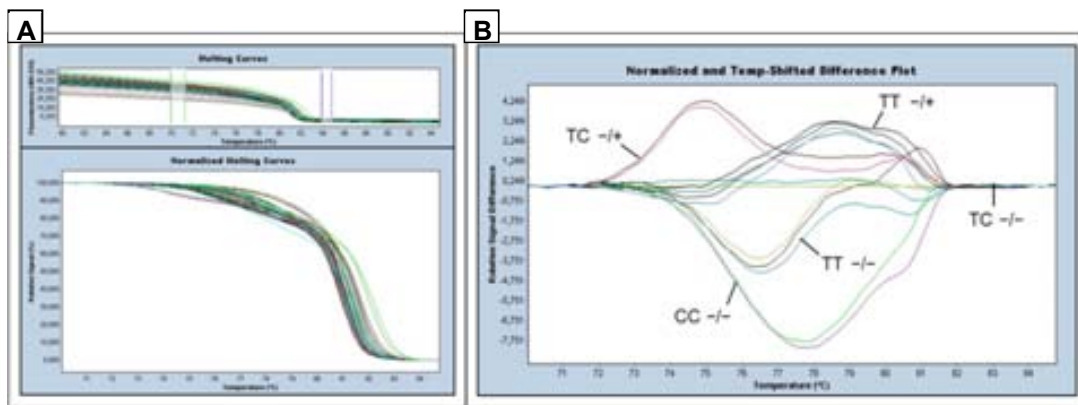


Figura R-7. Análisis mediante HRM de los polimorfismos c.*68_69insCTT y c.*82T>C de *TFF2*. A. Representación gráfica de la de curva de fusión. Arriba: Curva de fusión. Abajo: Normalización de la fluorescencia para la curva de fusión. B. Perfiles de la curva de fusión. Para cada perfil se indica el haplotipo correspondiente a los polimorfismos c.*82T>C y c.*68_69insCTT (la presencia de la inserción se indica con “+” y la ausencia con “-”). El perfil de referencia es el correspondiente al TC -/-

También se analizó, mediante SCCPs, un fragmento de 371 pb correspondiente al exón 2, detectándose diferentes patrones de bandas que mediante secuenciación confirmamos que se debían a la presencia de 2 polimorfismos no descritos previamente: c.196G>C (un polimorfismo de cambio de aminoácido V66L) y c.229+81C>T que se localiza en el intrón 3 (FIGURA R-8). El análisis de un fragmento de 244 pb del exón 3 no reveló ningún cambio de movilidad electroforética.

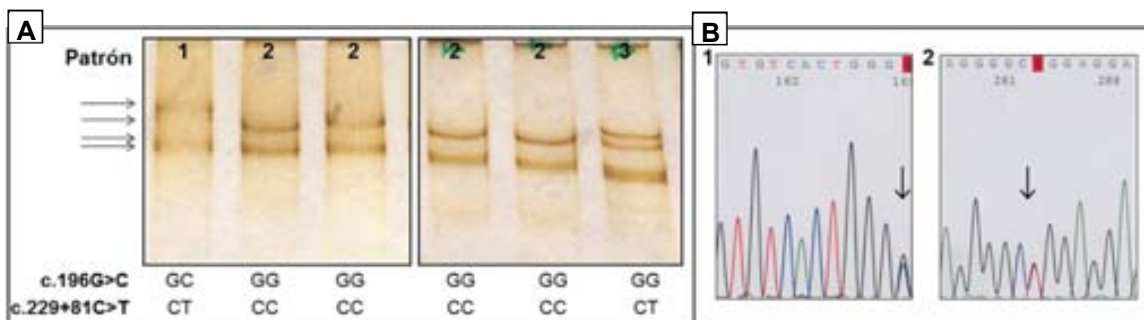


Figura R-8. A. SCCPs del exón 2 del gen *TFF2*. Se distinguen tres patrones diferentes correspondiente a los polimorfismos c.196G>C y c.229+81C>T. B. Secuencias correspondientes a los polimorfismos detectados en la figura A: 1- c.196G>C, heterocigoto GC. 2- c.229+81C>T, heterocigoto CT.

Finalmente, el VNTR de repeticiones de 25 pb, localizado en el intrón 2 de *TFF2*, se validó mediante PCR convencional y electroforesis en gel de agarosa, donde se distinguían dos bandas diferentes. El análisis de la secuencia permitió verificar que se trataba de los dos alelos (40 y 48 repeticiones) del VNTR descrito.

1.2.3. TFF3

En el gen *TFF3* se seleccionó un SNP de la región 5'UTR c.-3A>G (rs11701143) y dos SNPs no sinónimos: c.83C>T (A28V, rs11546960) y c.176C>T (P59L, rs11546961).

Los SNPs c.-3A>G y c.83C>T se validaron mediante análisis de restricción con las enzimas *Avall* y *HaeIII*, respectivamente, de un fragmento de 235 pb correspondiente al exón 1 (FIGURA R-9). Se analizaron un total de 50 controles obteniéndose una MAF de 0.05 para el c.106A>G, similar a la frecuencias descrita dbSNP (MAF=0.08). Todas las muestras analizadas fueron homocigotas para el alelo C del polimorfismo c.83C>T, variante que no estaba validada en el dbSNP. Mediante la digestión con *HaeIII* se confirmó nuevamente la existencia del polimorfismo c.-3A>G ya que también destruye una diana de restricción para dicha enzima.

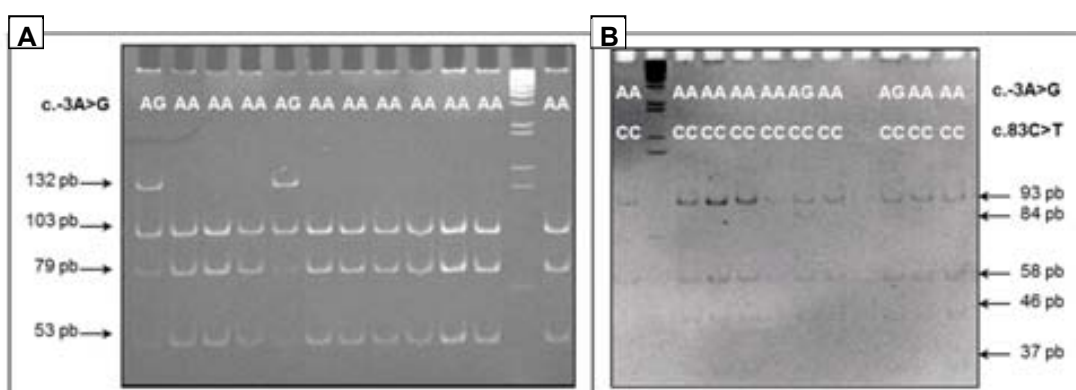


Figura R-9. A. Digestión enzimática con *Avall* de c.-3A>G, en el exón 1 de *TFF3*. Alelo G: 132pb+ 103pb. Alelo A: 103pb+79pb+53pb. B. Digestión enzimática con *HaeIII* de c.-3A>G y c.83C>T, en el exón 1 de *TFF3*. Haplotipo AC: 93pb+58pb+46pb+37pb. Alelo GC: 84pb+58pb+46pb+37pb+9pb. Los haplotipos AT (130pb+58pb+47pb) y GT (121+pb58pb+42pb+11pb) no se observaron en el gel.

Este mismo fragmento también se analizó mediante SSCPS para buscar otros polimorfismos no descritos (FIGURA R-10A). En el gel se observó un cambio de patrón que, mediante secuenciación, se comprobó que correspondía al polimorfismo ya validado, no hallándose nuevos polimorfismos. La validación del SNP c.176C>T se realizó mediante SSCPs de un fragmento de 295 pb correspondiente al exón 2 de *TFF3* (FIGURA R-11A). Sólo se observó 1 cambio de patrón en el heterodúplex de una muestra. Este cambio no corresponde a c.176C>T sino a una variante no descrita, c.238 G>A, de cambio de aminoácido (V80M) (FIGURA R-11B)

Finalmente, el análisis del exón 3, incluido en un fragmento de 364 pb, reveló un patrón diferente al resto en el gel de SSCPs (FIGURA R-10B). Se caracterizó el cambio mediante secuenciación hallándose una nueva variante en la región 3'UTR (c.*41G>A).

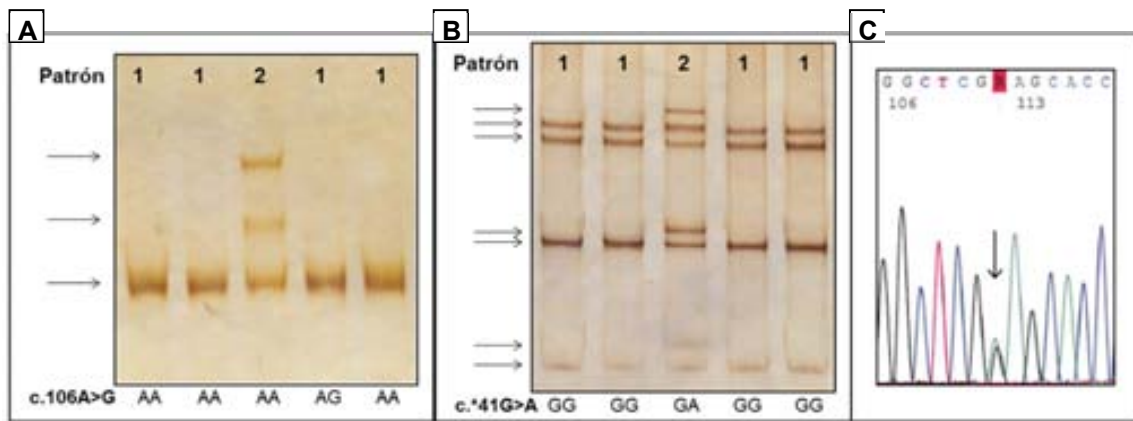


Figura R-10. A. SSCPs del exón 1 de *TFF3*. Se muestra el genotipo del polimorfismo c.-3A>G (rs11701143) ya analizado previamente por digestión con *Avall*. B. SSCPs del exón 3 de *TFF3*. Se muestra el genotipo del nuevo polimorfismo identificado c.*41G>A. C. B. Secuencia correspondiente al heterocigoto del SNP c.*41G>A.

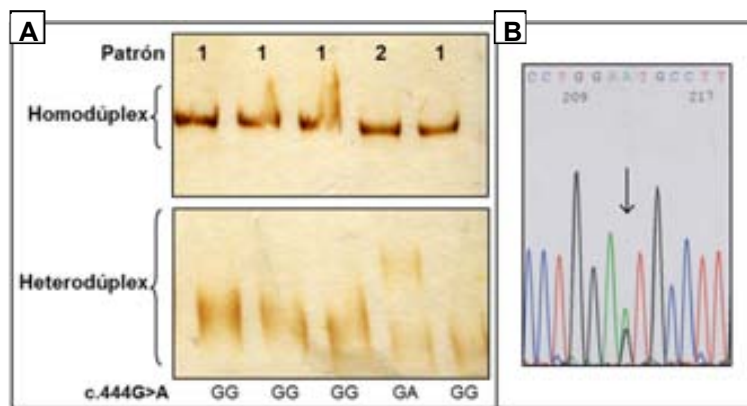


Figura R-11. A. SSCPs del exón 2 de *TFF3*. Se observa un patrón diferente en el heterodúplex correspondiente al polimorfismo c.238G>A. B. Secuencia correspondiente al heterocigoto del SNP c.238G>A.

Tabla R-1. Validación de SNPs seleccionados y nuevas variantes identificadas en TFFs.

Gen	SNP	rs dbSNP	Tipología	Localización	Validación	MAF ^a
<i>TFF1</i>	c.-29C>T	ND	5'UTR	Exón 1	ND	0.07
<i>TFF1</i>	c.-2C>T	rs2156310	5'UTR	Exón 1	Si	0.12
<i>TFF1</i>	c.239-9C>T	ND	Intrónico	Intrón 2	ND	0.02
<i>TFF1</i>	c.*41G>A	rs13276	3'UTR	Exón 3	Si	0.03
<i>TFF1</i>	c.*62C>T	ND	3'UTR	Exón 3	ND	0.02
<i>TFF2</i>	c.-450G>T	rs13052596	Promotor	Genómico 5'	Si	0.27
<i>TFF2</i>	c.7C>T	rs7277409	R3W	Exón 1	No	0
<i>TFF2</i>	c.20A>C	ND	Q7P	Exón 1	ND	0.03
<i>TFF2</i>	c.196 G>C	ND	V66L	Exón 2	ND	0.03
<i>TFF2</i>	c.229+81C>T	ND	Intrónico	Intrón 2	ND	0.03
<i>TFF2</i>	c.*68_69insCTT	ND	3'UTR	Exón 4	ND	0.06
<i>TFF2</i>	c.*82T>C	rs225334	3'UTR	Exón 4	Si	0.40
<i>TFF2</i>	c.*160C>T	rs4920084	3'UTR	Exón 4	Si	0.02
<i>TFF3</i>	c.-3A>G	rs11701143	5'UTR	Exón 1	Si	0.05
<i>TFF3</i>	c.83C>T	rs11546960	A28V	Exón 1	No	0
<i>TFF3</i>	c.176C>T	rs11546961	P59L	Exón 2	No	0
<i>TFF3</i>	c.238G>A	ND	V80M	Exón 2	ND	0.01
<i>TFF3</i>	c.*41G>A	ND	3'UTR	Exón 3	ND	0.03

a, MAF, frecuencia del alelo menor obtenida a partir del análisis inicial de ~50 muestras. ND, No Descrito en la base de datos dbSNP v126, en el momento de su identificación.

1.2.4. *MUC1*

Se analizaron mediante SSCPs los fragmentos correspondientes a los exones donde se localizaban los 10 SNPs potencialmente funcionales previamente seleccionados de las bases de datos para su validación (TABLA A-1 del apéndice). El análisis del exón 1 (c.-71G>A, rs11465204), exón 3 (S66T, rs35819649) y exón 5 (V126M, rs1611770 y N131K, rs41264913) no reveló ningún cambio de patrón (FIGURA R-12). El exón 8, donde se localizaba un SNP sinónimo (Y252Y, rs11465209) y tres SNPs en la región 3'UTR (c.*6C>T rs1611772, c.*79C>T rs1611773 y c.*197G>A rs11465210), se analizó en tres fragmentos debido a su gran tamaño y los requisitos de las técnicas utilizadas. No obstante, tampoco se observaron cambios en ninguno de los tres fragmentos (FIGURA R-12).

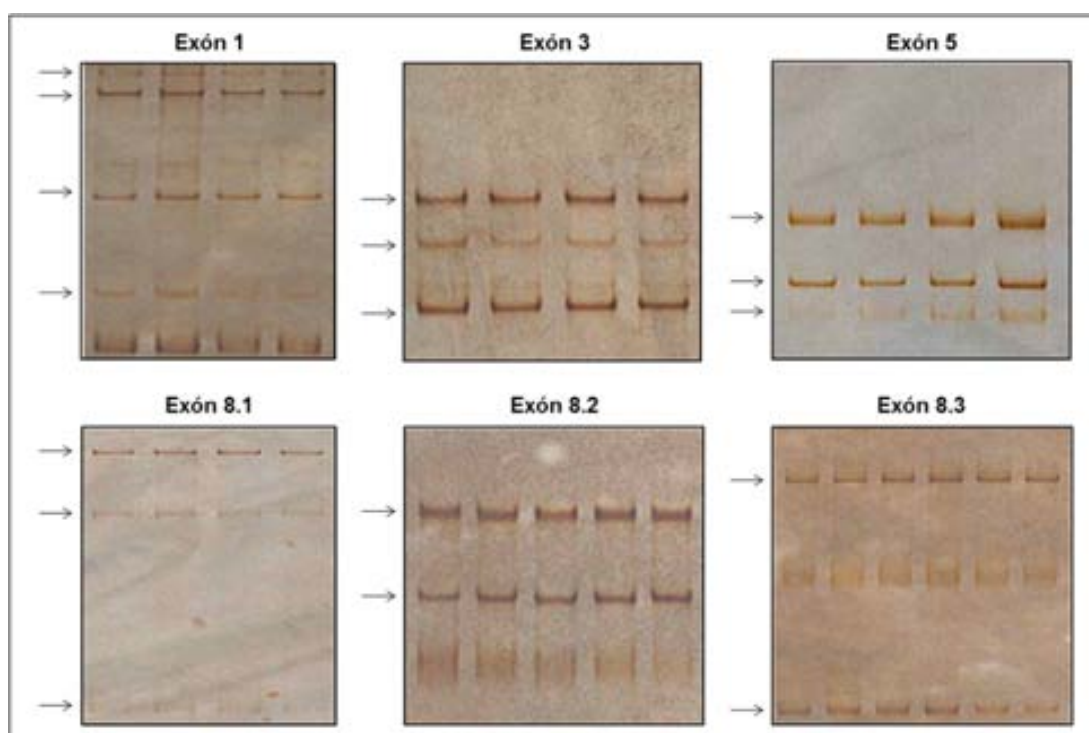


Figura R-12. SSCPs de los fragmentos correspondientes a los exones 1, 3, 5 y 8 de *MUC1*. No se observó ningún cambio en el patrón de movilidad electroforética.

En el análisis del exón 2, donde se localizaba el SNP sinónimo T31T (rs4072037), descrito previamente en la literatura, se observaron diferentes patrones que, mediante secuenciación, confirmaron la presencia de este polimorfismo (FIGURA R-13). A pesar de la existencia de un SNP no sinónimo en el exón 6 (N151S, rs11465207), este fragmento se analizó únicamente mediante HRM ya que este polimorfismo estaba descrito sólo en población asiática y africana, siendo su frecuencia 0 en población caucásica según dbSNP.

Todos los fragmentos correspondientes a los exones y regiones flanqueantes de *MUC1*, incluyendo aquellos analizados previamente por SSCPs, se analizaron mediante la técnica HRM. A pesar de haber diseñado específicamente los cebadores utilizados en la amplificación de los distintos fragmentos para obtener amplicones pequeños y, así, aumentar la sensibilidad de esta

técnica para el cribado de variantes (APARTADO 2.1.1 de materiales y métodos), no se observaron cambios en el perfil de la curva de fusión. En la FIGURA R-14A se muestra el perfil obtenido para el exón 7, el resto de exones mostraron perfiles similares. La única excepción fue el análisis del exón 2, que de nuevo confirmó la presencia de rs4072037, validado previamente mediante SSCPs (FIGURA R-14B).

En resumen, el análisis de los 8 exones de *MUC1* mediante SSCPs y/o HRM únicamente permitió validar el SNP c.93A>G T31T (rs4072037) ya descrito previamente. En el resto de fragmentos no se observaron cambios de patrón o de perfil mediante las técnicas utilizadas.

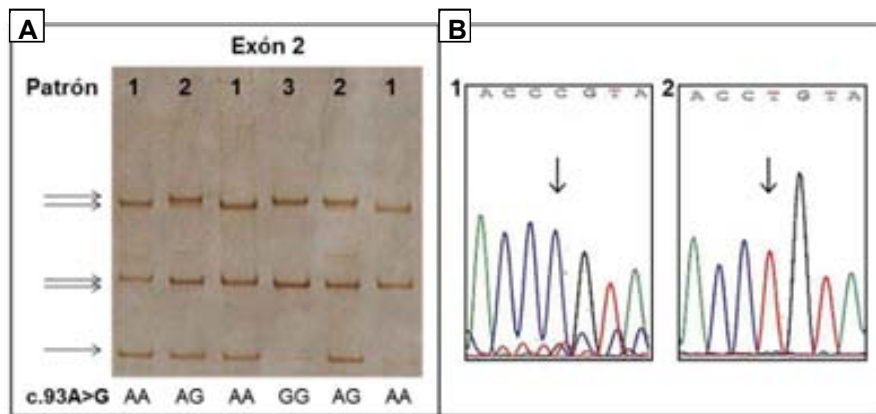


Figura R-13. A. SSCPs del exón 2 de *MUC1*. Se distinguen tres patrones correspondientes a los diferentes genotipos del polimorfismo c.93G>A (rs4072037). B. Secuencia correspondiente al SNP c.93G>A. 1- Homocigoto CC, cadena reversa. 2- Homocigoto TT, cadena reversa.

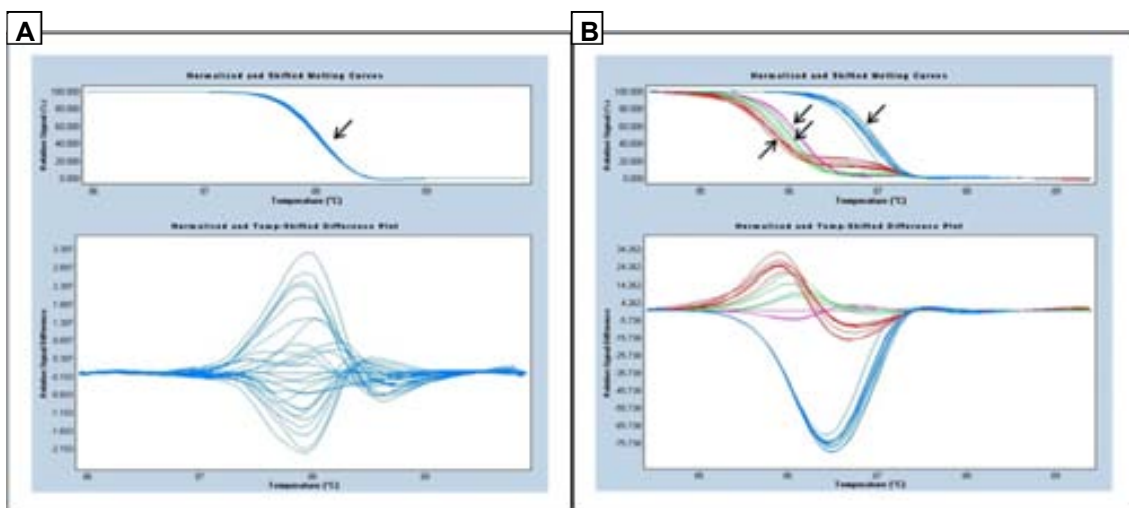


Figura R-14. A. Perfil de la curva de fusión del HRM del exón 7 de *MUC1*. No se observan diferencias en los perfiles de fusión de las muestras analizadas. B. Perfiles de la curva de fusión del HRM del exón 2 de *MUC1*. En azul homocigotos AA, en color verde homocigotos GG, en rojo heterocigotos GA, para el SNP rs4072037. En fucsia, muestras no amplificadas y control negativo.

1.3. Análisis de desequilibrio de ligamiento, identificación de haplotipos y selección de tagSNPs

El análisis de LD de los genes *TFF* y *MUC* y la selección de tagSNPs de los principales bloques de haplotipos se realizó con el programa Haploview, analizando los datos de genotipado del proyecto HapMap (versión 24, fase II, nov08), siguiendo la estrategia descrita en el APARTADO 2.1.2 de materiales y métodos. Los tagSNPs que no tenían un diseño óptimo de genotipado en las plataformas a utilizar se sustituyeron, siempre que fue posible, por otros en total LD para optimizar el proceso. También se priorizaron los tagSNPs que, además, podían ser funcionales.

1.3.1. *TFFs*

Debido al pequeño tamaño de estos tres genes (*TFF1* 4.2 kb, *TFF2* 4.7 kb y *TFF3* 3.9 kb) y al no encontrar bloques de LD claros al analizar cada gen individualmente, se procedió a analizar conjuntamente todo el clúster de *TFF*. Se obtuvieron los datos de genotipado de HapMap de 121 kb de la región cromosómica 21q22.3 que incluía el clúster de una longitud aproximada de 55 kb. En esta región se seleccionaron un total seis bloques de LD: dos, de 4 y 7 kb, en la región que incluía el gen *TFF3* y cuatro bloques, de longitudes comprendidas entre 1 y 18 kb, que incluían los genes *TFF2* y *TFF1* (FIGURA R-15A). Se seleccionaron tagSNPs que marcaran haplotipos con una frecuencia mayor del 0.05 en todos los bloques excepto en el número 3, el más extenso de todos con una longitud de 18 kb, para el que se seleccionaron tagSNPs que marcaran haplotipos con frecuencias mayores de 0.01. De esta manera se optimizó la relación entre número de tagSNPs seleccionados, que se incrementó debido al poco LD presente en la región, y la cobertura de la variabilidad genética. Mediante esta estrategia se conseguía capturar en todos los casos más del 90% de posibles haplotipos y se minimizó el número de SNPs a genotipar.

En la FIGURA R-15B se muestran los principales bloques de haplotipos y los tagSNPs identificados y seleccionados para cada bloque, los cuales también se detallan en la TABLA R-2 con su correspondiente “rs” y frecuencia alélica, incluyendo aquellos que se eliminaron o sustituyeron. El SNP 49 del bloque 2 se eliminó de la selección inicial porque no disponía de un buen diseño para el genotipado pero fue sustituido por el 51 ($D'=1$, $r^2=1$) en el mismo bloque. También se priorizaron los tagSNPs que, además, podían ser funcionales por lo que se substituyó el 91 del bloque 3 por el 90 del mismo bloque, un SNP (rs225334) localizado en la región 3'UTR del gen de *TFF3* y validado previamente (TABLA R-1). El SNP 128, también en el bloque 3, se eliminó de la selección porque no presentaba un diseño óptimo para su genotipado aunque estaba marcado por un SNP muy próximo, el 125.

En total se seleccionaron 19 tagSNPs: siete en los bloques 1 y 2, donde se localiza *TFF3*, cinco en el bloque 3, correspondiente a *TFF2* y parte de *TFF1*, y siete en los bloques del 4 al 6, que incluye a *TFF1* y región genómica a 5' del gen (TABLA R-2).

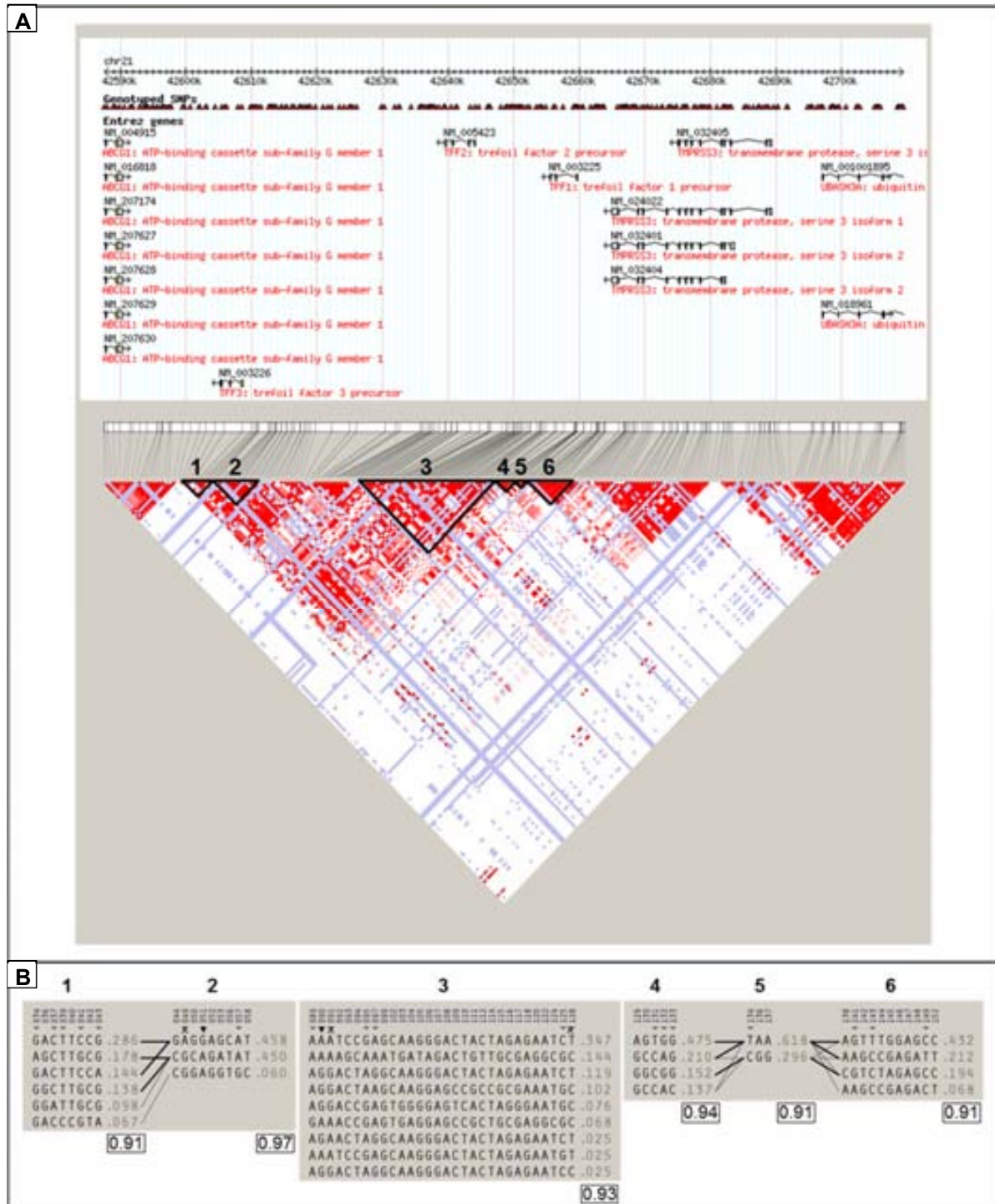


Figura R-15. Desequilibrio de ligamiento en el clúster de *TFFs* en la región genómica 21q22.3, según los datos obtenidos para población CEU en HapMap. **A.** Representación gráfica del LD. Esquema de color: cuadrados en blanco, $D' < 1$ y $LOD < 2$; en azul, $D' = 1$ y $LOD < 2$; en tonos rosa y rojo, $D' < 1$ y $LOD \geq 2$ y en rojo fuerte, $D' = 1$ y $LOD \geq 2$. El rojo y azul es indicativo de una fuerte evidencia de LD mientras que el blanco indica una fuerte evidencia de recombinación, en tonalidad rosa se muestran valores intermedios. Se indican los bloques seleccionados y la estructura de los genes presentes en la región. **B.** Bloques de haplotipos y tagSNPs seleccionados (▼). Las X indican aquellos tagSNPs que fueron eliminados o substituidos. También se indica el porcentaje de posibles haplotipos capturado en cada bloque por el conjunto de tag SNPs seleccionados.

Tabla R-2. Selección de tagSNPs en el clúster de *TFFs*

Región	Bloque ^a	Posición Bloque	SNP	Alelos	MAF ^b	Selección
<i>TFF3</i>	1	34	rs8133510	G/A	0.190	Si
<i>TFF3</i>	1	37	rs2236705	C/A	0.119	Si
<i>TFF3</i>	1	39	rs11701143	T/C	0.083	Si
<i>TFF3</i>	1	41	rs225363	G/C	0.467	Si
<i>TFF3</i>	1	43	rs9981660	G/A	0.181	Si
<i>TFF3</i>	2	49	rs225366	G/A	0.458	Sustituido por 51
<i>TFF3</i>	2	51	rs692356	A/G	0.455	Si
<i>TFF3</i>	2	57	rs11203197	A/G	0.075	Si
<i>TFF2</i>	3	89	rs12483635	A/G	0.083	Si
<i>TFF2</i>	3	90	rs225334	A/G	0.368	Si
<i>TFF2</i>	3	91	rs225335	A/G	0.350	Sustituido por 90
<i>TFF2</i>	3	95	rs1079380	G/A	0.442	Si
<i>TFF2</i>	3	97	rs3814896	A/G	0.178	Si
<i>TFF1</i>	3	125	rs225353	C/G	0.458	Si
<i>TFF1</i>	3	128	rs225355	T/C	0.465	Eliminado
<i>TFF1</i>	4	131	rs225357	C/T	0.492	Si
<i>TFF1</i>	4	132	rs4920094	G/A	0.347	Si
<i>TFF1</i>	4	133	rs13051704	G/C	0.157	Si
<i>TFF1</i>	5	134	rs225358	T/C	0.364	Si
<i>TFF1</i>	6	141	rs9976977	G/A	0.283	Si
<i>TFF1</i>	6	143	rs424694	C/T	0.492	Si
<i>TFF1</i>	6	149	rs13047838	C/T	0.217	Si

a, El número de bloque y la posición detallada hace referencia a la figura R-15; b, MAF según la versión 24, fase II de HapMap.

1.3.2. MUCs

En la región genómica de 1q21 que contiene *MUC1* (4.4 kb) y parte del gen *TRIM46*, a 855 pb a 3' de *MUC1*, se definieron dos bloques de haplotipos abarcando 51 kb (FIGURA R-16A). En esta región se seleccionaron un total de cuatro tagSNPs, dos por cada bloque, que marcaban haplotipos con una frecuencia superior al 3% (FIGURA R-16B). Uno de los SNPs, el 31 en el bloque 2, fue sustituido por el 27 que corresponde a un SNP sinónimo (T31T, rs4072037) asociado a cáncer gástrico en publicaciones recientes y validado previamente en nuestro estudio. El porcentaje de haplotipos capturados en el bloque 1 fue del 99% y en bloque 2 del 92%.

En la región cromosómica de 11p15 donde se localiza *MUC6* (22.9 kb), presentaba un elevado LD por lo que se definió un único bloque de 29 kb que contenía íntegramente el gen (FIGURA R-17A). En este bloque se seleccionaron un total de seis SNPs que marcaban haplotipos con una frecuencia superior a 0.05. (FIGURA R-17B). El tagSNP 14 se sustituyó por el 29, un polimorfismo de cambio de aminoácido, M619V (rs7481521), localizado en el exón 16 del gen.

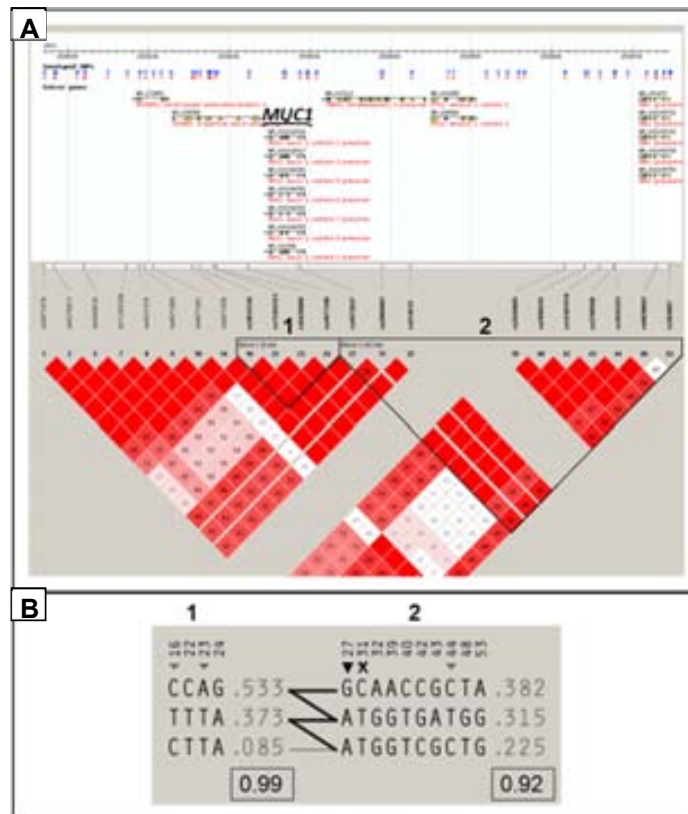


Figura R-16. A. Representación gráfica del LD y bloques de haplotipos seleccionados en la región de *MUC1*, en 1q21, según los datos obtenidos para población CEU en HapMap. Esquema de color: cuadrados en blanco, $D' < 1$ y $LOD < 2$; en azul, $D' = 1$ y $LOD < 2$; en tonos rosa y rojo, $D' < 1$ y $LOD \geq 2$ y en rojo fuerte, $D' = 1$ y $LOD \geq 2$. Se indica la estructura de los genes presentes en la región. B. Bloques de haplotipos y tagSNPs seleccionados (▼). Las X indican aquellos tagSNPs que fueron eliminados o sustituidos. También se indica el porcentaje de posibles haplotipos capturado en cada bloque.

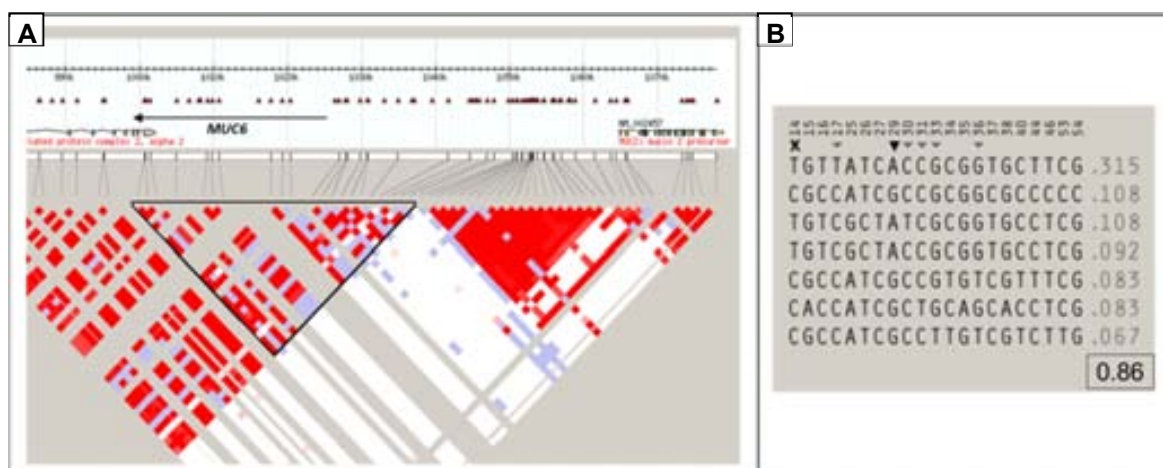


Figura R-17. A. Representación gráfica del LD y bloques de haplotipos seleccionados en la región de *MUC6*, en 11p15, según los datos obtenidos para población CEU en HapMap. Esquema de color: cuadrados en blanco, $D' < 1$ y $LOD < 2$; en azul, $D' = 1$ y $LOD < 2$; en tonos rosa y rojo, $D' < 1$ y $LOD \geq 2$ y en rojo fuerte, $D' = 1$ y $LOD \geq 2$. Se indica la estructura de los genes presentes en la región, excepto para *MUC6*, que no estaba disponible en la versión de HapMap consultada. B. Bloques de haplotipos y tagSNPs seleccionados (▼). Las X indican aquellos tagSNPs que fueron eliminados o sustituidos. También se indica el porcentaje de posibles haplotipos capturado en cada bloque por el conjunto de SNPs seleccionados.

También en 11p15 se analizó una región de 47 kb donde se localiza *MUC2* (29.5 kb). En este caso el gen no queda incluido en un único bloque de LD, ya que existen diversos puntos de recombinación en la región. Se definieron un total de siete bloques con tamaños comprendidos entre 1 y 7 kb (FIGURA R-18A). El primero de ellos, el más compacto y de mayor longitud, abarcaba íntegramente la región genómica a 5' del gen. El resto de bloques incluían polimorfismos localizados propiamente en el gen. La selección final incluyó un total de 15 tagSNPs (FIGURA R-18B) que marcaban haplotipos con una frecuencia superior a 0.05 y el porcentaje de posibles haplotipos cubiertos por la selección fue superior al 90% en todos los casos. El SNP 65 del bloque 3, se sustituyó por el 66, un SNP de cambio de aminoácido S832G (rs11245936) localizado en el mismo bloque. La selección de tagSNPs para el gen *MUC5AC* no se pudo realizar ya que no existe ningún SNP genotipado en HapMap en la región 11p15.5 donde se localiza el gen, que se tuvo que excluir del estudio. Todos los polimorfismos de *MUCs* seleccionados para esta tesis, tanto en el estudio de seguimiento de pacientes con lesiones gástricas como en el estudio caso-control de cáncer gástrico, fueron tagSNPs.

En la tabla TABLA R-3 se detallan los tagSNPs identificados y seleccionados para cada bloque de haplotipos de los genes *MUC*.

Tabla R-3. Selección de tagSNPs en el clúster de *MUCs*

Región	Bloque ^a	Posición Bloque	SNP	Alelos	MAF ^b	Selección
<i>MUC1</i>	1	16	rs3814316	C/T	0.373	Si
<i>MUC1</i>	1	23	rs9426886	A/T	0.458	Si
<i>MUC1</i>	2	27	rs4072037	A/G	0.421	Si
<i>MUC1</i>	2	31	rs2066981	T/C	0.417	Sustituido por 27
<i>MUC1</i>	2	44	rs1045253	C/T	0.325	Si
<i>MUC6</i>	1	14	rs1128413	T/C	0.408	Sustituido por 29
<i>MUC6</i>	1	17	rs4076950	C/T	0.333	Si
<i>MUC6</i>	1	29	rs7481521	A/G	0.408	Si
<i>MUC6</i>	1	30	rs11604757	C/T	0.125	Si
<i>MUC6</i>	1	31	rs11246384	C/T	0.092	Si
<i>MUC6</i>	1	33	rs6597947	G/T	0.075	Si
<i>MUC6</i>	1	36	rs9794921	G/T	0.200	Si
<i>MUC2</i>	1	30	rs10902073	C/A	0.283	Si
<i>MUC2</i>	1	41	rs10794281	T/C	0.350	Si
<i>MUC2</i>	2	57	rs2856082	G/C	0.117	Si
<i>MUC2</i>	2	58	rs2071174	T/C	0.317	Si
<i>MUC2</i>	3	64	rs7396030	C/T	0.183	Si
<i>MUC2</i>	3	65	rs11245935	C/T	0.050	Sustituido por 66
<i>MUC2</i>	3	66	rs11245936	G/A	0.075	Si
<i>MUC2</i>	4	72	rs7944723	C/G	0.200	Si
<i>MUC2</i>	4	74	rs10902089	G/A	0.492	Si
<i>MUC2</i>	4	75	rs6421972	G/A	0.425	Si
<i>MUC2</i>	4	76	rs10794293	C/T	0.367	Si
<i>MUC2</i>	5	77	rs11245954	A/G	0.078	Si
<i>MUC2</i>	5	78	rs7480563	A/A	0.500	Si
<i>MUC2</i>	6	79	rs7126405	A/G	0.100	Si
<i>MUC2</i>	7	83	rs3924453	A/G	0.317	Si
<i>MUC2</i>	7	85	rs4077759	T/C	0.392	Si

a, El número de bloque y la posición detallada hace referencia a las figuras R-16 (*MUC1*), R-17 (*MUC6*) y R-18 (*MUC2*); b, MAF según la versión 24, fase II de HapMap.

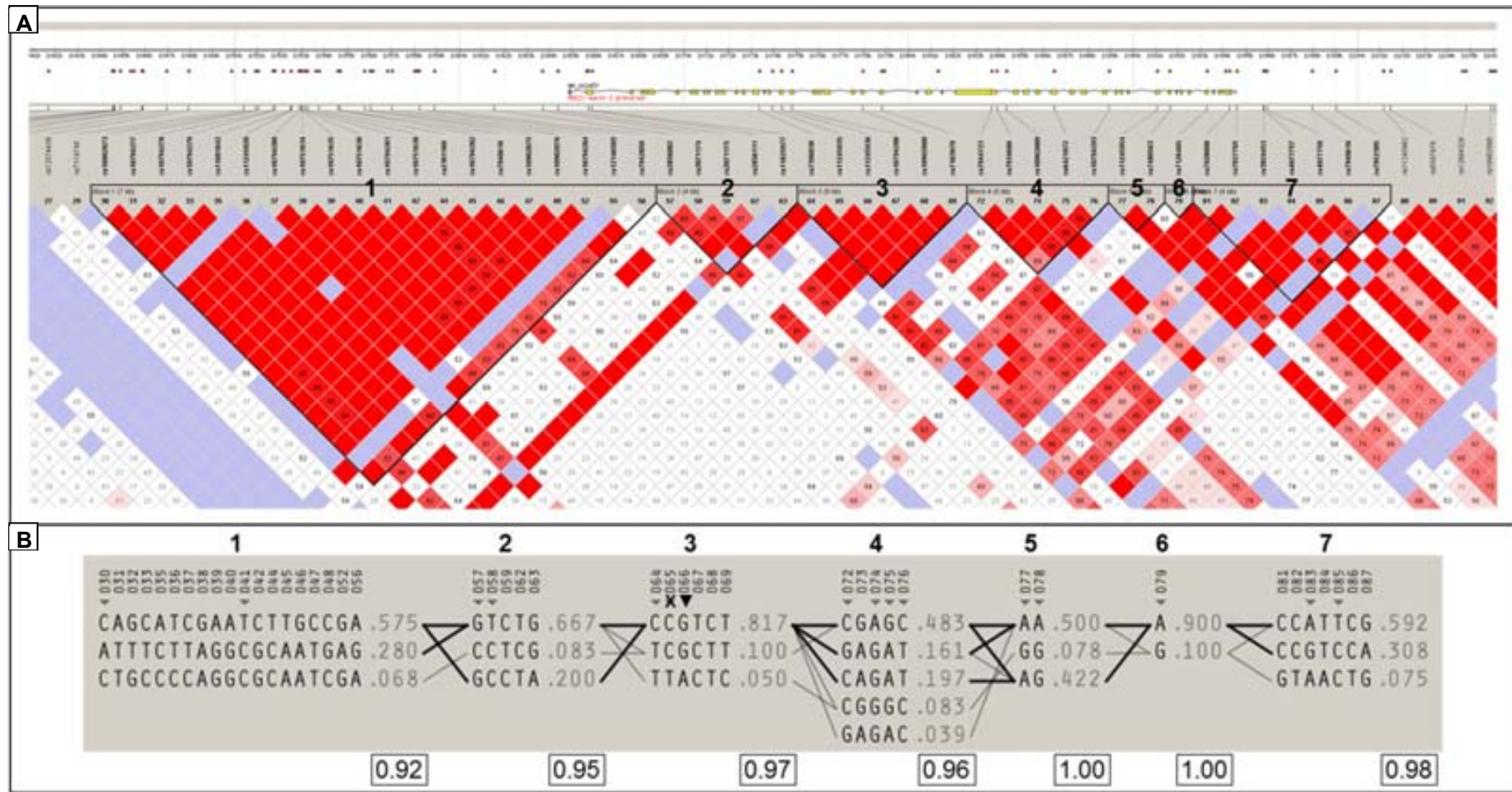


Figura R-18. A. Representación gráfica del LD y bloques de haplotipos seleccionados en la región 11p15 que incluye *MUC2*, según los datos obtenidos para población CEU en HapMap. Esquema de color: cuadrados en blanco, $D' < 1$ y $LOD < 2$; en azul, $D' = 1$ y $LOD < 2$; en tonos rosa y rojo, $D' < 1$ y $LOD \geq 2$ y en rojo fuerte, $D' = 1$ y $LOD \geq 2$. Se indica la estructura del gen *MUC2*. B. Bloques de haplotipos y tagSNPs seleccionados (▼). Las X indican aquellos tagSNPs que fueron eliminados o sustituidos. También se indica el porcentaje de posibles haplotipos capturado en cada bloque.

1.4. Análisis funcional del efecto de polimorfismos en *TFFs*

1.4.1. Análisis funcional *in silico*

Se analizó más detalladamente el posible efecto funcional de algunos de los SNPs identificados en *TFFs* mediante herramientas bioinformáticas.

Mediante el análisis con PupaSuite se identificó el SNP c.-450G>T (rs13052596) en *TFF2* como potencialmente funcional, ya que se localizaba en una posible secuencia consenso para el sitio de unión de NFκB. Se consultó la herramienta is-rSNP, basada en las bases de datos TRANSFAC y JASPAR, para verificar la predicción de PupaSuite y comprobar que otros posibles factores de transcripción podrían ver alterada su unión por la presencia del SNP c.-450G>T. Este algoritmo verificó que el SNP c.-450G>T podría alterar la unión de NFκB, ya que obtuvo el menor valor p de todos los factores de transcripción identificados mediante TRANSFAC ($p=0.0001$). Sin embargo, este mismo factor de transcripción obtuvo un valor p de 0.104 cuando la base de datos utilizada fue JASPAR. Según esta base de datos, el factor de transcripción con mayor probabilidad de interactuar, y verse afectado por la presencia del SNP, fue Sp1 ($p=0.007$) que, además, obtuvo un valor p de 0.002 según TRANSFAC. Seguidamente, se utilizó la herramienta regSNP para comprobar el efecto de la variante en la unión de estos dos factores de transcripción, NFκB y Sp1, y, efectivamente, se observó una diferencia en la puntuación para cada uno de los alelos del SNP c.-450G>T (TABLA R-4), siendo esta diferencia mayor para Sp1.

Tabla R-4. Efecto del SNP c.-450G>T en la interacción de NFκB y Sp1 según regSNP.

Factor de transcripción	Alelo	Secuencia	Puntuación Matriz	Diferencia Puntuación
NFκB	G	GGGACTTTTCC	0.7091	0.1091
	T	TGGACTTTTCC	0.6000	
Sp1	G	AGGGGACTTT	0.7223	0.2854
	T	AGTGGACTTT	0.4369	

A pesar de no haber seleccionado inicialmente el SNP c.85+8G>A (rs225358) de *TFF1* como potencialmente funcional, este polimorfismo se seleccionó como tagSNP. Debido a su localización, a tan sólo 8 pb del primer exón, se decidió comprobar su posible funcionalidad en relación al procesamiento (*splicing*) alternativo. Según la herramienta *Variant Effect Predictor* de Ensembl, no disponible en el momento de la selección, este polimorfismo está descrito como una variante de región de *splicing*. Para analizar más detalladamente el posible efecto de la variante se utilizaron las herramientas FSPLICE y *Human Splicing Finder*. Según FSPLICE el alelo c.85+8A crearía un nuevo sitio donador (gtaagGCatgct, donde a es el alelo menos frecuente del SNP), pero del tipo GC, muy poco frecuente en humanos. La herramienta *Human Splicing Finder* también identificó un nuevo posible sitio donador (AAGgcatgc) aunque obtuvo una puntuación de 1.8, mucho menor que la del sitio donador nativo (5.4). También se analizó otra de las variantes

intrónicas de *TFF1*, c.239-9C>T, identificadas mediante SSCPs pero no se identificaron posibles sitios alternativos de *splicing*.

También se analizó el posible efecto funcional de las variantes de cambio de aminoácido. Mediante el programa PolyPhen se realizó un alineamiento de la secuencia del péptido maduro de TFF3 humano con el de otras especies para comprobar el grado de conservación de la valina correspondiente al polimorfismo c.238G>A (V80M) y, como se puede observar en la FIGURA R-19A, este aminoácido parece estar altamente conservado entre especies. En la FIGURA R-19B se indica la posición del aminoácido, que en la proteína madura se localiza en el residuo 45. El cambio V80M se calificó como seguramente perjudicial en PolyPhen, con una puntuación de 0.999, en un rango de 0 a 1. El programa SIFT también calificó a este cambio como muy perjudicial, con una puntuación de 0.002 (se considera perjudicial a partir de una puntuación menor a 0.05).

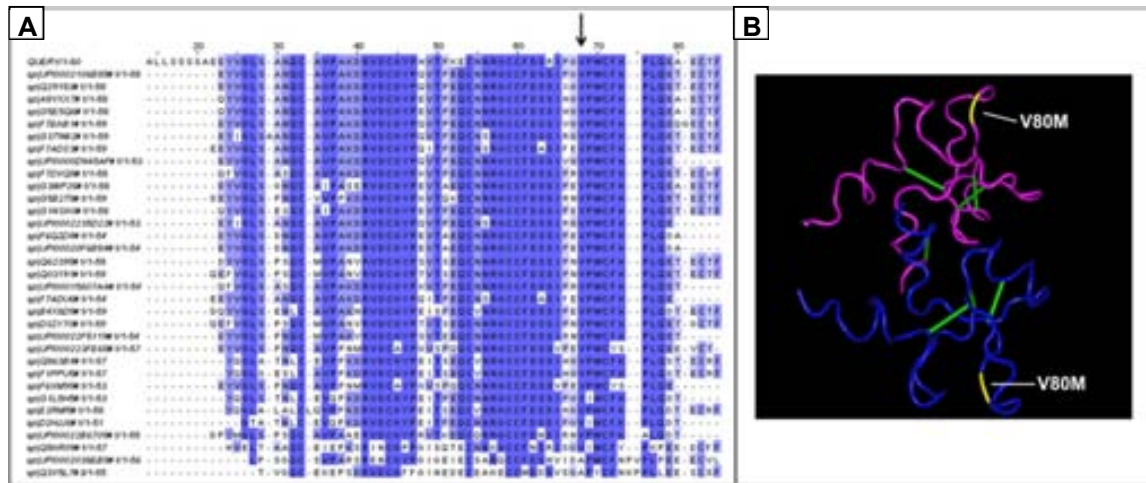


Figura R-19. A. Alineamiento múltiple de la secuencia proteica de TFF3 de diferentes especies. La primera fila muestra la secuencia de TFF3 en humanos. La flecha indica la posición de la valina correspondiente al polimorfismo V80M. El azul intenso es indicativo de un elevado grado de identidad. B. Estructura en 3D de la forma dimérica del péptido TFF3. En azul y fucsia, cada uno de los monómeros de TFF3. En verde se indican los enlaces disulfuro que mantienen la estructura en forma de bucle del péptido. En amarillo se indica la posición de la valina correspondiente al polimorfismo V80M.

Sin embargo, los SNP c.196 G>C (V66L) y c.20A>C (Q7P) de *TFF2*, se calificaron como benignos en PolyPhen, con unas puntuaciones de 0.094 y 0, respectivamente. La herramienta SIFT también calificó la variante Q7P que no forma parte de la proteína madura, como tolerable, con una puntuación de 0.229. Sin embargo, consideró la variante V66L como perjudicial con una puntuación de 0.008, sensiblemente menor al límite establecido de 0.005.

El efecto funcional de la mayoría de polimorfismos de 3'UTR de *TFFs* (*TFF1* c.*41G>A, *TFF2* c.*68_69InsCTT, *TFF2* c.*82T>C y *TFF3* c.*41G>A) y su posible interacción con miRNAs (ver APARTADO 1.4.1.1) se analizó más detalladamente mediante ensayos luciferasa (ver APARTADO 1.4.2) ya que estos polimorfismos fueron los primeros en validarse o identificarse en una población control y podían ser testados utilizando la misma estrategia experimental. No obstante, se realizó un alineamiento múltiple de las secuencias de la región 3'UTR de los tres genes *TFF* de humanos y también de la región 3'UTR de *TFF1* de diferentes especies con ClustalW

ya que estos genes presentan importantes similitudes de secuencia y variabilidad en la región. Como se observa en la FIGURA R-20A, la secuencia 3'UTR de *TFF1* presenta un elevado grado de identidad entre especies, especialmente en el extremo 3'. La posición correspondiente al SNP c.*41G>A en el alineamiento presenta cierto grado de indentidad FIGURA R-20A, y todas las especies presentan un nucleótido G, la mayoría, o A en dicha posición. Las regiones 3'UTR de los tres miembros de *TFF*, presentan una identidad menor entre sus secuencias, aunque tienen un 50%, aproximadamente, de similitud entre ellas. En la figura FIGURA R-20B se indica la posición relativa en el alineamiento de los SNPs de 3'UTR identificados.

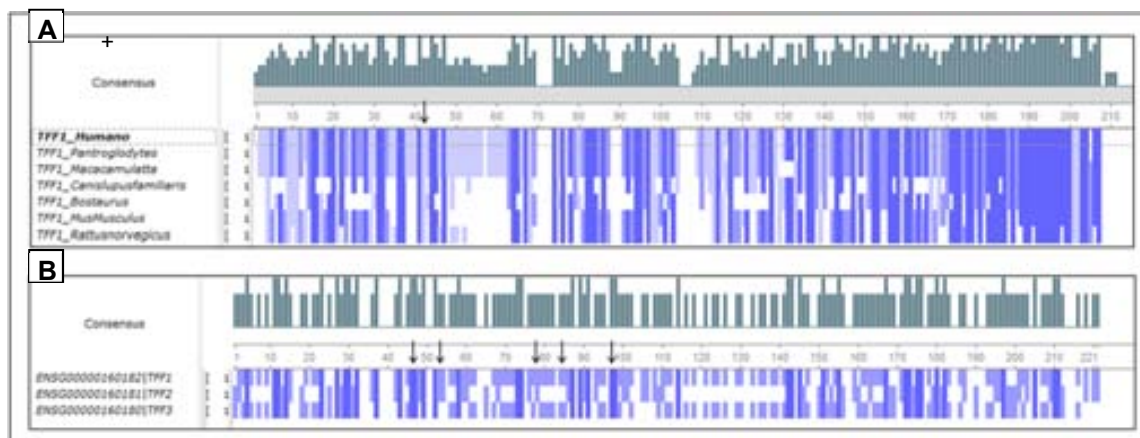


Figura R-19. A. Alineamiento múltiple de la secuencia nucleotídica de la región 3'UTR de *TFF1* de diferentes especies. La primera fila muestra la secuencia de *TFF1* en humanos. La flecha indica la posición relativa del SNP c.*41G>A. El azul intenso es indicativo de un elevado grado de identidad. B. Alineamiento múltiple de las secuencias nucleotídicas de las regiones 3'UTR de los tres genes *TFF*. El azul intenso es indicativo de un elevado grado de identidad. Las flechas indican la posición relativa de los polimorfismos de 3'UTR de *TFFs* en el siguiente orden: c.*41G>A de *TFF3*, c.*41G>A y c.*62C>T de *TFF1*, c.*68_69InsCTT y c.*82C>T de *TFF2*.

1.4.1.1 Predicción de sitios diana de miRNAs en la región 3'UTR de *TFFs*

Se realizó una selección de microRNAs que tuvieran como posible diana las secuencias de la región 3'UTR de genes *TFF*. Se utilizaron diversos programas de predicción de dianas de miRNAs (APARTADO 2.2 de metodología) y la premisa principal para la selección fue que interaccionaran específicamente en la región donde se localizaban los polimorfismos identificados. Dos de las herramientas consultadas no generaron ningún resultado para estos genes: TarBase y Pic-Tar. TarBase es una base de datos que engloba una serie de dianas de miRNAs confirmadas experimentalmente por lo que se deduce que, hasta ahora, dicha confirmación no se ha reportado para *TFFs*. Pic-Tar es una herramienta limitada ya que los datos de entrada, en los que se basa su predicción, son un alineamiento múltiple de secuencias ortólogas 3'UTR y un grupo predefinido de miRNAs maduros. Aunque las aplicaciones DIANA-micro-T y MicroRNA.org sí que identificaron miRNAs que probablemente interaccionaran en la región 3'UTR de genes *TFF*, ninguno de ellos se localizó en el sitio correspondiente a las variantes identificadas, por lo que sus resultados no se tuvieron en cuenta. Finalmente, la selección se realizó a partir de los resultados obtenidos con las aplicaciones de TargetScan, MicroInspector y MicroCosm que fueron muy diversos en cuanto al número de miRNAs predichos.

Mediante targetScan se identificó, como máximo, un miRNA por cada uno de los polimorfismos identificados mientras que con MicroInspector se llegaron a predecir hasta 15 miRNAs que interaccionarían en la región donde se localizaba un único SNP, en este caso el c.*41G>A de *TFF1* (rs13276). Se seleccionaron todos los miRNAs que fueron identificados por, al menos, dos de estas aplicaciones: hsa-miR-92b en c.*41G>A de *TFF1*, hsa-miR-671-5p en c.*82T>C de *TFF2* (rs225334) y hsa-miR-615-3p c.*41G>A de *TFF3* (TABLA R-5).

Tabla R-5. Selección de miRNAs en la región 3'UTR polimórfica de los genes TFF.

Gen	Polimorfismo	hsa-miR	Diferencias alelos ^a	Bases de datos ^b			Otros SNPs ^c
				Target Scan	Micro Inspector	Micro Cosm	
<i>TFF1</i>	c.*41G>A	92b	Si		X	X	No
		652	Si			X	No
		760	Si		X		Si
<i>TFF2</i>	c.*67_68InsCTT	524-3p	Si	X			No
		127-5p	Si			X	Si
	c.*82T>C	671-5p	Si		X	X	No
		127-5p	Si			X	Si
<i>TFF3</i>	c.*41G>A	615-3p	Si	X		X	No

a, Diferencia entre alelos de cada SNP en la interacción de los miRNAs, según miRanda; b, Las X indican que el miRNA interacciona en la región del SNP indicado según la predicción de la herramienta utilizada en cada caso. c, El miRNA indicado interacciona con otros SNPs identificados en la región 3'UTR de *TFF*.

Además, se incluyeron miRNAs que interaccionaran también con otros SNPs en el mismo gen o en otros genes *TFF*. Como se observa en la TABLA R-5 algunos de los SNPs ya seleccionados presentaron esta característica. El hsa-miR-760, que está predicho interaccionar en la secuencia donde se localiza el *TFF1* c.*41G>A, se seleccionó inicialmente porque también interaccionaba con el c.*62G>A, en el mismo gen. Este último SNP, al no validarse en la población de pacientes con lesiones gástricas, no se incluyó finalmente en el análisis funcional mediante ensayos luciferasa. En el caso de la inserción de *TFF2*, al no encontrarse miRNAs en más de una aplicación, se seleccionó el hsa-miR-524, único miRNA identificado con TargetScan, y hsa-miR-127-5p que también interaccionaba con la variante c.*82T>C en el mismo gen. Además se seleccionó el hsa-miR-652 que está predicho interaccionar en la zona donde se localiza el *TFF1* c.*41G>A, ya que se también se identificó en el análisis mediante el programa PupaSuite, utilizado para determinar el posible efecto funcional de las variantes en la selección de SNPs potencialmente funcionales (APARTADO 1.1 de resultados).

El requisito indispensable que debían cumplir todos los miRNAs seleccionados fue que su interacción con la región de interés estuviera alterada por la presencia de las variantes localizadas en el sitio de unión (TABLA R-5). Para comprobar esta premisa se compararon las puntuaciones de complementariedad (*score*) para cada alineamiento y la energía libre de la interacción entre las dos cadenas miRNA-UTR obtenidas con el algoritmo miRanda (APARTADO 2.2 de metodología) para cada uno de los alelos a testar (FIGURA R-20) y se seleccionaron únicamente aquellos que presentaron diferencias en ambos parámetros.

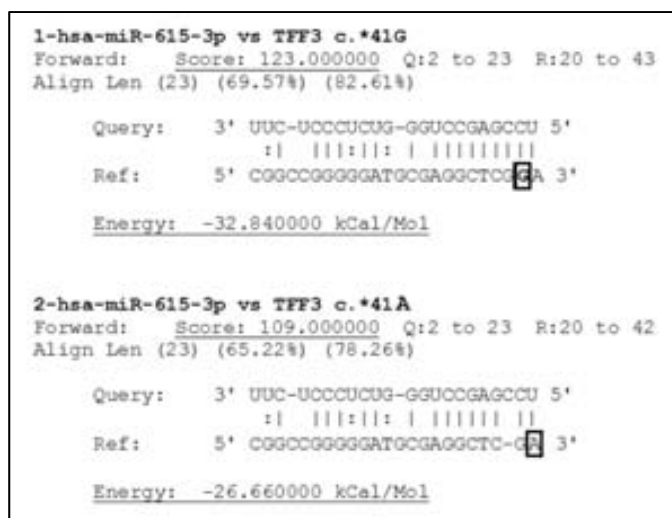


Figura R-20. Análisis de la interacción del miRNA hsa-miR-615-3p con la región 3'UTR de *TFFs* donde se localiza el SNP c.*41G>A. Alineamiento y energía libre de la interacción entre 1- hsa-miR-615-3p y *TFF3* c.*41G y 2- hsa-miR-615-3p y *TFF3* c.*41A

1.4.2. Análisis funcional de los polimorfismos de la región 3'UTR de *TFFs*

1.4.2.1 Efecto de los polimorfismos en la expresión de la luciferasa

Mediante ensayos luciferasa se analizó el efecto que podía tener cada uno de los alelos de los polimorfismos identificados en la región 3'UTR de *TFFs* sobre la expresión del gen reportero. Para ello se transfectaron en células HEK-293T (*Human Embryonic Kidney 293*) y AGS (procedentes de un adenocarcinoma gástrico) las construcciones con las regiones 3'UTR, con cada uno de los alelos de las diversas variantes, clonadas en el vector pGL4 que contenía la luciferasa *Firefly* (APARTADO 8 y 13.2 de materiales y métodos). Conjuntamente con los vectores reporteros se transfectó el vector pGL4.74 que expresa constitutivamente la luciferasa Renilla, utilizado como normalizador de los resultados. Para cada experimento, se calculó la media y desviación estándar de los datos de luminiscencia relativa Firefly/Renilla obtenida en cada una de las tres réplicas y se calculó un ratio con la media obtenida para la construcción que portara el alelo/haplotipo más frecuente de cada gen. Posteriormente, se hizo la media de los ratios obtenidos en, como mínimo, tres días diferentes y se calculó el error estándar para cada una de ellas. En las FIGURAS R-21 y R-22 se muestran los resultados obtenidos en células HEK-293T y AGS, respectivamente. En células HEK-293T se puede observar que tanto el alelo C del SNP c.*82T>C, como la inserción de 3 pb c.*67_68insCTT de *TFF2* provocan un aumento significativo de expresión del gen de luciferasa respecto al haplotipo más frecuente c.*82T sin inserción (FIGURAS R-21). Además, el efecto causado por el alelo C es independiente de la presencia la inserción ya que los haplotipos c.*82C (sin inserción) y c.*82C InsCTT presentan la misma expresión. Por el contrario, las variantes c.*41G>A de *TFF1* y *TFF3* provocaron una reducción de la expresión de la

luciferasa que fue muy significativa en el caso de *TFF1* y mucho menor en *TFF3*. En este último caso no se observaron diferencias significativas aunque por poco (t-test $p=0.069$).

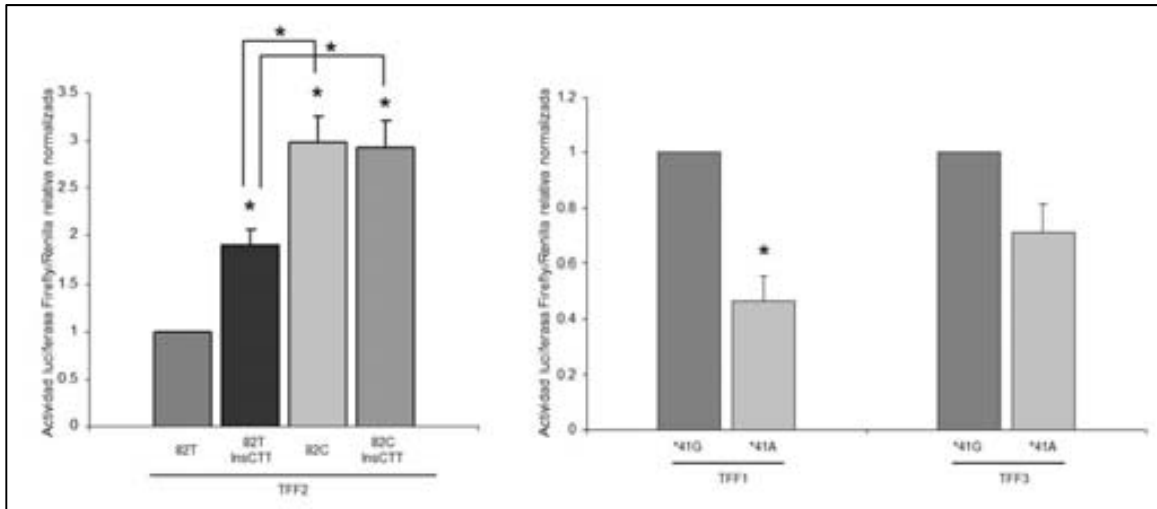


Figura R-21. Efecto de los SNPs en 3'UTR de genes *TFF*, en la expresión de la luciferasa en células HEK-293T. Los ratios de la luminiscencia de Firefly y Renilla se presentan después de la normalización. Los datos están reportados como ratios de las medias \pm SEM de todos los experimentos realizados (un mínimo de 3), en referencia al genotipo más frecuente. Los asterísticos indican un valor $p<0.05$ en el t-test para muestras independientes.

Cuando se realizó el análisis en células AGS, específicas de cáncer gástrico, se observaron resultados muy similares (FIGURA R-22).

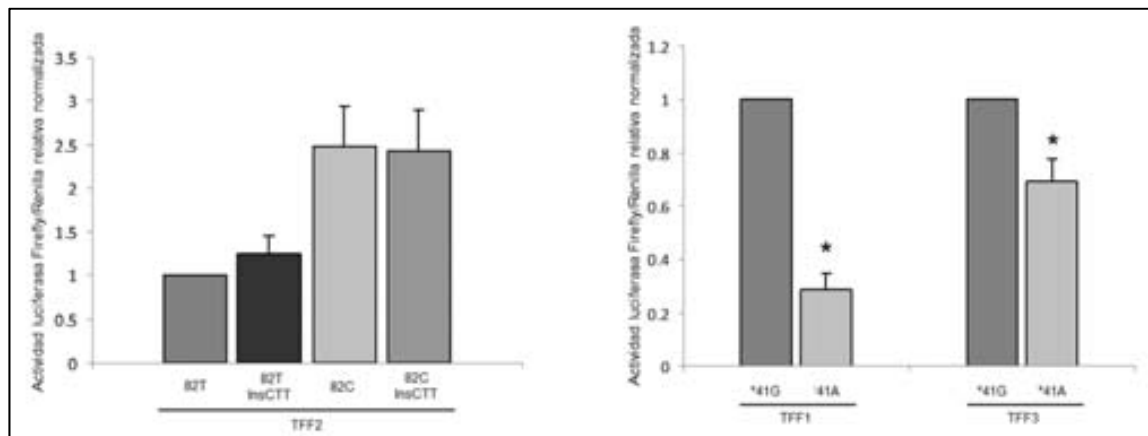


Figura R-22. Efecto de los SNPs en 3'UTR de genes *TFF*, en la expresión de la luciferasa en células AGS, de adenocarcinoma gástrico. Los ratios de la luminiscencia de Firefly y Renilla se presentan después de la normalización. Los datos están reportados como ratios de las medias \pm SEM de todos los experimentos realizados (un mínimo de 3), en referencia al genotipo más frecuente. Los asterísticos indican un valor $p<0.05$ en el t-test para muestras independientes.

En el caso de *TFF2*, el alelo C del SNP situado en c.*82 causó de nuevo un aumento de la expresión muy significativo, sin hallarse tampoco diferencias en función de la presencia de la otra variante. En este caso no se observó ningún efecto de la inserción ya que la expresión para el alelo T con la inserción fue prácticamente idéntico al alelo T sin la inserción. A pesar de las

diferencias observadas, el test de comparación de medias no fue significativo probablemente debido a la desviación y error estándar de los ratios. Por otro lado, la reducción de la expresión provocada por el SNP de c.*41G>A en *TFF1* y en *TFF3* fue prácticamente idéntico en células AGS. En el caso de *TFF1* fue incluso más acusado y en ambos genes fue significativo ($p < 0.05$)

1.4.2.2 Interacción de miRNAs con *TFFs* en la región 3'UTR

Se realizaron ensayos luciferasa para analizar si los polimorfismos de 3'UTR de *TFFs* modificaban el efecto silenciador de los miRNAs previamente seleccionados. Para ello se clonaron los miRNAs en el vector de expresión pSIL y se contrasfectaron con los vectores de expresión pGL4 que contenían cada uno de los alelos de las variantes de la región 3'UTR en células HEK-293T. Los miRNAs hsa-miR-92b y 127-5p no se pudieron clonar con éxito debido a que las secuencias obtenidas de la amplificación con los cebadores diseñados específicamente a partir de las secuencias publicadas en miRBase para el pre-miRNA no coincidieron (hsa-miR-92b) o la ligación de los oligos adaptadores que se utilizaron para la clonación en la diana BmHI de pSIL no funcionaron (hsa-miR-127-5p). En el caso de miR-524 se utilizó una molécula de RNA de doble cadena comercial (miRIDIAN, APARTADO 2.2 de materiales y métodos) que mimetiza a los microRNAs nativos. También se transfectó en todos los experimentos el vector que expresa constitutivamente la luciferasa Renilla que se utiliza como normalizadora de los resultados. Tal y como hay se ha descrito en el APARTADO 3.2.1, se hizo la media de los ratios obtenidos, en relación a la construcción que portara el alelo/haplotipo más frecuente, en, como mínimo, tres días diferentes y se calculó el error estándar para cada una de ellas. Los miRNAs 524 y 671 aumentaron ligeramente la expresión de la luciferasa debido a su interacción con la región 3'UTR de *TFF2* (FIGURA R-23). Respecto el miR-524, se observó un aumento similar para los dos alelos, con o sin inserción, del polimorfismo c*68_69insCTT aunque sólo fue significativo para el alelo que no contenía la inserción (InsCTT $p=0.088$, no InsCTT $p=0.006$), probablemente debido a su menor error estándar. En el caso del miR-671, no se observó interacción con el alelo T del SNP c.*82T>C ya que la expresión en el control negativo y el microRNA fue prácticamente idéntica. Sin embargo, si se observó un ligero aumento de expresión con el alelo C, que fue significativo en el haplotipo que forma con la inserción, y casi significativo en el haplotipo que no la lleva.

Para la región 3'UTR de *TFF1* se analizaron dos miRNAs, el miR-652 y el miR-760. Para el miR-652 no se observó ni interacción con la región 3'UTR, determinada por la diferencia de expresión de la luciferasa, ni diferencias entre alelos en esa interacción. Sin embargo, para el miR-760 se observó una disminución de la expresión de la luciferasa para el alelo más común del SNP c.*41G>A ($p=0.066$) (FIGURA R-23). Este efecto del miRNA para el alelo G es poco concluyente debido a la amplitud de la desviación estándar. Finalmente, el alelo A del SNP c.*41G>A de *TFF3* también provocó una disminución de la expresión de la luciferasa casi significativa ($p=0.088$) debido a la interacción con el miR-615 (FIGURA R-23).

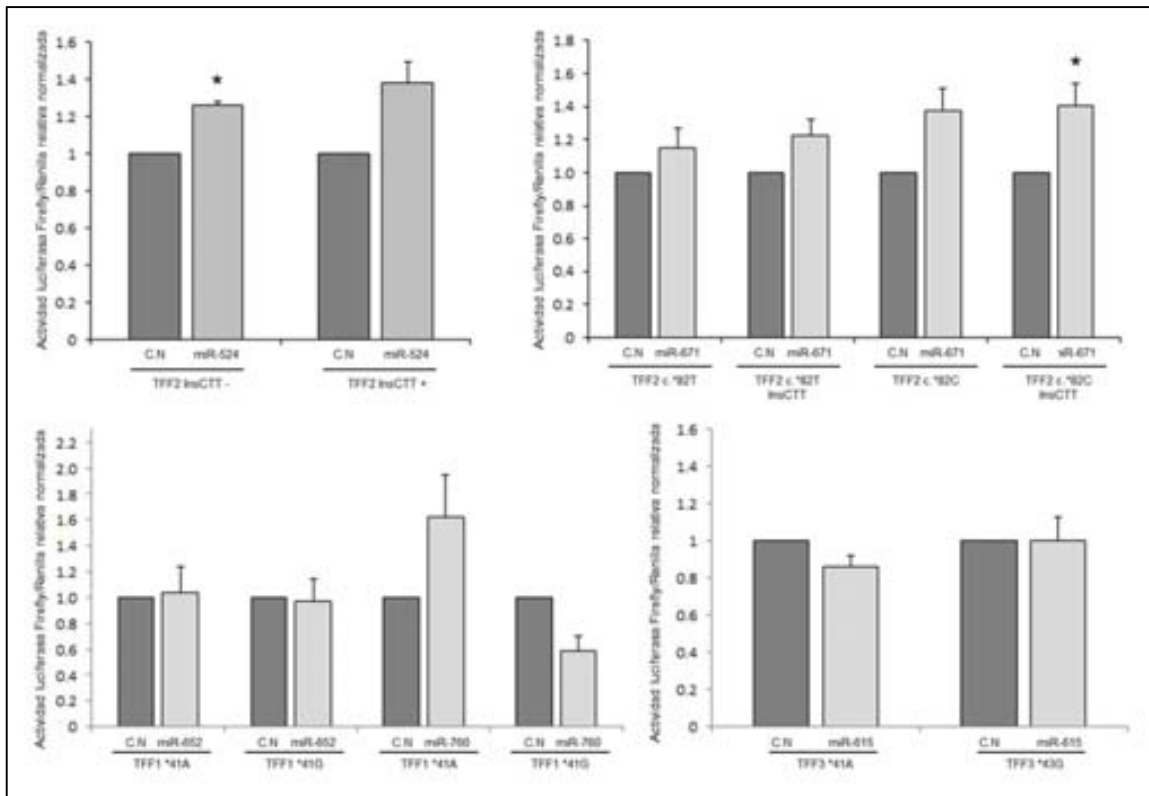


Figura R-23. Efecto de los miRNAs, en función de los SNPs en 3'UTR de los genes *TFF*, en la expresión de la luciferasa en células HEK-293T. Los ratios de la luminiscencia de *Firefly* y *Renilla* se presentan después de la normalización. Los datos están reportados como medias \pm SEM de estos ratios en referencia al vector control pSIL (CN o Control Negativo). Los asteriscos indican un valor $p < 0.05$ en el t-test entre el ensayo con el miRNA y el control negativo.

1.5. Polimorfismos seleccionados para los estudios de asociación

Las principales características y localización de todos los SNPs en genes *MUC* y *TFF* (tagSNPs y/o SNPs potencialmente funcionales) seleccionados en esta tesis se muestra en la TABLA R-6 del apéndice, donde también se detalla el estudio en el cual se incluyeron y las técnicas de genotipación utilizadas en cada caso: PCR, sondas de hibridación FRET, sondas de hibridación TaqMan, HRM y/o MassArray de Sequenom (ver APARTADO 4 y 7 de materiales y métodos).

Se seleccionaron inicialmente un total de 109 polimorfismos para los análisis de asociación en los estudios de evolución en pacientes con lesión gástricas y caso-control de cáncer gástrico: 25 en *MUCs*, 33 en *TFFs*, 27 en *ILs* y 24 en genes de reparación del DNA, inflamación y de respuesta a la infección por *H. pylori* (TABLA R-6).

En el caso de *TFFs*, se incluyeron tanto polimorfismos potencialmente funcionales (5 en *TFF1*, 7 en *TFF2* y 2 en *TFF3*), validados o identificados en este estudio, como los tagSNPs (8 en *TFF1*, 4 en *TFF2*, y 7 en *TFF3*) seleccionados a partir del análisis de desequilibrio de ligamiento e identificación de haplotipos. El rs225334 de *TFF2*, en la región 3'UTR, y el rs11701143 de *TFF3*,

en 5'UTR, son polimorfismos funcionales que se priorizaron como tagSNPs en la selección y se han contabilizado, por tanto, como tales. Uno de los cinco SNPs potencialmente funcionales de *TFF1*, el rs34795821 (T22I), se incluyó posteriormente al revisar la base de datos dbSNP, en su versión más actual. Como se indica en el apartado anterior (APARTADO 1.3.2 de resultados), la totalidad de polimorfismos en genes *MUC* incluidos en este estudio fueron tag SNPs (4 en *MUC1*, 15 en *MUC2* y en 6 *MUC6*). Referente a polimorfismos en otros genes involucrados en la carcinogénesis gástrica (ver APARTADO 2.1 de materiales y métodos), se realizó una búsqueda bibliográfica y se seleccionaron polimorfismos para los cuales ya había antecedentes de su asociación con cáncer gástrico, lesiones precursoras o procesos relacionados. Además, en el caso de las interleucinas, se incluyeron otros SNPs adicionales localizados en las posibles regiones de interés, previamente identificadas en la bibliografía, y que formaran parte del mismo bloque de LD. En esta tesis, estos polimorfismos sólo se incluyeron en estudio de asociación con la evolución de lesiones gástricas (TABLA R-6).

Tabla R-6. Características generales de todos los polimorfismos incluidos en esta tesis en los genes *MUCs*, *TFFs*, *ILs* y genes relacionados con la infección por *H.pylori*.

Polimorfismo	Gen	Cambio ^a	Tipo ^b	Cr	Posición ^c	Tag SNP	Técnica genotipado ^d	Estudio ^e
rs3814316	<i>TRIM46</i>	C/T	T287T	1	153416342	Si	Seq, Illu	1,2
rs9426886	<i>TRIM46</i>	A/T	Intrónico	1	153418378	Si	Seq, Illu	1,2
rs4072037	<i>MUC1</i>	A/G	T31T	1	153428691	Si	Seq, Illu	1,2
rs1045253	<i>MUC1</i>	C/T	5'-up	1	153467859	Si	Seq, Illu	1,2
rs4076950	<i>MUC6</i>	C/T	Intrónico	11	1005021	Si	Seq, Illu	1,2
rs7481521	<i>MUC6</i>	A/G	V619M	11	1017811	Si	Seq, Illu	1,2
rs11604757	<i>MUC6</i>	C/T	G395R	11	1019320	Si	Seq, Illu	1,2
rs11246384	<i>MUC6</i>	C/T	Intrónico	11	1020448	Si	Seq, Illu	1,2
rs6597947	<i>MUC6</i>	G/T	5'-up	11	1027029	Si	Seq, Illu	1,2
rs9794921	<i>MUC6</i>	G/T	5'-up	11	1029867	Si	Seq, Illu	1,2
rs10902073	<i>MUC2</i>	C/A	5'-up	11	1050934	Si	Seq, Illu	1,2
rs10794281	<i>MUC2</i>	T/C	5'-up	11	1053149	Si	Seq, Illu	1,2
rs2856082	<i>MUC2</i>	C/G	5'-up	11	1061562	Si	Seq, Illu	1,2
rs2071174	<i>MUC2</i>	T/C	5'-up	11	1063712	Si	Seq, Illu	1,2
rs7396030	<i>MUC2</i>	C/T	Intrónico	11	1073364	Si	Seq	1
rs11245936	<i>MUC2</i>	G/A	G832S	11	1074362	Si	Seq, Illu	1,2
rs7944723	<i>MUC2</i>	C/G	P1832P	11	1083710	Si	Seq, Illu	1,2
rs10902089	<i>MUC2</i>	A/G	Intrónico	11	1084357	Si	Seq, Illu	1,2
rs6421972	<i>MUC2</i>	G/A	I2154T	11	1086494	Si	Seq, Illu	1,2
rs10794293	<i>MUC2</i>	C/T	Intrónico	11	1088939	Si	Seq, Illu	1,2
rs11245954	<i>MUC2</i>	A/G	V2459V	11	1091078	Si	Seq, Illu	1,2
rs7480563	<i>MUC2</i>	A/G	T2524P	11	1091649	Si	Seq, Illu	1,2
rs7126405	<i>MUC2</i>	A/G	Q2653P	11	1093296	Si	Seq, Illu	1,2
rs3924453	<i>MUC2</i>	G/A	3'-down	11	1095806	Si	Seq, Illu	1,2
rs4077759	<i>MUC2</i>	T/C	3'-down	11	1095976	Si	Seq, Illu	1,2
rs8133510	<i>TFF3</i>	G/A	3' down	21	42603048	Si	Seq, Illu	1,2
c.*41G>A ^f	<i>TFF3</i>	G/A	3'UTR	21	42605394	No	FRET	1,2
rs2236705	<i>TFF3</i>	C/A	Intrónico	21	42605897	Si	Seq, Illu	1,2
V80M ^f	<i>TFF3</i>	G/A	V80M	21	42606697	No	FRET, Illu	1,2
rs11701143	<i>TFF3</i>	A/G	5'UTR	21	42608598	Si	FRET, Seq, Illu	1,2

Polimorfismo	Gen	Cambio ^a	Tipo ^b	Cr	Posición ^c	Tag SNP	Técnica genotipado ^d	Estudio ^e
rs225363	<i>TFF3</i>	G/C	5' up	21	42610443	Si	Seq, Illu	1,2
rs9981660	<i>TFF3</i>	G/A	5' up	21	42611007	Si	Seq, Illu	1,2
rs692356	<i>TFF3</i>	A/G	5' up	21	42613698	Si	Seq, Illu	1,2
rs11203197	<i>TFF3</i>	A/G	5' up	21	42615302	Si	Illu	2
rs12483635	<i>TFF2</i>	A/G	3'down	21	42639446	Si	Illu	2
rs4920084	<i>TFF2</i>	C/T	3'UTR	21	42639551	No	FRET	1
rs225334	<i>TFF2</i>	T/C	3'UTR	21	42639629	Si	Seq, FRET, HRM	1,2
c.*67_68insCTT ^f	<i>TFF2</i>	-/InsCTT	3'UTR	21	42639641	No	HRM	1,2
VNTR	<i>TFF2</i>	A/B	Intrónico	21	42640000	No	PCR	1
V66L ^f	<i>TFF2</i>	G/C	V66L	21	42643092	No	FRET, Illu	1,2
rs1079380	<i>TFF2</i>	A/G	Intrónico	21	42643352	Si	Seq, Illu	1,2
Q7P ^f	<i>TFF2</i>	A/C	Q7P	21	42644116	No	FRET	1
rs7277409	<i>TFF2</i>	G/A	R3W	21	42644129	No	Seq, Illu	1,2
rs13052596	<i>TFF2</i>	G/T	5' up	21	42644585	No	FRET	1
rs3814896	<i>TFF2</i>	A/G	5' up	21	42644780	Si	Seq, Illu	1,2
c.*62C>T ^f	<i>TFF1</i>	C/A	3'UTR	21	42655595	No	FRET	1
rs13276	<i>TFF1</i>	G/A	3'UTR	21	42655616	No	FRET, Illu	1,2
rs225353	<i>TFF1</i>	C/G	Intrónico	21	42656124	Si	Seq, Illu	1,2
rs225356	<i>TFF1</i>	G/A	Intrónico	21	42659137	Si	Illu	2
rs2839488	<i>TFF1</i>	G/C	Intrónico	21	42659255	Si	Illu	2
rs225357	<i>TFF1</i>	C/T	Intrónico	21	42659382	Si	Seq	1
rs4920094	<i>TFF1</i>	G/A	Intrónico	21	42659402	Si	Seq	1
rs13051704	<i>TFF1</i>	G/C	Intrónico	21	42659440	Si	Seq, Illu	1,2
rs225358	<i>TFF1</i>	T/C	Intrónico	21	42659581	Si	Seq, Illu	1,2
rs34795821	<i>TFF1</i>	C/T	T22I	21	42659609	No	Seq	1
rs2156310	<i>TFF1</i>	C/T	5'UTR	21	42659675	No	FRET	1
c.-29C>T ^f	<i>TFF1</i>	G/A	5'UTR	21	42659703	No	FRET	1,2
rs9976977	<i>TFF1</i>	G/A	5' up	21	42662543	Si	Seq, Illu	1,2
rs424694	<i>TFF1</i>	C/T	5' up	21	42662841	Si	Seq, Illu	1,2
rs13047838	<i>TMPRSS3</i>	C/T	3'UTR	21	42665938	Si	Seq, Illu	1,2
rs1800872	<i>IL10</i>	G/T	5' up	1	205013030	No	TaqMan	1
rs1800896	<i>IL10</i>	T/C	5' up	1	205013520	No	TaqMan	1
rs2856836	<i>IL1A</i>	A/G	3'UTR	2	113248554	No	TaqMan	1
rs3783546	<i>IL1A</i>	C/G	Intrónico	2	113251301	No	TaqMan	1
rs17561	<i>IL1A</i>	C/A	A114S	2	113253694	No	TaqMan	1
rs1800587	<i>IL1A</i>	G/A	5'UTR	2	113259431	No	TaqMan	1
rs1143634	<i>IL1B</i>	G/A	F105F	2	113306861	No	TaqMan	1
rs1143627	<i>IL1B</i>	A/G	5' up	2	113310858	No	TaqMan	1
rs16944	<i>IL1B</i>	G/A	5' up	2	113311338	No	TaqMan	1
rs1143623	<i>IL1B</i>	C/G	5' up	2	113312300	No	TaqMan	1
rs4251961	<i>IL1RN</i>	T/C	5' up	2	113590938	No	TaqMan	1
rs2637988	<i>IL1RN</i>	A/G	Intrónico	2	113593250	No	TaqMan	1
rs4251985	<i>IL1RN</i>	G/T	Intrónico	2	113593884	No	TaqMan	1
rs4073	<i>IL8</i>	T/A	5' up	4	74824888	No	TaqMan	1
rs2243250	<i>IL4</i>	C/T	5' up	5	132037053	No	TaqMan	1
rs2070874	<i>IL4</i>	C/T	5'UTR	5	132037609	No	TaqMan	1
rs2227284	<i>IL4</i>	G/T	Intrónico	5	132040624	No	TaqMan	1
rs3181226	<i>IL12B</i>	C/G	3' down	5	158673108	No	TaqMan	1
rs3181224	<i>IL12B</i>	T/G	3' down	5	158673428	No	TaqMan	1
rs1368439	<i>IL12B</i>	T/G	3'UTR	5	158674592	No	TaqMan	1
rs3212227	<i>IL12B</i>	A/C	3'UTR	5	158675528	No	TaqMan	1
rs1800797	<i>IL6</i>	G/A	5' up	7	22732746	No	TaqMan	1

Polimorfismo	Gen	Cambio ^a	Tipo ^b	Cr	Posición ^c	Tag SNP	Técnica genotipado ^d	Estudio ^e
rs2057768	<i>IL4R</i>	C/T	5' up	16	27229596	No	TaqMan	1
rs1805010	<i>IL4R</i>	A/G	I75L	16	27263704	No	TaqMan	1
rs1805011	<i>IL4R</i>	A/C	E400A	16	27281373	No	TaqMan	1
rs1805013	<i>IL4R</i>	C/T	S436L	16	27281481	No	TaqMan	1
rs1801275	<i>IL4R</i>	A/G	Q576R	16	27281901	No	TaqMan	1
rs28362491	<i>NFKB1</i>	A/B	5' NFKB1	4	103641187	No	Seq	1
rs5743708	<i>TLR2</i>	G/A	Q753Q	4	154845767	No	Seq	1
rs2569190	<i>CD14</i>	A/G	5'UTR	5	139993100	No	Seq	1
rs5744455	<i>CD14</i>	C/T	5' CD14	5	139993491	No	Seq	1
rs909253	<i>LTA</i>	T/C	Intrónico	6	31648292	No	TaqMan	1
rs1800629	<i>TNF</i>	G/A	5' TNF	6	31651010	No	TaqMan	1
rs361525	<i>TNF</i>	G/A	5' TNF	6	31651080	No	TaqMan	1
rs2234711	<i>IFNGR1</i>	T/C	5'UTR	6	137582213	No	Seq	1
rs6472812	<i>LY96</i>	G/A	G56R	8	75079638	No	Seq	1
rs11466004	<i>LY96</i>	C/T	P157S	8	75103829	No	Seq	1
rs2976392	<i>PSCA</i>	G/A	Intrónico	8	143759934	No	Seq	1
rs4986790	<i>TLR4</i>	A/G	D299G	9	119515123	No	Seq	1
rs4986791	<i>TLR4</i>	C/T	T399I	9	119515423	No	Seq	1
rs2430561	<i>IFNG</i>	A/T	Intronic	12	66838787	No	Seq	1
rs2301756	<i>PTPN11</i>	T/C	Intrónico	12	111375159	No	Seq	1
rs2066842	<i>NOD2</i>	C/T	P268S	16	49302125	No	Seq	1
rs2066844	<i>NOD2</i>	C/T	R702W	16	49303427	No	Seq	1
rs16260	<i>CDH1</i>	A/C	5' up	16	67328535	No	Seq	1
rs1125557	<i>CDH1</i>	A/G	Intrónico	16	67367100	No	Seq	1
rs2297518	<i>NOS2A</i>	G/A	S608L	17	23120724	No	Seq	1
rs25487	<i>XRCC1</i>	G/A	Q399R	19	48747566	No	FRET	1
rs1799782	<i>XRCC1</i>	C/T	R194W	19	48749414	No	FRET	1
rs13181	<i>ERCC2</i>	A/C	K751Q	19	50546759	No	FRET	1
rs1799793	<i>ERCC2</i>	C/T	D312N	19	50559099	No	FRET	1

a, El alelo más común en población de origen europeo se indica primero; b, Localización del SNP relativa a cada gen; c, Posición cromosómica correspondiente al *Genome Build 36.3*; d, Técnicas de genotipado: Seq (MassArray de Sequenom), I (Illumina GoldenGate), FRET (sondas de hibridación FRET), TaqMan (sondas de hidrólisis TaqMan), HRM (High Resolution Melting), PCR (reacción en cadena de la polimerasa); e, 1- Seguimiento de pacientes con lesiones gástricas y 2- Caso-control de cáncer gástrico anidado en la cohorte EPIC; f, Polimorfismos potencialmente funcionales identificados en este estudio.

2. ASOCIACIÓN DE POLIMORFISMOS DE TFFS Y MUCS CON LA CARCINOGENÉISIS GÁSTRICA

Una vez finalizado el análisis de la variabilidad genética en *TFFs* y *MUCs*, que incluía la búsqueda de SNPs potencialmente funcionales en los genes *TFF* y *MUC1* y la selección de tagSNPs, los polimorfismos identificados y validados se genotiparon en las dos poblaciones de estudio, pacientes con lesiones gástricas y casos y controles de cáncer gástrico. Posteriormente, se realizaron análisis de asociación genética con las distintas variables respuesta, considerándose tanto los genotipos de cada una de las variantes seleccionadas como los haplotipos que dichas variantes forman. Las tablas y figuras indicadas con R se localizan en el apartado de resultados y las tablas y figuras suplementarias indicadas como A se localizan en el apéndice II (tablas) y III (figuras).

2.1. Estudio de seguimiento de pacientes con lesiones gástricas

Para determinar la implicación de genes de protección de la mucosa gástrica y genes relacionados con la respuesta a la infección por *H.pylori* en la evolución de las lesiones gástricas preneoplásicas, se realizó un estudio de asociación entre polimorfismos seleccionados en estos genes (TABLA R-6) y la progresión o regresión de las lesiones en un grupo de 453 pacientes con lesiones gástricas procedentes de un estudio observacional longitudinal (APARTADO 1.1 de materiales y métodos).

2.1.1. Descriptiva de la población y asociación de covariables con la evolución de lesiones

En primer lugar, se realizó una estadística descriptiva de las principales características de la población de pacientes con lesiones precursoras de cáncer gástrico (N=453) y se determinó que variables estaban asociadas con evolución de las lesiones gástricas mediante regresión logística. En la TABLA R-7 se muestra el diagnóstico histopatológico en relación con la edad, el sexo y otros posibles factores de riesgo al inicio del seguimiento.

Tabla R-7. Principales características y diagnóstico histopatológico de pacientes con lesiones gástricas en el inicio del reclutamiento.

Características	Normal (N=8)	NAG (N=51)	MAG (N=215)	Metaplasia Intestinal		Total (N=453)
				Compl. (N=97)	Incompl. (N=82)	
Hombre (%)	62.5	45.1	41.4	49.5	52.4	45.9
Edad media (SD)	36.3 (11.9)	45.0 (12.9)	48.0 (12.1)	51.2 (10.8)	56.1 (9.7)	49.6 (12.1)
<i>H.pylori</i> positivo (%)	0	84.0	81.1	78.1	67.1	77.1
Fumador ^a (%)	50.0	53.1	36.8	45.6	41.3	41.5
Consumo AINES (%)	50.0	40.8	54.5	41.9	49.3	49.3
Antecedentes CG (%)	12.5	16.3	12.9	14.1	28.6	16.2

a, El grupo de fumadores incluye a los ex fumadores. NAG, gastritis no atrófica; MAG, gastritis atrófica multifocal ; CG, Cáncer gástrico. Adaptado de González et al, 2010(45).

Los pacientes de mayor edad presentaron lesiones más avanzadas y hasta el 84% de casos diagnosticados con gastritis no atrófica fueron positivos para la infección por *H.pylori*, porcentaje que disminuyó ligeramente en lesiones más severas. La prevalencia de los fumadores fue del 41.5% y no se observaron diferencias muy claras entre grupos histopatológicos. El consumo de antiinflamatorios no esteroideos (AINES) varió ligeramente en función del tipo de lesión diagnosticada, siendo los pacientes diagnosticados con gastritis atrófica multifocal los que presentaron un mayor consumo (54.5%). La prevalencia de historia familiar con antecedentes de cáncer gástrico fue particularmente alta (28.6%) en los pacientes con metaplasia intestinal incompleta.

A partir del diagnóstico, según el índice de Correa, de las lesiones gástricas en el reclutamiento y al final de seguimiento, se determinó la evolución de éstas tal y como se describe en el APARTADO 1.1 de Materiales y Métodos. Se identificaron 191 (42.2%) pacientes cuyas lesiones se mantuvieron estables, 111 (29.3%) cuyas lesiones experimentaron una regresión y 109 (28.5%) que progresaron. En la TABLA R-8 se resumen las principales características de los pacientes con lesiones gástricas, tanto a nivel global como subdividas por categoría de evolución.

Tabla R-8. Principales características y evolución de los pacientes con lesiones gástricas (N=453).

Características	Total		Estables		Regresión		Progresión		Valor p ^a
	n	%	n	%	n	%	n	%	
Sexo									
Hombre	208	45.92	86	45.03	51	38.35	71	55.04	0.024
Mujer	245	54.08	105	54.97	82	61.65	129	44.96	
Edad									
<40	120	26.49	48	25.13	40	30.08	32	24.81	0.576
40-50	94	20.75	39	20.42	30	22.56	25	19.38	
50-60	139	30.68	58	30.37	42	31.58	39	30.23	
>60	100	22.08	46	24.08	21	15.79	33	25.58	
Consumo AINES									
No	220	50.69	84	45.65	64	48.48	72	61.02	0.028
Si	214	49.31	100	54.35	68	51.52	46	38.98	
Historia familiar									
No	353	83.85	156	86.19	104	82.54	93	81.58	0.516
Si	68	16.15	25	13.81	22	17.46	21	18.42	
Consumo tabaco									
Nunca	255	58.49	111	60.33	78	60.00	66	54.10	0.798
Habitual	109	25.00	43	23.37	33	25.38	33	27.05	
Ex-fumador	72	16.51	30	16.30	19	14.62	23	18.85	
Infección <i>H.pylori</i>									
No	102	22.87	44	23.16	26	19.55	32	26.02	0.465
Si	344	77.13	146	76.84	107	80.45	91	73.98	
Tipo de lesión^b									
Normal	8	1.77	3	1.57	0	0.00	5	3.88	3.56x10⁻⁵
NAG	51	11.26	22	11.52	3	2.26	26	20.16	
MAG	215	47.46	91	47.64	74	55.64	50	38.76	
IM Completa	97	21.41	38	19.90	26	19.55	33	25.58	
IM Incompleta	82	18.10	37	19.37	30	22.56	15	11.63	

a, Valor p de las diferencias entre los grupos de evolución para cada variable basado en el test χ^2 ; b, Tipo de lesión diagnosticada en el reclutamiento. NAG, gastritis no atrófica; MAG, gastritis atrófica multifocal; IM, Metaplasia intestinal.

En esta tabla también se muestran los valores p del test χ^2 de Pearson utilizado para determinar las diferencias existentes entre los grupos de evolución para cada una de las variables analizadas. Además, se evaluó la asociación de estos factores con la variable respuesta mediante regresión logística univariante, tal y como se muestra en la TABLA R-9.

La población total de pacientes se distribuyó de manera equitativa entre hombres (45.9%) y mujeres (54.1%). Esta distribución varió ligeramente en el grupo de progresión y significativamente en el grupo de regresión. En concordancia con estos resultados, en el análisis mediante regresión logística, las mujeres presentaron tanto un menor riesgo significativo de progresión como una mayor probabilidad significativa de regresión. La edad de los participantes en el reclutamiento estuvo comprendida entre los 22 y 71 años, con una media global de 49.6 ± 12.1 años. Los diferentes grupos de evolución presentaron una media similar: estables 50 ± 12.1 años, regresión 47.8 ± 11.7 años y progresión 50.9 ± 12.4 años. En general no se observaron diferencias significativas de edad entre los pacientes (TABLA R-8) pero los mayores de 60 años presentaron una menor probabilidad significativa de regresión (TABLA R-9). Globalmente la mitad de los pacientes incluidos en el estudio eran consumidores de AINES aunque el porcentaje de pacientes no consumidores de este tipo de fármaco aumentó de manera significativa en el grupo de progresión (TABLA R-8). De hecho, en el análisis mediante regresión logística, el consumo de AINES se asoció significativamente con un menor riesgo de progresión. La TABLA R-8 muestra también los diferentes grupos de evolución en función de la lesión gástrica diagnosticada al inicio de reclutamiento. El tipo de lesión inicial puede ser determinante en la posterior evolución de la misma observándose diferencias muy significativas entre tipo de lesión al reclutamiento y los diferentes grupos de evolución y siendo también la asociación significativa en el análisis de regresión logística (TABLA R-9).

Un pequeño porcentaje de pacientes (16.1%) indicó la existencia de antecedentes familiares de cáncer gástrico y casi la mitad de la población se declaró fumadora o exfumadora. La mayoría de los pacientes incluidos en el estudio resultaron positivos en el análisis de infección por *H.pylori* (77.1%), tal y como era previsible al tratarse de pacientes con lesiones gástricas relacionadas con este tipo de infección. No obstante, no se observaron diferencias ni asociación significativa entre los diferentes grupos de evolución e historia familiar de cáncer gástrico, consumo de tabaco e infección por *H.pylori* o alguna de sus cepas (TABLAS R-8 y R-9).

A partir de estos resultados, el sexo, la edad, el consumo de AINES y el tipo de lesión al reclutamiento se incluyeron en el modelo final como variables de ajuste.

Tal y como se comentó en el APARTADO 7.4.1 de Materiales y Métodos, para el genotipado de 67 polimorfismos en la plataforma de Sequenom, sólo se dispuso de 387 muestras de pacientes con lesiones gástricas. A pesar del menor número de individuos disponibles para el estudio en estos casos, tanto la distribución como las diferencias entre grupos de evolución de las principales características analizadas mantuvieron la misma tendencia (ver TABLA A-2 del apéndice II).

Tabla R-9. Asociación de posibles factores de riesgo con la evolución de las lesiones gástricas

Variables	Categorías	Regresión					Progresión				
		Contr	Casos	OR	(CI 95%) ^a	Valor p	Contr	Casos	OR	(CI 95%) ^a	Valor p
Sexo	Hombre	157	51	1.00			137	71	1.00		
	Mujer	163	82	1.55	(1.02-2.34)	0.038	187	58	0.60	(0.40-0.90)	0.014
Edad	<40	80	40	1.00			88	32	1.00		
	40-50	64	30	0.94	(0.52-1.67)	0.826	69	25	1.00	(0.54-1.83)	0.991
	50-60	97	42	0.87	(0.51-1.46)	0.591	100	39	1.07	(0.62-1.86)	0.802
	>60	79	21	0.53	(0.29-0.98)	0.043	67	33	1.35	(0.76-2.42)	0.306
Consumo AiNES	No	156	64	1.00			148	72	1.00		
	Si	146	68	1.13	(0.75-1.71)	0.543	168	46	0.56	(0.37-0.87)	0.009
Historia familiar	No	249	104	1.00			260	93	1.00		
	Si	46	22	1.14	(0.66-2.00)	0.634	47	21	1.25	(0.71-2.20)	0.441
Consumo tabaco	Nunca	177	78	1.00			189	66	1.00		
	Habitual	76	33	0.81	(0.45-1.46)	0.491	76	33	1.34	(0.76-2.37)	0.389
	Ex fumador	53	19	0.98	(0.60-1.60)	0.953	49	23	1.24	(0.76-1.04)	0.308
Infección <i>H.pylori</i>	No	76	26	1.00			70	32	1.00		
	Si	237	107	1.32	(0.80-2.18)	0.277	253	91	0.79	(0.49-1.27)	0.330
Tipo de lesión^b	Normal+NAG	56	3	1.00			28	31	1.00		
	MAG	141	74	9.80	(2.96-32.37)	1.8x10⁻⁴	165	50	0.27	(0-15-0.50)	2.3x10⁻⁵
	IM Completa	71	26	6.84	(1.97-23.75)	0.002	64	33	0.47	(0.24-0.90)	0.024
	IM Incompleta	52	30	10.77	(3.10-37.42)	1.8x10⁻⁴	67	15	0.20	(0.09-0.43)	3.6x10⁻⁵

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento; b, Tipo de lesión diagnosticada en el reclutamiento. Debido al pequeño tamaño muestral de la categoría normal, ésta se ha reagrupado con la categoría más similar, NAG. NAG, gastritis no atrófica; MAG, gastritis atrófica multifocal; IM, Metaplasia intestinal.

2.1.2. Genotipado de los polimorfismos

Las frecuencias alélicas obtenidas en el genotipado para cada uno de los polimorfismos en la totalidad de la población así como por grupo de evolución se detallan en las TABLA A-3 del apéndice II.

Dos de los polimorfismos inicialmente seleccionados (n=109), se eliminaron de este estudio porque no permitían un diseño adecuado para su genotipado en la plataforma de Sequenom: los tagSNPs rs12483635 de *TFF2* y rs11203197 de *TFF3*. Además, tal y como se indica en la TABLA A-3, se excluyeron 2 SNPs por ser monomórficos (1 en *TFF1* y 1 en *TFF2*) y 7 (1 en cada uno de los siguientes genes: *MUC1*, *MUC2*, *MUC6*, *TFF1*, *TFF3*, *NOD2*, *CDH1*) por problemas de genotipado (no amplificación o, en la mayoría de casos, deficiente separación de los clústeres de cada genotipo) en la plataforma Sequenom. Un polimorfismo, el rs13181 del gen *ERCC2*, se excluyó por no superar el control de calidad establecido (mostraba un 5.8% de discordancia entre muestras duplicadas). Los 97 polimorfismos restantes (22 en *MUCs*, 27 en *TFFs* y 27 en *ILs* y 21 en genes relacionados con *H.pylori*) cumplieron el equilibrio Hardy-Weinberg ($p > 10^{-4}$, TABLA A-3) en la población analizada, condición indispensable para incluirlos en el análisis de asociación y presentaron una concordancia del 100% en el control de calidad, a excepción de la inserción de *TFF2* que fue del 99.6%.

La media del porcentaje de muestras genotipadas mediante las distintas metodologías utilizadas en este estudio fue del 98.91%. Las frecuencias alélicas fueron muy variables, desde frecuencias muy bajas a partir de 0.01, generalmente polimorfismos potencialmente funcionales en *TFFs*, a frecuencias mucho más elevadas, de hasta 0.48, que corresponderían en su mayoría a polimorfismos seleccionados como tagSNPs. En el caso de mucinas todos los polimorfismos son marcadores y tuvieron frecuencias mayores de 0.05 ya que era un criterio indispensable para su selección, y, en general, son polimorfismos muy frecuentes.

Uno de los polimorfismos, el rs225334 de *TFF2*, se genotipo mediante tres técnicas diferentes: HRM, conjuntamente con la inserción de 3'UTR c*68_69insCTT, sondas de hibridación FRET y MassArray de Sequenom. Se realizó una comparativa de los resultados de genotipado entre todos los individuos para los que se pudo determinar el genotipo mediante sondas FRET y las demás técnicas para comprobar el grado de concordancia (FIGURA R-24).

Se utilizó el método Cohen-Kappa que se basa en comparar la concordancia observada entre dos métodos cuando son aplicados al mismo grupo de individuos, respecto a la que podría ocurrir por azar. Se obtuvo un índice kappa de 0.97 entre sondas FRET y MassArray y de 0.99 entre sondas FRET y HRM. Estos valores indican una concordancia casi perfecta, correspondiente a los valores entre 0.81 y 1.00, según la escala de interpretación de Landis y Koch (378).

		FRET			
		TT	TC	CC	
Sequenom	TT	138	0	0	138
	TC	1	148	1	150
	CC	0	4	51	55
		139	152	52	343
		Kappa=0.9877			

		FRET			
		TT	TC	CC	
HRM	TT	152	0	0	152
	TC	1	195	1	197
	CC	0	1	55	56
		153	196	56	405
		Kappa=0.9717			

Figura R-24. Comparativas de los resultados de genotipado del SNP rs225334 de *TFF2*, mediante distintas tecnologías. Se indican los valores del índice kappa de concordancia. A. Comparativa entre el genotipado con sondas FRET y MassArray de Sequenom. B. Comparativa entre el genotipado con sondas FRET y HRM.

2.1.3. Análisis de LD y determinación de haplotipos

Con los datos obtenidos en el genotipado de polimorfismos de *TFFs* y *MUCs* en pacientes con lesiones precursoras de cáncer gástrico se calcularon los coeficientes de LD, D' y r^2 , entre pares de SNPs de cada uno de los genes y se infirieron los haplotipos, seleccionando únicamente aquellos con frecuencias superiores a 0.01. Los haplotipos inferidos se muestran en las TABLAS R-16 (*TFFs*) y R-18 (*MUCs*) del APARTADO 2.1.5 de resultados, conjuntamente con los resultados de asociación de estos haplotipos con la evolución de lesiones gástricas.

A pesar de que los genes *TFF* son pequeños, alrededor de las 4 kb, y se localizan en un único clúster en el cromosoma 21, no se observa un fuerte desequilibrio de ligamiento ni en el locus de cada gen (FIGURA R-25) ni globalmente en la región cromosómica (FIGURA A-1A y A-1B del apéndice III), confirmando los datos que obtuvimos del HapMap en la selección de tagSNPs (APARTADO 1.3.1 de resultados.). Por esa razón, el análisis detallado de LD no se realizó para todo el clúster sino individualmente para cada gen *TFF* y regiones adyacentes.

Aunque existe cierto LD entre algunos polimorfismos en zonas determinadas de la región que engloba a *TFF1*, este no se mantiene a lo largo del locus por lo que no se observan bloques de LD compactos (FIGURA R-25). El tagSNP inicialmente seleccionado rs4920094, en el intrón 1 de *TFF1*, se eliminó debido a problemas en el genotipado. Este SNP marcaba, junto dos tagSNPs próximos situados en el mismo intrón, un bloque de ~1 kb localizado en la región 5' del gen. Aunque se pierde un 40% de cobertura de la variabilidad en este bloque, debido a su pequeño tamaño la pérdida de información en la región es mínima, sobre todo si se tiene cuenta que se seleccionaron 3 bloques más de LD en *TFF1*. Los 11 polimorfismos analizados en *TFF1* forman hasta un total de 17 haplotipos con frecuencias comprendidas entre 0.011 y 0.277 y sólo tres de ellos superan una frecuencia del 10% (TABLA R-16), reflejando el poco LD existente en esta zona. Una situación similar se da en *TFF2*, donde no existe un único bloque claro, en valores de D' , que englobe todo el gen (FIGURA R-25). Este gen se encuentra adyacente a *TFF1* separado únicamente por 10 kb y, aunque la región analizada para *TFF2* (5.3 kb) abarca un tamaño poco mayor que el

gen (4.7 kb), los tagSNPs analizados, juntamente con los de *TFF1*, cubren toda la región intergénica según el análisis realizado a partir de los datos del HapMap (APARTADO 1.3.1 de resultados). A pesar de que se tuvo que excluir el rs12483635, un tagSNP poco frecuente (MAF=0.08 según datos del HapMap) seleccionado en el bloque 3 del clúster de *TFFs* (TABLA R-2) y que no marcaba a ningún otro SNP, la reducción de la variabilidad fue mínima, tan sólo de un 3%. En *TFF2* se infirieron 10 haplotipos, formados por los 9 SNPs analizados, con frecuencias comprendidas entre 0.011 y 0.324 (TABLA R-16).

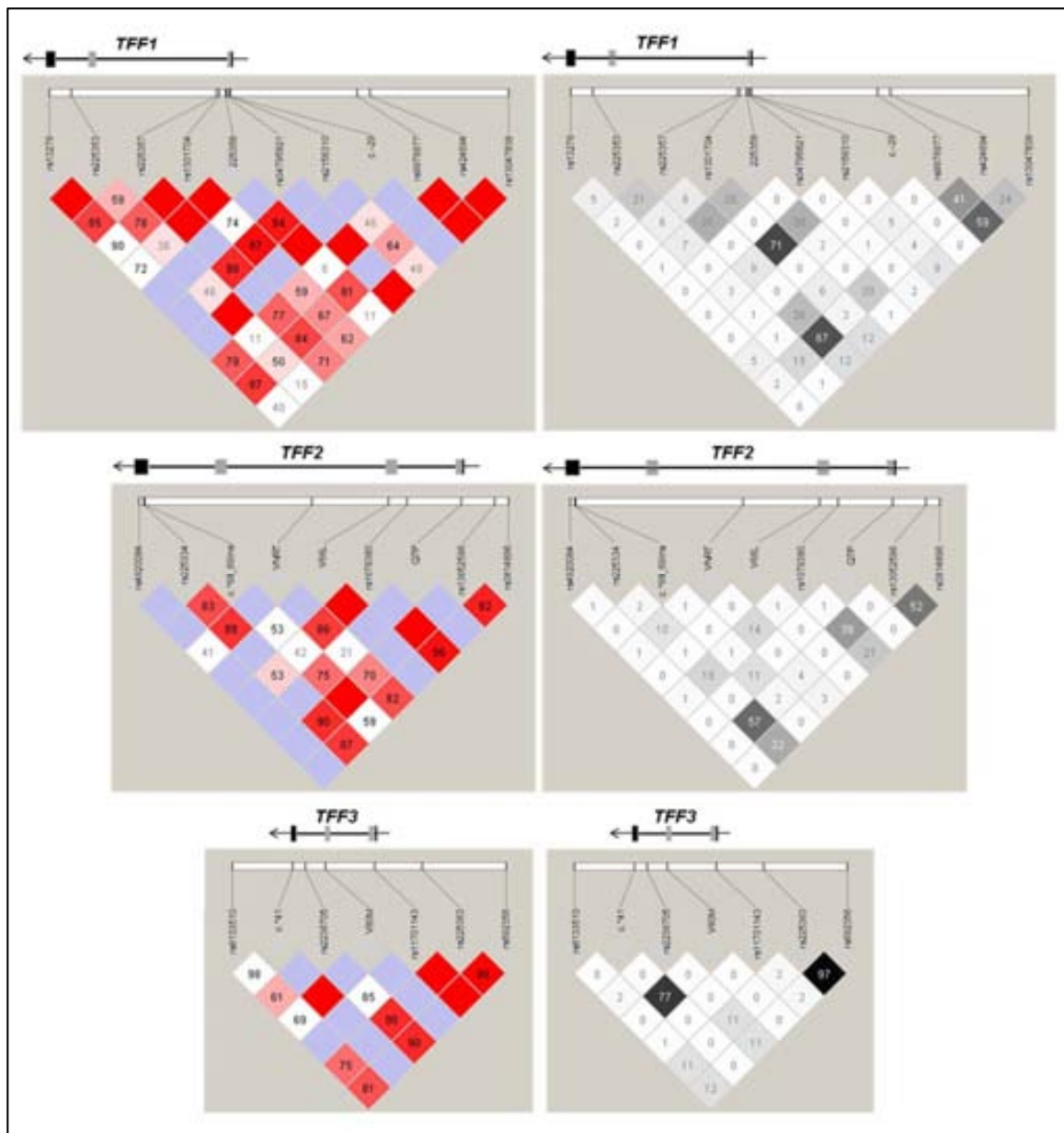


Figura R-25. Análisis del desequilibrio de ligamiento entre polimorfismos de las regiones en 21q22.3 donde se localizan *TFF1*, *TFF2* y *TFF3*, a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico. En la parte superior de cada figurase indica la estructura del gen correspondiente. Izquierda, representación gráfica del coeficiente D' . Los valores D' de 1 no se muestran. Esquema de color: cuadrados en blanco, $D' < 1$ y $LOD < 2$; en azul, $D' = 1$ y $LOD < 2$; en tonos rosados, $D' < 1$ y $LOD \geq 2$ y en rojo fuerte, $D' = 1$ y $LOD \geq 2$. Derecha, representación gráfica del coeficiente r^2 . Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

TFF3 es el gen del clúster donde se observa un mayor LD, con valores de D' mayores de 0.6 en todos los casos, y en la gran mayoría superiores al 0.8 (FIGURA R-25). Este hecho se refleja también en la formación de haplotipos; 7 SNPs configuran 6 haplotipos diferentes con frecuencias entre 0.015 y 0.376 (TABLA R-16). La ausencia de datos de genotipado para el rs9981660, uno de los cinco tagSNPs de un bloque de 7 kb que abarca todo el gen *TFF3* (TABLA R-2), resultó en una reducción de la cobertura de la variabilidad genética en un 12% en ese bloque y sólo del 6% si se considera toda la región donde se localiza el gen. Aunque también se excluyó del estudio el tagSNP rs11203197, seleccionado en el bloque 2 (TABLA R-2) de la región genómica a 5' de *TFF3*, no afectó a la cobertura de la variabilidad ya que está marcado por un SNP del bloque adyacente (rs11701143, el 39 del bloque 1 en TABLA R-2). El LD observado a partir de los valores de r^2 es también muy bajo entre la mayoría de polimorfismos de *TFFs*. En muchos casos este coeficiente adquiere el valor de 0 debido a la diferencias entre las frecuencias alélicas de los polimorfismos seleccionados, tagSNPs con frecuencias relativamente elevadas y SNPs funcionales con frecuencias generalmente muy bajas. Además, debido al criterio escogido para la selección de SNPs marcadores ($r^2 > 0.8$) era de esperar no encontrar valores superiores al 0.8 entre este tipo de polimorfismos.

MUC1 es el gen *MUC* de menor tamaño (~4.5 kb) y se encuentra en una zona de relativamente elevado LD por lo que se analizó un reducido número de tagSNPs (FIGURA R-26). Los dos tagSNPs a 3' del gen, que se localizan en el gen adyacente *TRIM46*, muestran un fuerte LD que no se mantiene con el tercer SNP analizado en la región genómica a 5' del gen *MUC1*. La imposibilidad de genotipar el único tagSNP localizado en el gen y recientemente asociado a cáncer gástrico, el rs4072037, disminuyó la cobertura de la variabilidad genética de la región en un 40%. Con los tres tag SNPs genotipados se infirieron seis haplotipos con frecuencias comprendidas entre 0.015 y 0.466, siendo el haplotipo CAC, que porta los alelos más comunes de cada uno de los SNPs que lo forman, el haplotipo más frecuente con diferencia (TABLA R-18). El gen *MUC2* (~35 kb) es mucho más complejo a causa de su gran tamaño y bajo grado de LD. En general, la estructura de LD obtenida a partir de los genotipos de pacientes de lesiones gástricas (FIGURA R-26) es muy similar a la observada anteriormente a partir de los datos de HapMap (FIGURA R-18). Uno de los tagSNPs, el rs10902089, se descartó por problemas de genotipado en la plataforma de Sequenom, por lo que el número final de polimorfismos analizados en este se gen se redujo a 14. No obstante, el porcentaje de la variabilidad genética cubierta por los tagSNPs analizados permaneció prácticamente inalterada ya que la reducción fue sólo de un 7% en toda la región 3' del gen. Debido a la complejidad de *MUC2*, y al poco desequilibrio de ligamiento existente (FIGURA R-26), el análisis de haplotipos se realizó de manera independiente para las regiones 5' y 3' del gen, separadas por la presencia del exón central que codifica para el dominio VNTR. Desde la región genómica a 5' del gen hasta el exón 30 se analizaron seis polimorfismos conformando diez haplotipos diferentes con frecuencias entre 0.345 y 0.014. En la región 3' del gen, a partir del exón 30, se analizaron otros ocho polimorfismos configurando también diez haplotipos con frecuencias comprendidas entre 0.322 y 0.018. Los haplotipos inferidos se muestran conjuntamente con los resultados de asociación en la TABLA R-18 del APARTADO 2.1.5.

De entre todos los genes *MUC* analizados, *MUC6* es el que presenta un mayor LD, tal y como ya se observó con el análisis realizado con los datos de genotipado de HapMap (FIGURA R-17). La totalidad del gen *MUC6* (~24 kb) está marcado por tres polimorfismos (rs4076950 rs7481521, rs11246384) entre los que existe un elevado LD, con valores D' entre 0.9 y 1 (FIGURA R-26).

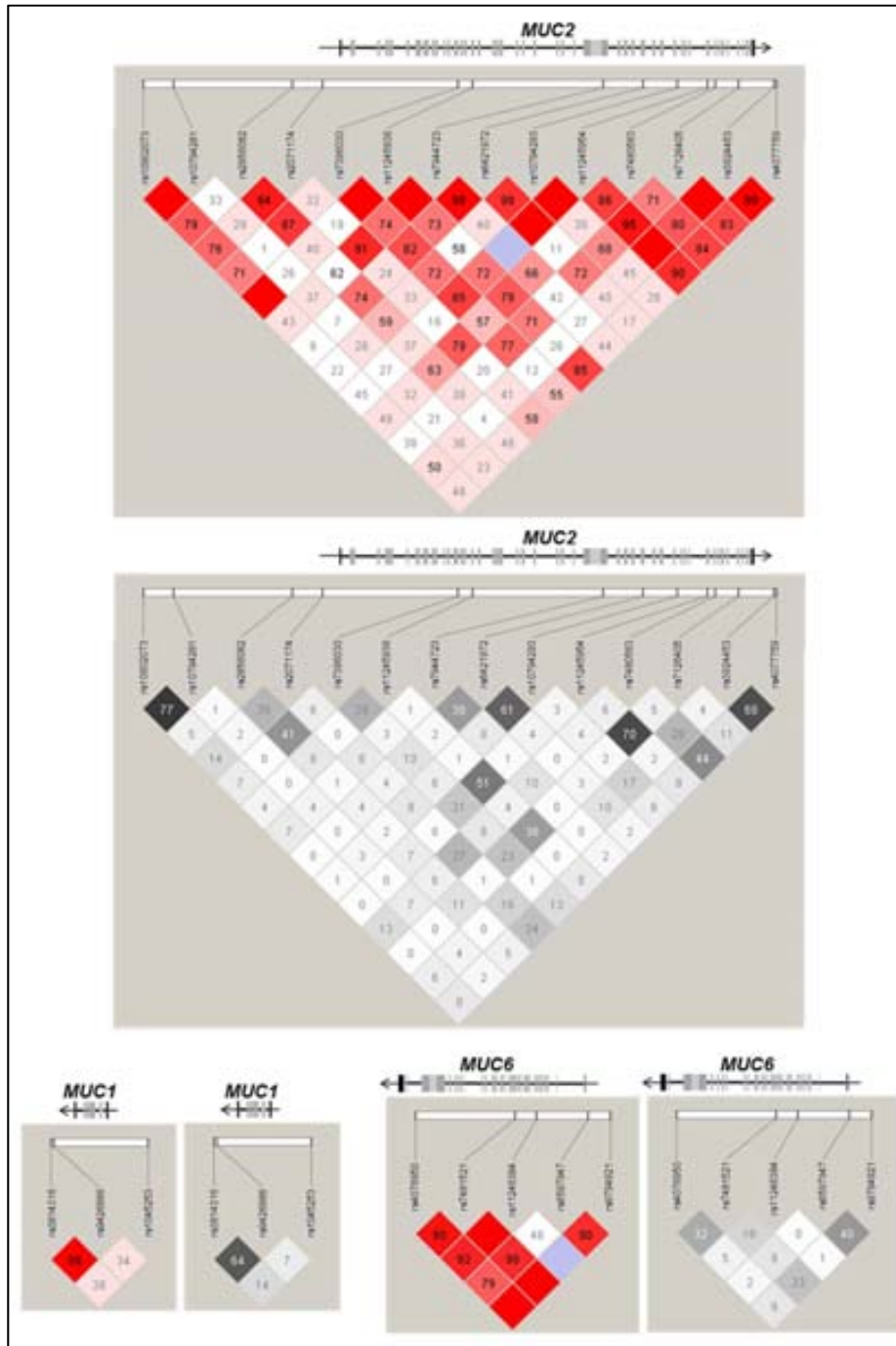


Figura R-26. Análisis de desequilibrio de ligamiento entre polimorfismos de las regiones donde se localizan *MUC1*, en 1q21, y *MUC2* y *MUC6*, en 11p15, a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico. En la parte superior de cada figurase indica la estructura del gen correspondiente. En color, representación gráfica del coeficiente D' . Los valores D' de 1 no se muestran. Esquema de color: cuadrados en blanco, $D' < 1$ y $\text{LOD} < 2$; en azul, $D' = 1$ y $\text{LOD} < 2$; en tonos rosados, $D' < 1$ y $\text{LOD} \geq 2$ y en rojo fuerte, $D' = 1$ y $\text{LOD} \geq 2$. En blanco y negro, representación gráfica del coeficiente r^2 . Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

Uno de los tagSNPs inicialmente seleccionados, rs11604757, se descartó por problemas en la plataforma de genotipación resultando en una reducción de la cobertura de la variabilidad genética del 10%. Los otros dos polimorfismos analizados (rs6597947, rs9794921) se localizan en la región genómica a 5' del gen y muestran también un fuerte LD ($D'=0.9$). Se infirieron un total de seis haplotipos, a partir de la combinación de los cinco polimorfismos genotipados, con frecuencias entre 0.069 y 0.373 (TABLA R-18) Referente a los valores observados para el coeficiente r^2 , no se encontraron valores superiores al 0.8, como era previsible. En muchos casos, estos valores no superaron el 0.1 reflejando la variabilidad entre las frecuencias alélicas de los SNPs y/o el poco LD existente en algunas regiones, especialmente donde se localiza *MUC2*.

En general, la estructura y los valores de LD (D' y r^2) para cada uno de los genes observados en el grupo de pacientes con lesiones gástricas está en concordancia con los obtenidos a partir de los datos de HapMap tal y como se ejemplifica en el caso de *MUC2* (FIGURA A-2 del apéndice II).

2.1.4. Asociación polimorfismos con lesiones precursoras

El análisis de asociación de polimorfismos (n=97) de genes involucrados en la protección de la mucosa gástrica y la respuesta a la infección por *H.pylori* con la evolución de lesiones precursoras se realizó mediante regresión logística incondicional. En el modelo final se incluyeron como variables de ajuste la edad, sexo, consumo de AINES y tipo de lesión en el reclutamiento (APARTADO 2.1.1 de resultados). En la TABLA A-4 del apéndice II se indica la significación (determinada por el valor p) de todos los resultados obtenidos en el análisis de asociación con la evolución de lesiones gástricas.

Además, el análisis de asociación se estratificó según la infección por *H.pylori*, el sexo y el consumo de AINES para determinar relación de los genes analizados con estas tres variables.

2.1.4.1 *TFFs*

Se analizó la asociación de 27 polimorfismos en el clúster de genes *TFF* (11 en *TFF1*, 9 en *TFF2* y 7 en *TFF3*) con la progresión y regresión de las lesiones precursoras de cáncer gástrico. Seis polimorfismos de *TFFs* (ver TABLA A-4) se asociaron significativamente con la evolución de las lesiones: dos con la progresión (rs225358 en *TFF1* y rs1079380 en *TFF2*) y cuatro con la regresión (rs9976977, rs424694 y rs13047838 en *TFF1* y c.*41G>A en *TFF3*).

El rs225358, en el intrón 1 de *TFF1*, se asoció con un mayor riesgo de progresión bajo los modelos de herencia codominante y sobredominante (TABLA R-10). El rs225358 es un tagSNP que marca parte del extremo 5' de *TFF1*, incluida la región 5'UTR, y está en elevado LD ($D'=0.9-1$) con otros SNPs próximos analizados en esta región (FIGURA R-25).

Tabla R-10. Asociación del SNP rs225358 de *TFF1* con la progresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por *H.pylori*.

Modelo	Progresión			<i>H.pylori</i> +			<i>H.pylori</i> -			P-Inter.
	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	
Codominante										
TT	134/41	1.00		109/35	1.00		24/5	1.00		
TC	106/54	1.75 (1.05 - 2.93)		78/35	1.43 (0.80 - 2.57)		28/18	6.18 (1.50 - 25.53)		
CC	32/9	0.72 (0.30 - 1.72)	.0320	23/8	1.04 (0.40 - 2.66)	.4645	9/1	0.22 (0.02 - 2.70)	.0012	.0591
Dominante										
TT	134/41	1.00		109/35	1.00		24/5	1.00		
TC-CC	138/63	1.48 (0.91 - 2.40)	.1146	101/43	1.34 (0.77 - 2.33)	.2999	37/19	2.92 (0.86 - 9.97)	.0739	.3447
Recesivo										
TT-TC	240/95	1.00		187/70	1.00		52/23	1.00		
CC	32/9	0.55 (0.24 - 1.25)	.1384	23/8	0.88 (0.36 - 2.16)	.7766	9/1	0.09 (0.01 - 0.94)	.0145	.0488
Sobredominante										
TT-CC	166/50	1.00		132/43	1.00		33/6	1.00		
TC	106/54	1.86 (1.14 - 3.04)	.0120	78/35	1.42 (0.81 - 2.49)	.2164	28/18	8.60 (2.19 - 33.73)	.0006	.0330
Aditivo										
TT,TC,CC	272/104	1.09 (0.77 - 1.55)	.6287	210/78	1.14 (0.76 - 1.71)	.5255	61/24	1.04 (0.47 - 2.27)	.9276	.7176

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento.

Cuando el efecto del rs225358 se analizó en función de la infección por *H.pylori*, la asociación con la progresión fue significativa sólo en los individuos no infectados y la interacción fue estadísticamente significativa (TABLA R-10). No obstante, en ambos grupos se observa la misma tendencia aunque con ORs en los no infectados muy superiores. Esto puede ser debido al amplio rango del intervalo de confianza causado por el pequeño tamaño muestral de este grupo. En el análisis estratificado por sexo, aunque sólo se mantuvo la significación en hombres en los modelos codominante (CT, OR=2.64, 95% IC=1.21-5.76; TT, OR=1.36, 95% IC=0.43-4.28, $p=0.0438$), dominante (CT-CC, OR=2.25, 95% IC=1.08-4.69, $p=0.0265$) y sobredominante (CT, OR=2.45, 95% IC=1.18-5.10, $p=0.0145$), no se observó una interacción positiva (p de la interacción=0.2120, 0.1467 y 0.3631, respectivamente). De manera similar, en la estratificación por consumo de AINES, únicamente los consumidores de este tipo de fármacos se asociaron con la progresión en el modelo sobredominante (CT, OR=2.47, 95% IC=1.16-5.23, $p=0.0164$), sin ser la interacción significativa ($p=0.3443$).

El alelo G del rs1079380, un tagSNP en el intrón 1 de *TFF2*, se asoció como factor protector de la progresión en los modelos codominante (AG, OR=1.05, 95% IC=0.61-1.81; GG, OR=0.48, 95% IC=0.24-0.98, $p=0.0482$) y recesivo (GG, OR=0.47, 95% IC=0.25-0.88, $p=0.0140$). En este gen no se observó ninguna otra asociación con la evolución de lesiones gástricas y el análisis estratificado por sexo, consumo de AINES e infección por *H.pylori* no reveló ninguna interacción significativa.

Otros cuatro polimorfismos se asociaron significativamente con la regresión de lesiones precursoras de cáncer gástrico. Tres de estos polimorfismos (rs9976977, rs424694, rs13047838) son tagSNPs localizados en la región genómica a 5' del gen *TFF1* (FIGURA R-25) y se asociaron con una mayor probabilidad de regresión en los modelos dominante (rs9976977, rs13047838) y/o sobredominante (rs424694, rs13047838) tal y como muestra la TABLA R-11. Los tres SNPs se asociaron con una mayor regresión únicamente en los pacientes consumidores de AINES en el análisis estratificado aunque la interacción no fue significativa (TABLA R-11). Cuando el efecto de estos polimorfismos se analizó en función del sexo y la infección por *H.pylori* no se observaron diferencias relevantes en la asociación con la regresión de los diferentes subgrupos ni tampoco se hallaron interacciones estadísticamente significativas.

El SNP localizado en la región 3'UTR de *TFF3* c*41G>A también se asoció con una mayor probabilidad de regresión en el modelo dominante y aditivo ($p=0.0377$ en ambos casos). En este caso se trata de un SNP muy poco frecuente, MAF=0.01, y no se pudieron calcular los OR, debido a que, en el grupo de casos sólo había presencia de homocigotos GG, ni realizar análisis posteriores como la estratificación por diferentes factores de riesgo.

Tabla R-11. Asociación de los tagSNPs de la región génomica a 5' de *TFF1* con la regresión de lesiones gástricas precursoras de cáncer y estratificación por consumo de AINES.

SNP	Modelo	Genotipo	Regresión			No Consumo AINES			Consumo AINES			P-Inter.		
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p			
rs9976977	Cod	GG	112/35	1.00		56/20	1.00		56/15	1.00				
		GA	113/55	1.74	(1.03-2.95)	60/24	1.39	(0.66-2.91)	53/31	2.06	(0.97-4.37)			
		AA	42/20	1.42	(0.72-2.80)	.1101	25/11	1.15	(0.46-2.83)	.6780	17/9	1.91	(0.67-5.44)	.1444
	Dom	GG	112/35	1.00		56/20	1.00		56/15	1.00				
		GA-AA	155/75	1.64	(1.01-2.69)	.0448	85/35	1.30	(0.66-2.56)	.4406	70/40	2.02	(0.99-4.16)	.0498
	Rec	GG-GA	225/90	1.00		116/44	1.00		109/46	1.00				
		AA	42/20	1.05	(0.57-1.94)	.8709	25/11	0.97	(0.43-2.20)	.9429	17/9	1.24	(0.49-3.16)	.6494
	Sob	GG-AA	154/55	1.00		81/31	1.00		73/24	1.00				
		GA	113/55	1.56	(0.97-2.51)	.0649	60/24	1.33	(0.68-2.59)	.4061	53/31	1.70	(0.87-3.31)	.1194
	Adi	GG,GA,AA	267/110	1.27	(0.92-1.75)	.1524	141/55	1.11	(0.72-1.71)	.6640	126/55	1.50	(0.91-2.45)	.1062
rs424694	Cod	CC	100/37	1.00		50/18	1.00		50/19	1.00				
		CT	101/55	1.46	(0.87-2.46)		58/26	1.21	(0.58-2.53)		43/29	1.83	(0.87-3.86)	
		TT	50/13	0.65	(0.31-1.36)	.0589	25/7	0.65	(0.23-1.84)	.4509	25/6	0.67	(0.23-1.99)	.0946
	Dom	CC	100/37	1.00		50/18	1.00		50/19	1.00				
		CT-TT	151/68	1.18	(0.72-1.94)	.5000	83/33	1.03	(0.51-2.08)	.9260	68/35	1.42	(0.71-2.87)	.3193
	Rec	CC-CT	201/92	1.00		108/44	1.00		93/48	1.00				
		TT	50/13	0.52	(0.26-1.04)	.0566	25/7	0.58	(0.22-1.50)	.2472	25/6	0.49	(0.18-1.33)	.1458
	Sob	CC-TT	150/50	1.00		75/25	1.00		75/25	1.00				
		CT	101/55	1.66	(1.03-2.68)	.0380	58/26	1.38	(0.71-2.71)	.3422	43/29	2.04	(1.03-4.08)	.0407
	Adi	CC,CT,TT	251/105	0.91	(0.66-1.28)	.5972	133/51	0.88	(0.54-1.41)	.5843	118/54	0.99	(0.62-1.58)	.9561
rs1304783	Cod	CC	144/46	1.00		71/26	1.00		73/20	1.00				
		CT	89/49	1.83	(1.10-3.03)		52/21	1.17	(0.58-2.38)		37/28	2.74	(1.32-5.68)	
		TT	22/9	1.20	(0.50-2.86)	.0635	11/4	0.91	(0.26-3.27)	.8807	11/5	1.54	(0.46-5.13)	.0230
	Dom	CC	144/46	1.00		71/26	1.00		73/20	1.00				
		CT-TT	111/58	1.69	(1.04-2.73)	.0319	63/25	1.12	(0.57-2.20)	.7378	48/33	2.45	(1.23-4.89)	.0100
	Rec	CC-CT	233/95	1.00		123/47	1.00		110/48	1.00				
		TT	22/9	0.91	(0.39-2.11)	.8325	11/4	0.86	(0.25-2.94)	.8026	11/5	0.97	(0.31-3.04)	.9538
	Sob	CC-TT	166/50	1.00		82/30	1.00		84/25	1.00				
		CT	89/55	1.78	(1.09-2.89)	.0207	52/21	1.19	(0.60-2.36)	.6276	37/28	2.55	(1.27-5.10)	.0078
	Adi	CC,CT,TT	255/104	1.33	(0.93-1.90)	.1245	134/51	1.04	(0.62-1.75)	.8797	121/53	1.63	(0.98-2.71)	.0572

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobresominante; Adi, aditivo.

2.1.4.2 MUCs

En el caso de genes *MUC*, no se encontraron asociaciones significativas entre la evolución de lesiones gástricas precursoras y las variantes genotipadas de *MUC1* y *MUC6* (TABLA A-4).

Sin embargo, tal y como se muestra en la TABLA R-12, tres polimorfismos en *MUC2*, rs10794293, en el intrón 37 y rs3924453 y rs4077759, en la región genómica a 3' del gen, se asociaron con un menor riesgo de progresión de lesiones. El alelo T del rs10794293 en el intrón 37 se asoció con un menor riesgo de progresión en los modelos de herencia aditivo y dominante. Este polimorfismo es tagSNPs de un bloque de haplotipos localizado a 3' de la región central de repeticiones en tándem del gen. El alelo A del rs3924453 y el alelo C del rs4077759 también se asociaron como factores protectores para la progresión de lesiones y este efecto fue significativo en los modelos de herencia recesivo y aditivo, así como en el modelo codominante para el rs4077759. Ambos polimorfismos son tagSNPs de un bloque de haplotipos de 4 kb que abarca el extremo 3' del gen *MUC2*, incluida la región 3'UTR, y región genómica adyacente (FIGURAS R-18 y R-26).

Además, otros cuatro polimorfismos localizados en la región central y 5' del gen *MUC2* (rs10902073, rs10794281, rs2071174 y rs7944723), se asociaron significativamente con la regresión de lesiones, tal y como se muestra en la TABLA R-13. Los SNP rs10902073 y rs10794281, que marcan el principal bloque de haplotipos de *MUC2* de 7 kb en la región genómica a 5' del gen, se asociaron con una menor probabilidad de regresión en los modelos dominante y aditivo. También se observó una asociación significativa con menor regresión para el rs7944723, un SNP codificante sinónimo en el exon central 30, en el modelo recesivo. Por otra parte, el alelo C del rs2071174, localizado en la región genómica a 5' de *MUC2* y que marca un bloque de haplotipos que contiene la región promotora y los primeros exones del gen, se asoció con una mayor probabilidad de regresión de lesiones en los modelos codominante y recesivo.

Debido al elevado número de SNPs asociados a evolución de las lesiones en un mismo gen se realizó un análisis de regresión logística progresiva (*stepwise*), con cada uno de los modelos de herencia, para intentar determinar los efectos independientes de los múltiples marcadores asociados. Mediante este método no se pudo descartar ninguno de los SNPs asociados, teniendo en cuenta todos los modelos, por lo que se deduce que su efecto es independiente y las asociaciones observadas no son debidas al elevado LD existente entre alguno de estos polimorfismos.

Cuando el efecto de los polimorfismos asociados con la evolución de lesiones precursoras se analizó en función de la infección por *H. pylori*, se observó que en todos los casos la asociación fue significativa únicamente en los individuos infectados por la bacteria (TABLAS R-12 y R-13), siendo la interacción significativa para el polimorfismo rs4077759, bajo los modelos recesivo y codominante, así como para el rs10902073 y rs10794281, bajo los modelos aditivo, dominante y codominante.

Tabla R-12. Asociación de los SNPs de *MUC2* rs10794293, rs3924453 y rs4077759 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por *H.pylori*.

SNP	Modelo	Genotipo	Progresión			<i>H.pylori</i> +			<i>H.pylori</i> -			P-Inter.		
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p			
rs10794293	Cod	CC	117/57	1.00		90/45	1.00		26/10	1.00				
		CT	125/40	0.66	(0.40 - 1.10)	102/29	0.52	(0.29 - 0.94)	23/11	1.48	(0.48 - 4.62)			
		TT	33/7	0.43	(0.17 - 1.08)	.0905	21/4	0.32	(0.10 - 1.06)	.0328	12/3	0.71	(0.14 - 3.58)	.6123
	Dom	CC	117/57	1.00		90/45	1.00		26/10	1.00				
		CT-TT	158/47	0.61	(0.38 - 1.00)	.0476	123/33	0.49	(0.28 - 0.87)	.0133	35/14	1.22	(0.42 - 3.53)	.7117
	Rec	CC-CT	242/97	1.00		192/74	1.00		49/21	1.00				
		TT	33/7	0.52	(0.21 - 1.28)	.1382	21/4	0.45	(0.14 - 1.44)	.1538	12/3	0.59	(0.13 - 2.63)	.4731
	Adi	CC,CT,TT	275/104	0.66	(0.45 - 0.96)	.0284	213/78	0.54	(0.34 - 0.87)	.0092	61/24	0.96	(0.47 - 1.98)	.9131
rs3924453	Cod	GG	135/60	1.00		112/47	1.00		23/12	1.00				
		GA	112/40	0.77	(0.47 - 1.27)	82/30	0.83	(0.47 - 1.46)	29/9	0.70	(0.20 - 2.43)			
		AA	28/4	0.30	(0.10 - 0.93)	.0597	19/1	0.12	(0.01 - 0.94)	.0292	9/3	0.36	(0.07 - 1.97)	.4600
	Dom	GG	135/60	1.00		112/47	1.00		23/12	1.00				
		GA-AA	140/44	0.67	(0.42 - 1.00)	.1074	101/31	0.70	(0.40 - 1.22)	.2053	38/12	0.56	(0.19 - 1.70)	.3066
	Rec	GG-GA	247/100	1.00		194/77	1.00		52/21	1.00				
		AA	28/4	0.34	(0.11 - 1.01)	.0320	19/1	0.13	(0.02 - 1.00)	.0099	9/3	0.40	(0.08 - 2.11)	.2665
	Adi	GG,GA,AA	275/104	0.66	(0.44 - 0.97)	.0317	213/78	0.63	(0.39 - 1.01)	.0488	61/24	0.62	(0.29 - 1.34)	.2203
rs4077759	Cod	TT	105/47	1.00		89/36	1.00		16/10	1.00				
		TC	119/48	0.86	(0.52 - 1.44)	85/38	1.05	(0.59 - 1.87)	34/9	0.53	(0.14 - 1.97)			
		CC	46/7	0.35	(0.14 - 0.85)	.0437	35/2	0.14	(0.03 - 0.65)	.0045	10/5	0.56	(0.12 - 2.53)	.5782
	Dom	TT	105/47	1.00		89/36	1.00		16/10	1.00				
		TC-CC	165/55	0.72	(0.44 - 1.18)	.1897	120/40	0.80	(0.45 - 1.40)	.4267	44/14	0.54	(0.17 - 1.72)	.2966
	Rec	TT-TC	224/95	1.00		174/74	1.00		50/19	1.00				
		CC	46/7	0.37	(0.16 - 0.88)	.0148	35/2	0.14	(0.03 - 0.62)	.0010	10/5	0.74	(0.18 - 3.00)	.6685
	Adi	TT,TC,CC	270/102	0.68	(0.47 - 0.98)	.0338	209/76	0.64	(0.42 - 1.00)	.0439	60/24	0.71	(0.34 - 1.51)	.3732

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla R-13. Asociación de los SNPs de *MUC2* rs10902073, rs10794281, rs2071174 y 7944723 con la regresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por *H.pylori*.

SNP	Modelo	Genotipo	Regresión			<i>H.pylori</i> +			<i>H.pylori</i> -			P-Inter.		
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p			
rs10902073	Cod	CC	88/50	1.00		57/45	1.00		29/5	1.00				
		CA	134/48	0.59	(0.36 - 0.97)	102/37	0.41 (0.23 - 0.73)		31/11	2.03	(0.58 - 7.04)			
		AA	41/11	0.47	(0.22 - 1.04)	.0549	36/9	0.34	(0.14 - 0.82)	.0034	5/2	1.45	(0.19 - 11.1)	.5249
	Dom	CC	88/50	1.00		57/45	1.00		29/5	1.00				
		CA-AA	175/59	0.56 (0.35 - 0.91)	.0190	138/46	0.39 (0.23 - 0.68)	.0008	36/13	1.92	(0.57 - 6.44)	.2791	.0100	
	Rec	CC-CA	222/98	1.00		159/82	1.00		60/16	1.00				
		AA	41/11	0.64	(0.31 - 1.33)	.2191	36/9	0.56	(0.25 - 1.27)	.1520	5/2	0.95	(0.15 - 6.17)	.9559
	Adi	CC,CA,AA	263/109	0.65 (0.46 - 0.94)	.0195	195/91	0.52 (0.34 - 0.79)	.0016	65/18	1.43	(0.59 - 3.44)	.4294	.0272	
rs10794281	Cod	TT	72/41	1.00		45/35	1.00		26/6	1.00				
		TC	121/45	0.60	(0.35 - 1.04)	89/37	0.47	(0.25 - 0.87)		30/8	1.26	(0.35 - 4.52)		
		CC	55/17	0.53	(0.27 - 1.06)	.1037	48/12	0.32 (0.14 - 0.71)	.0080	7/5	2.38	(0.50 - 11.2)	.5459	.0486
	Dom	TT	72/41	1.00		45/35	1.00		26/	1.00				
		TC-CC	176/62	0.58 (0.35 - 0.96)	.0361	137/49	0.42 (0.23 - 0.75)	.0032	37/6	1.53	(0.47 - 5.01)	.4787	.0375	
	Rec	TT-TC	193/86	1.00		134/72	1.00		56/13	1.00				
		CC	55/17	0.71	(0.38 - 1.33)	.2800	48/12	0.50	(0.24 - 1.03)	.0508	7/5	2.09	(0.53 - 8.20)	.2969
	Adi	TT,TC,CC	248/103	0.71 (0.50-1.00)	.0463	182/84	0.55 (0.37 - 0.82)	.0024	63/19	1.51	(0.69 - 3.29)	.2972	.0095	
rs2071174	Cod	TT	138/53	1.00		105/44	1.00		31/9	1.00				
		TC	110/40	0.93	(0.56 - 1.54)	78/34	0.98	(0.56 - 1.73)		31/6	0.77	(0.23 - 2.61)		
		CC	17/17	2.76 (1.24 - 6.13)	.0275	14/14	2.66	(1.09 - 6.50)	.0785	3/3	4.36	(0.61 - 31.3)	.2304	.7917
	Dom	TT	138/53	1.00		105/44	1.00		31/9	1.00				
		TC-CC	127/57	1.17	(0.73-1.86)	.5186	92/48	1.21	(0.71 - 2.05)	.4784	34/9	4.36	(0.34 - 3.20)	.9471
	Rec	TT-TC	248/93	1.00		183/78	1.00		62/15	1.00				
		CC	17/17	2.85 (1.32 - 6.14)	.0077	14/14	2.68 (1.13 - 6.32)	.0241	3/3	4.98	(0.76 - 32.7)	.0967	.6144	
	Adi	TT,TC,CC	265/110	1.36	(0.95 - 1.93)	.0893	197/92	1.37	(0.92 - 2.04)	.1168	65/18	1.45	(0.59 - 3.53)	.4181
rs7944723	Cod	CC	169/69	1.00		125/57	1.00		42/12	1.00				
		CG	83/33	0.91	(0.54 - 1.51)	61/27	0.90	(0.51 - 1.60)		21/6	0.91	(0.28 - 2.95)		
		GG	12/1	0.15	(0.02 - 1.22)	.0803	9/0	Na	Na	.0188	3/1	0.84	(0.07 - 9.76)	.9807
	Dom	CC	169/69	1.00		125/67	1.00		42/12	1.00				
		CG-GG	95/34	0.79	(0.48 - 1.31)	.3667	70/32	0.77	(0.44 - 1.35)	.3554	24/7	0.90	(0.29 - 2.77)	.8508
	Rec	CC-CG	252/102	1.00		186/84	1.00		63/18	1.00				
		GG	12/1	0.16 (0.02 - 1.25)	.0268	9/0	Na	Na	.0052	3/1	0.87	(0.08 - 9.73)	.9098	.0814
	Adi	CC,CG,GG	264/103	0.72	(0.47 - 1.12)	.1416	195/84	0.67	(0.41 - 1.12)	.1191	66/19	0.91	(0.36 - 2.30)	.8437

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Además, tres de los SNPs asociados con una menor (rs10902073, rs10794281) o mayor (rs2071174) probabilidad de regresión también se asociaron con un mayor (rs10902073, rs10794281) o menor (rs2071174) riesgo de progresión únicamente en los infectados por *H. pylori*, aunque sin ser la interacción positiva (TABLA R-14). El rs10794293, asociado a un menor riesgo de progresión (TABLA R-12), también se asoció con una mayor probabilidad de regresión únicamente en el grupo de infectados, en el modelo dominante (CT-TT, OR=1.76, 95% IC=1.02-3.03, $p=0.0400$, p interacción=0.1875).

Cuando se analizó el efecto de los polimorfismos de *MUC2* asociados significativamente con la evolución de lesiones precursoras en función del consumo de AINES, se observó que todos los SNPs, excepto el rs10902073 y el rs7944723, se asociaban significativamente con la evolución de lesiones únicamente en los no consumidores de este tipo de fármacos. De esta manera, los 3 SNPs significativamente asociados con un menor riesgo de progresión de lesiones (rs10794293, rs3924453 y 4077759, TABLA R-12) también mostraron un papel protector en los no consumidores de AINES (TABLA A-5). De los 4 SNPs asociados con la regresión de lesiones (TABLA R-13), el rs10794281 se asoció tanto con una menor probabilidad de regresión en los modelos codominante (TC OR=0.43, 95% IC=0.20-0.91; CC OR=0.33, 95% IC=0.12-0.90, $p=0.0284$), dominante (TC-CC, OR=0.40, 95% IC=0.20-0.80, $p=0.0089$) y aditivo (OR=0.54, 95% IC=0.33-0.88, $p=0.0107$), como con un mayor riesgo de progresión de lesiones en los modelos dominante (TC-CC, OR=2.44, 95% IC=1.16-5.11, $p=0.0151$) y aditivo (OR=1.64, 95% IC=1.04-2.58, $p=0.0303$, TABLA A-5) en los no consumidores. El SNP rs2071174, asociado previamente a una mayor probabilidad de regresión (TABLA R-13), se asoció como factor protector para la progresión en el grupo de no consumidores bajo los modelos de herencia dominante y aditivo (TABLA A-5). No obstante, no se observó una interacción significativa entre ninguna de las variantes de *MUC2* y el consumo de AINES.

Cuando se analizó el efecto de los SNPs de *MUC2* en la progresión de las lesiones en función del sexo, el rs10794293 se asoció únicamente en hombres y el rs4077759 únicamente en mujeres aunque la interacción no fue significativa y la tendencia en ambos grupos fue muy similar (TABLA A-6). Tres de los SNPs asociados con la regresión de las lesiones gástricas (rs10794281, rs2071174 y rs7944723, TABLA R-13) se asociaron únicamente en hombres en la estratificación por sexo aunque tampoco se observaron interacciones significativas (TABLA A-7).

Tabla R-14. Asociación de los SNPs de *MUC2* rs10902073, rs10794281 y rs2071174 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por *H.pylori*.

SNP	Modelo	Genotipo	Progresión			<i>H.pylori</i> +			<i>H.pylori</i> -			P-Inter.		
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p			
rs10902073	Cod	CC	107/31	1.00		83/19	1.00		23/11	1.00				
		CA	128/54	1.63	(0.95 - 2.80)	98/41	1.98	(1.03 - 3.81)	30/12	1.04	(0.34 - 3.13)			
		AA	35/17	1.69	(0.80 - 3.56)	.1586	29/16	2.14	(0.93 - 4.95)	.0747	6/1	1.04	(0.06 - 9.38)	.9713
	Dom	CC	107/31	1.00		83/19	1.00		23/11	1.00				
		CA-AA	163/71	1.64	(0.98 - 2.75)	.0552	127/57	2.02	(1.08 - 3.77)	.0233	36/13	1.01	(0.34 - 2.97)	.9893
	Rec	CC-CA	235/85	1.00		181/60	1.00		53/23	1.00				
		AA	35/17	1.26	(0.65 - 2.47)	.4964	29/16	1.41	(0.69 - 2.91)	.3529	6/1	0.75	(0.07 - 8.56)	.8160
Adi	CC,CA,AA	270/102	1.36	(0.95 - 1.93)	.0893	210/76	1.52	(1.01 - 2.27)	.0423	59/24	0.96	(0.39 - 2.39)	.9373	.2237
rs10794281	Cod	TT	89/24	1.00		66/14	1.00		22/10	1.00				
		TC	119/47	1.64	(0.90 - 2.96)	92/34	1.93	(0.93 - 4.02)	27/11	0.93	(0.29 - 2.95)			
		CC	49/23	1.79	(0.88 - 3.64)	.1685	39/21	2.41	(1.05 - 5.52)	.0818	10/2	0.70	(0.11 - 4.49)	.9278
	Dom	TT	89/24	1.00		66/14	1.00		22/10	1.00				
		TC-CC	168/70	1.68	(0.96 - 2.94)	.0621	131/55	2.08	(1.04 - 4.16)	.0313	37/13	0.88	(0.29 - 2.66)	.8172
	Rec	TT-TC	208/71	1.00		158/48	1.00		49/21	1.00				
		CC	49/23	1.33	(0.73 - 2.42)	.3545	39/21	1.58	(0.81 - 3.08)	.1800	10/2	0.73	(0.29 - 2.66)	.7154
Adi	TT,TC,CC	257/94	1.35	(0.96 - 1.91)	.0854	197/69	1.54	(1.03 - 2.32)	.0343	59/23	0.86	(0.38 - 1.98)	.7276	.1983
rs2071174	Cod	TT	132/59	1.00		102/47	1.00		29/11	1.00				
		TC	112/38	0.72	(0.43 - 1.19)	87/25	0.60	(0.33 - 1.09)	25/12	0.97	0.32 2.91			
		CC	28/6	0.48	(0.18 - 1.28)	.1973	23/5	0.46	(0.16 - 1.35)	.1292	5/1	0.40	0.03 4.63	.7280
	Dom	TT	132/59	1.00		102/47	1.00		29/11	1.00				
		TC-CC	140/44	0.67	(0.42 - 1.09)	.1064	110/30	0.57	(0.32 - 1.00)	.0488	30/13	0.89	0.30 2.59	.8244
	Rec	TT-TC	244/97	1.00		189/72	1.00		54/23	1.00				
		CC	28/6	0.56	(0.21 - 1.44)	.2053	23/5	0.57	(0.20 - 1.63)	.2721	5/1	0.41	0.04 4.32	.4264
Adi	TT,TC,CC	272/103	0.71	(0.48 - 1.04)	.0721	212/77	0.64	(0.41 - 1.00)	.0423	59/24	0.80	0.33 1.93	.6102	.5662

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

2.1.4.3 Genes involucrados en la respuesta a la infección por *H.pylori*.

No se encontraron asociaciones significativas entre la evolución de lesiones gástricas precursoras, progresión o regresión, y las variantes genotipadas en los genes de distintas interleucinas (TABLA A-4).

Únicamente tres polimorfismos en genes de respuesta a la infección por *H. pylori* se asociaron con la progresión de lesiones (TABLA A-4).

El rs2301756 del gen *PTPN11*, se asoció a un menor riesgo de progresión en el modelo recesivo, ($p=0.0237$, TABLA R-15). Debido a la relativa baja frecuencia de este SNP, $MAF=0.10$, no se pudo calcular el *odd ratio* ya que solamente se encontraron seis individuos CC, el homocigoto del alelo menos frecuente, en el grupo de los estables, pero ninguno en el grupo de progresión.

El rs1800629 de *TNF* mostró asociación con un mayor riesgo de progresión en los modelos codominante y recesivo (TABLA R-15). En este caso se trata también de un SNP no muy frecuente, $MAF=0.13$, y el *odd ratio* para el homocigoto del alelo menor AA se basa en la presencia de un único individuo en el grupo de estables y cinco individuos en el grupo de los que progresaron. Este hecho explicaría el elevado *odd ratio* observado y el amplio rango del intervalo de confianza.

El alelo G del rs2569190, en la región 5'UTR de *CD14* se asoció como factor protector de la progresión en los modelos codominante, recesivo y aditivo (TABLA R-15).

En relación a la regresión de lesiones es de destacar la asociación significativa de los dos polimorfismos de cambio de aminoácido analizados en *TLR4* (rs4986790, D299G y rs4986791, T399I) y una menor probabilidad regresión (TABLA R-15). Como en los casos anteriores, se trata de polimorfismos poco frecuentes, rs4986790 $MAF=0.06$ y rs4986791 $MAF=0.07$, y la baja presencia de homocigotos para el alelo menor (únicamente un individuo en cada caso) no permitió calcular los *odd ratios* del modelo recesivo. Las bajas frecuencias alélicas de los polimorfismos anteriormente citados no permitieron realizar análisis estratificados por sexo, consumo de AINES o infección por *H.pylori*.

Finalmente, el polimorfismo rs1125557 del gen de la E-cadherina, *CDH1*, se asoció significativamente a una mayor regresión en el modelo recesivo y mostró la misma tendencia en el modelo aditivo (TABLA R-15). El efecto de este SNP sólo se mantuvo en los individuos no infectados por *H.pylori* en el análisis estratificado bajo el modelo recesivo (GG, OR=6.61, 95% IC=1.29-33.92, valor- $p=0.0210$), aunque la interacción no fue significativa ($p=0.0562$). No se observaron diferencias significativas cuando se analizó este SNP en función del sexo o del consumo de AINES.

Tabla R-15. Asociación de los SNPs de *PTPN11*, *TNF*, *CD14*, *TLR4* y *CDH1* asociados con la evolución de lesiones gástricas.

SNP (gen)	Modelo	Genotipo	Regresión			Progresión			
			Contr/ Casos	OR (CI 95%) ^a	Valor p	Contr/ Casos	OR (CI 95%) ^a	Valor p	
rs2301756 (<i>PTPN11</i>)	Cod	TT	217/90	1.00		219/88	1.00		
		TC	46/20	1.09 (0.59-2.00)		50/16	0.78 (0.40-1.50)	.0582	
		CC	5/1	0.47 (0.05-4.37)	.7399	6/0	Na	Na	
	Dom	TT	217/90	1.00		219/88	1.00		
		TC-CC	51/21	1.02 (0.56-1.85)	.9411	56/16	0.67 (0.35-1.28)	.2160	
	Rec	TT-TC	263/110	1.00		269/104	1.00		
		CC	5/1	0.46 (0.05-4.30)	.4657	6/0	Na	Na	.0237
Adi	TT,TC,CC	268/111	0.97 (0.57-1.64)	.8962	275/104	0.62 (0.34-1.12)	.1004		
rs1800629 (<i>TNF</i>)	Cod	GG	220/104	1.00		239/85	1.00		
		GA	75/28	0.82 (0.49-1.37)		76/27	0.96 (0.57-1.64)		
		AA	6/0	Na	Na	.0868	1/5	19.1 (2.07-176.4)	.0080
	Dom	GG	220/104	1.00		239/85	1.00		
		GA-AA	81/28	0.76 (0.46-1.26)	.2795	77/32	1.15 (0.69-1.91)	.5872	
	Rec	GG-GA	295/132	1.00		315/112	1.00		
		AA	6/0	Na	Na	.0381	1/5	19.3 (2.09-177.3)	.0019
Adi	GG,GA,AA	301/132	0.71 (0.44-1.15)	.1586	316/117	1.36 (0.86-2.16)	.1910		
rs2569190 (<i>CD14</i>)	Cod	AA	71/31	1.00		67/35	1.00		
		AG	126/50	0.87 (0.50-1.52)		128/48	0.75 (0.43-1.31)		
		GG	59/28	1.20 (0.63-2.31)	.5664	70/17	0.41 (0.20-0.65)	.0485	
	Dom	AA	71/31	1.00		67/35	1.00		
		AG-GG	185/78	0.97 (0.58-1.63)	.9074	198/65	0.62 (0.37-1.06)	.0819	
	Rec	AA-AG	197/81	1.00		195/83	1.00		
		GG	59/28	1.31 (0.75-2.28)	.3414	70/17	0.49 (0.26-0.94)	.0248	
Adi	AA,AG,GG	256/109	1.09 (0.78-1.51)	.6260	265/100	0.66 (0.46-0.93)	.0169		
rs4986790 (<i>TLR4</i>)	Cod	AA	222/102	1.00		238/86	1.00		
		AG	35/6	0.39 (0.15-0.97)		26/15	1.62 (0.78-3.37)		
		GG	1/0	Na	Na	.0690	1/0	Na	Na
	Dom	AA	222/102	1.00		238/86	1.00		
		AG-GG	36/6	0.38 (0.15-0.95)	.0242	27/15	1.55 (0.75-3.19)	.2460	
	Rec	AA-AG	257/108	1.00		264/101	1.00		
		GG	1/0	Na	Na	.0298	1/0	Na	Na
Adi	AA,AG,GG	258/108	0.38 (0.15-0.94)	.0221	265/101	1.44 (0.72-2.88)	.1979		
rs4986791 (<i>TLR4</i>)	Cod	CC	224/102	1.00		238/88	1.00		
		CT	43/9	0.21 (0.21-0.99)		36/16	1.20 (0.61-2.37)		
		TT	1/0	Na	Na	.0842	1/0	Na	Na
	Dom	CC	224/102	1.00		238/88	1.00		
		CT-TT	44/9	0.20 (0.20-0.97)	.0305	37/16	1.16 (0.59-2.27)	.6671	
	Rec	CC-CT	267/111	1.00		274/104	1.00		
		TT	1/0	Na	Na	.0356	1/0	Na	Na
Adi	CC,CT,TT	268/111	0.21 (0.21-0.96)	.0281	275/104	1.11 (0.58-2.13)	.7578		
rs1125557 (<i>CDH1</i>)	Cod	AA	98/33	1.00		101/30	1.00		
		AG	128/51	1.18 (0.69-2.01)		127/52	1.41 (0.81-2.43)		
		GG	34/23	2.10 (1.04-4.23)	.1111	41/16	1.30 (0.62-2.75)	.4643	
	Dom	AA	98/33	1.00		101/30	1.00		
		AG-GG	162/74	1.37 (0.83-2.26)	.2213	168/68	1.38 (0.82-2.33)	.2226	
	Rec	AA-AG	226/84	1.00		228/82	1.00		
		GG	34/23	1.91 (1.02-3.57)	.0448	41/16	1.06 (0.54-2.08)	.3023	
Adi	AA,AG,GG	260/107	1.40 (0.99-1.98)	.0542	269/98	1.19 (0.83-1.69)	.3481		

a, Análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

2.1.5. Asociación de haplotipos con lesiones precursoras

Se analizó la asociación de los haplotipos inferidos para cada uno de los genes *MUC* y *TFF* (APARTADO 2.1.3 de resultados) con la progresión y regresión de lesiones gástricas. El análisis de asociación de los haplotipos del clúster de *TFFs* se realizó de manera individual para cada gen y un primer análisis, considerando el modelo aditivo, no indicó la existencia de haplotipos que se asociaran significativamente con la progresión o regresión de las lesiones (TABLA R-16).

Tabla R-16. Haplotipos de *TFFs* y asociación con la evolución de lesiones gástricas.

Gen	Haplotipo	Frec.	Progresión		Regresión		
			OR (IC 95%) ^d	Valor p	OR (IC 95%) ^d	Valor p	
<i>TFF1</i>^a	CTGGCCTGTCCG	0.277	1.00	Referencia	1.00	Referencia	
	CCGGCCCGCGG	0.142	1.05	(0.59 - 1.84)	.8785	1.07 (0.62 - 1.86)	.7967
	TCAGCCTGCGG	0.103	1.18	(0.64 - 2.18)	.6059	1.43 (0.82 - 2.49)	.2021
	TCAGTCCCCCG	0.077	1.11	(0.54 - 2.26)	.7773	0.99 (0.49 - 2.01)	.9761
	CTGGCCTGTGG	0.068	0.98	(0.43 - 2.26)	.9633	1.13 (0.55 - 2.32)	.7297
	CCAGCCTGCCG	0.052	0.48	(0.16 - 1.46)	.1980	0.76 (0.31 - 1.87)	.5454
	CCAGCCTGCGA	0.037	1.00	(0.38 - 2.61)	.9997	1.28 (0.49 - 3.38)	.6119
	TCAGCCTGTCCG	0.026	1.09	(0.37 - 3.17)	.8780	2.01 (0.69 - 5.89)	.2013
	TCAGCCTGCCG	0.018	1.69	(0.37 - 7.70)	.4985	1.06 (0.26 - 4.39)	.9372
	TCAGCCCAGCGG	0.016	0.60	(0.12 - 3.08)	.5385	2.47 (0.62 - 9.79)	.1986
	CCGGCCTGTCCG	0.016	0.21	(0.02 - 1.87)	.1599	0.56 (0.10 - 3.06)	.5001
	TCAGCCTGCGA	0.012	0.45	(0.07 - 3.03)	.4110	1.92 (0.33 - 11.28)	.4724
	CCGACCCGCGG	0.012	1.51	(0.39 - 5.82)	.5466	0.37 (0.05 - 2.81)	.3347
	CTGGCCTGCGG	0.012	1.80	(0.30 - 10.79)	.5210	2.47 (0.44 - 13.87)	.3050
	CCGGTCCGCGG	0.011	1.88	(0.32 - 11.14)	.4869	1.51 (0.19 - 12.20)	.6994
CCGGTCCCCCG	0.011	2.81	(0.45 - 17.65)	.2699	0.89 (0.14 - 5.53)	.8983	
CCGGCCTGCGG	0.011	4.06	(0.67 - 24.58)	.1268	Na	Na	
<i>TFF2</i>^b	AGAGGAATC	0.324	1.00	Referencia	1.00	Referencia	
	GTAAGAACC	0.188	1.22	(0.75 - 1.98)	.4153	1.13 (0.70 - 1.80)	.6225
	AGAAGBATC	0.150	1.26	(0.75 - 2.11)	.3910	1.18 (0.72 - 1.93)	.5226
	ATAAGAACC	0.102	1.38	(0.77 - 2.49)	.2823	0.96 (0.53 - 1.75)	.8933
	AGAGGAACC	0.080	0.84	(0.40 - 1.76)	.6500	1.53 (0.81 - 2.91)	.1939
	AGAGGBATC	0.031	2.26	(0.79 - 6.50)	.1287	1.41 (0.48 - 4.14)	.5326
	AGAAGAATC	0.024	0.41	(0.06 - 2.95)	.3737	1.76 (0.67 - 4.67)	.2528
	AGAGCAATC	0.016	0.86	(0.20 - 3.69)	.8417	1.82 (0.47 - 6.96)	.3836
	GTAAGAATC	0.014	2.09	(0.47 - 9.28)	.3314	1.34 (0.28 - 6.46)	.7173
	AGAAGBATT	0.011	0.98	(0.15 - 6.43)	.9837	3.70 (0.74 - 18.38)	.1097
<i>TFF3</i>^c	GCAGCGG	0.376	1.00	Referencia			
	AGAGCGA	0.191	1.18	(0.74 - 1.88)	.4900	1.13 (0.71 - 1.78)	.6132
	AGAGCGG	0.182	1.32	(0.82 - 2.13)	.2495	0.98 (0.61 - 1.55)	.9216
	AGAGAGG	0.166	0.94	(0.56 - 1.59)	.8147	1.13 (0.70 - 1.84)	.6090
	AGGGCGG	0.034	1.89	(0.76 - 4.66)	.1686	1.17 (0.43 - 3.15)	.7558
	GCAGCGA	0.015	1.06	(0.26 - 4.22)	.9386	Na	Na

a, Los haplotipos de *TFF1* incluyen los SNPs en el siguiente orden: rs13047838, rs424694, rs9976977, c.-29, rs2156310, rs34795821, rs225358, rs13051704, rs225357, rs225353, rs13276; b, Los haplotipos de *TFF2* incluyen los SNPs en el siguiente orden rs3814896, rs13052596, Q7P, rs1079380, V66L, VNTR (40rep=A, 48rep=B), c.*67_68insCTT (alelo sin inserción=A, InsCTT=C), rs225334, rs4920084; c, Los haplotipos de *TFF3* incluyen los SNPs en el siguiente orden: , , y rs692356, rs225363, rs11701143, V80M, rs2236705, c.*41G>A, rs8133510. d, Análisis de regresión logística no condicional, ajustado por sexo, edad, consumo de AINES y tipo de lesión en el reclutamiento, modelo aditivo.

Los polimorfismos de *TFF1* que se asociaron significativamente con una mayor probabilidad de regresión de lesiones gástricas lo hicieron bajo los modelos dominante y/o sobredominante (TABLA R-11). Por esa razón, se analizó el efecto dominante de los haplotipos

formados por los tres tagSNPs asociados previamente a regresión y que forman parte de un bloque de LD independiente (FIGURA R-25), localizado en la región genómica a 5' de *TFF1*. El haplotipo ACT, que contiene los alelos de los SNPs rs9976977, rs13047838 y rs424694 previamente asociados a regresión, se asoció a una mayor probabilidad de regresión cuando se comparó con el haplotipo de referencia GTC (TABLA R-17).

Tabla R-17. Haplotipos de la región genómica a 5' de *TFF1* y asociación con la evolución de lesiones gástricas

Haplotipo ^a	Frec.	Progresión			Regresión		
		OR (IC 95%) ^b	Valor p	OR (IC 95%) ^b	Valor p		
GTC	0.393	1.00		1.00	Referencia		
ACT	0.276	1.09	(0.65-1.80)	0.750	1.75	(1.06 - 2.87)	0.0282
GCC	0.218	1.13	(0.66-1.92)	0.656	0.99	(0.58 - 1.68)	0.9708
ACC	0.113	0.86	(0.45-1.64)	0.640	1.23	(0.67 - 2.27)	0.5069

a, Los haplotipos incluyen los SNPs de *TFF1* en el siguiente orden: rs9976977, rs424694, rs13047838. b, Análisis de regresión logística no condicional, ajustado por sexo, edad, consumo de AINES y tipo de lesión en el reclutamiento, modelo dominante.

Ninguno de los haplotipos de *TFF2* se asoció con la progresión y/o regresión de lesiones gástricas. El alelo G del rs1079380, asociado previamente como factor protector de la progresión, está incluido en el haplotipo más común AGAGGAATC (TABLA R-16), que se toma como referencia en el análisis de regresión. En el caso de *TFF3*, en que tampoco se observaron haplotipos asociados significativamente a la evolución de las lesiones, un único SNP, c.*41G>A, se asoció previamente con una mayor probabilidad de regresión de las lesiones. Debido a su baja frecuencia (MAF=0.01), el alelo asociado A tampoco queda representado en ninguno de los haplotipos inferidos (TABLA R-16).

Referente a mucinas, no se encontraron asociaciones significativas con la progresión o regresión de lesiones precursoras y los haplotipos inferidos de *MUC1* y *MUC6* (TABLA R-18), como era probable teniendo en cuenta el análisis individual de SNPs. No obstante, el haplotipo menos frecuente de *MUC6* CGCTT mostró una tendencia, aunque no significativa (OR=2.09, $p=0.0536$, TABLA R-18), como factor de riesgo para la progresión.

Por otro lado, el haplotipo CTCCCG de *MUC2*, en la región 5', se asoció significativamente con un menor riesgo de progresión de lesiones así como con una mayor probabilidad de regresión (TABLA R-18). Este haplotipo contiene el alelo C del rs2071174 que también se asoció con una mayor probabilidad de regresión y los dos alelos comunes del SNP rs10902073 (C/A) y rs10794281 (T/C), cuyas variantes menos frecuentes se asociaron con una menor regresión en el análisis individual de SNPs. En la región 3', el haplotipo CATAGAAC también se asoció a un menor riesgo de progresión (TABLA R-18). Este haplotipo contenía los tres alelos (T, A y C de rs10794293, rs3924453 y rs4077759, respectivamente) que también fueron asociados con un menor riesgo de progresión en el análisis individual de SNP. El haplotipo resultante de la combinación de estos dos haplotipos significativamente asociados a la evolución

de lesiones gástricas, en las regiones 5' y 3' (CTCCCGCATAGAAC, con una frecuencia de 0.120), también mostró una tendencia como factor protector de la progresión (OR=0.53, 95% IC: 0.26-1.07, $p=0.075$) cuando se comparó con el haplotipo ACCTCGCGCAAAGT (frecuencia de 0.147), aunque no alcanzó la significación estadística. Finalmente, en relación con los genes involucrados en la infección por *H.pylori*, el haplotipo GT de *TLR4* (frecuencia de 0.057) se asoció a una menor probabilidad de regresión (OR= 0.38, 95% IC=0.16-0.91, valor- $p=0.0308$) cuando se comparó con el haplotipo más común AC (frecuencia de 0.938). El haplotipo GT porta los dos alelos asociados también a una menor probabilidad de regresión en el análisis de SNPs individuales (rs4986790 y rs4986791, TABLA R-15), confirmando los resultados obtenidos previamente. Estos dos SNPs, de cambio de aminoácido, no son muy frecuentes y están en elevado LD ($D'=0.98$).

Tabla R-18. Haplotipos de genes *MUC* y asociación con la evolución de lesiones gástricas.

Gen	Haplotipo ^a	Frec.	Progresión			Regresión		
			OR (IC 95%) ^e	Valor p	OR (IC 95%) ^c	Valor p		
<i>MUC1</i> ^a	CAC	0.466	1.00	Referencia	1.00	Referencia		
	TTT	0.181	1.21	(0.72 - 2.02)	.4741	1.05	(0.65 - 1.69)	.8434
	TTC	0.137	1.06	(0.60 - 1.86)	.8398	0.71	(0.39 - 1.28)	.2538
	CAT	0.115	1.25	(0.64 - 2.46)	.5123	0.95	(0.49 - 1.84)	.8793
	CTC	0.083	0.92	(0.46 - 1.87)	.8233	0.77	(0.39 - 1.53)	.4584
	CTT	0.015	1.03	(0.20 - 5.24)	.9724	0.83	(0.15 - 4.47)	.8306
<i>MUC6</i> ^b	TACGG	0.373	1.00	Referencia	1.00	Referencia		
	CACGG	0.251	1.08	(0.69 - 1.67)	.7453	1.19	(0.78 - 1.82)	.4186
	CGCGG	0.150	1.47	(0.88 - 2.45)	.1367	0.77	(0.44 - 1.35)	.3688
	CGTGG	0.087	0.92	(0.46 - 1.85)	.8210	1.11	(0.59 - 2.09)	.7379
	CGCGT	0.069	0.95	(0.44 - 2.05)	.9006	1.45	(0.73 - 2.87)	.2932
	CGCTT	0.061	2.09	(0.99 - 4.40)	.0536	1.43	(0.67 - 3.06)	.3510
<i>MUC2 5'</i> ^c	ACCTCG	0.345	1.00	Referencia	1.00	Referencia		
	CTCTCG	0.270	0.72	(0.46 - 1.12)	.1478	1.58	(0.99 - 2.52)	.0548
	CTCCCG	0.158	0.49	(0.28 - 0.85)	.0117	1.67	(1.02 - 2.73)	.0407
	CTGCTG	0.046	0.81	(0.33 - 1.99)	.6502	1.72	(0.71 - 4.17)	.2268
	CTCTTG	0.039	0.60	(0.22 - 1.61)	.3080	1.25	(0.45 - 3.52)	.6697
	CTCTTA	0.035	0.84	(0.32 - 2.21)	.7279	0.69	(0.23 - 2.08)	.5114
	CCGCTA	0.033	0.57	(0.20 - 1.63)	.2963	1.81	(0.76 - 4.32)	.1786
	CCGCTG	0.020	0.83	(0.21 - 3.30)	.7898	0.29	(0.03 - 2.64)	.2707
	ACCCCG	0.015	0.46	(0.09 - 2.42)	.3584	2.38	(0.44 - 13.01)	.3156
	ACCTTG	0.014	0.20	(0.02 - 1.88)	.1607	1.49	(0.19 - 11.56)	.7014
<i>MUC2 3'</i> ^d	CGCAAAGT	0.322	1.00	Referencia	1.00	Referencia		
	CATAGAAC	0.140	0.46	(0.25 - 0.86)	.0156	1.16	(0.69 - 1.95)	.5810
	CGCAGAGT	0.093	0.77	(0.39 - 1.54)	.4583	1.00	(0.51 - 1.97)	.9989
	CGCAGAAC	0.089	0.73	(0.36 - 1.45)	.3667	0.86	(0.45 - 1.65)	.6439
	GATAAAGT	0.077	0.49	(0.21 - 1.14)	.0991	1.19	(0.59 - 2.39)	.6293
	GACAAAGT	0.070	1.51	(0.74 - 3.06)	.2558	0.52	(0.23 - 1.17)	.1130
	CGCGGGGC	0.067	0.83	(0.39 - 1.79)	.6378	0.57	(0.26 - 1.24)	.1552
	CATAGAGT	0.034	0.93	(0.34 - 2.51)	.8835	1.34	(0.51 - 3.57)	.5523
	GATAAAC	0.015	1.26	(0.30 - 5.27)	.7532	0.86	(0.19 - 3.87)	.8436
GATAGAAC	0.018	0.86	(0.17 - 4.29)	.8543	0.43	(0.07 - 2.70)	.3679	

a, Los haplotipos de *MUC1* incluyen los SNPs en el siguiente orden: rs3814316, rs9426886 y rs1045253; b, *MUC6* incluye los SNPs en el siguiente orden: rs4076950, rs7481521, rs11246384, rs6597947 y rs9794921; c, La región 5' de *MUC2* los SNPs en el siguiente orden: rs10902073, rs10794281, rs2856082, rs2071174, rs7396030, rs11245936; d, La región 3' incluye los SNPs en el siguiente orden: rs7944723, rs6421972, rs10794293, rs11245954, rs7480563, rs7126405, rs3924453, rs4077759; e, Análisis de regresión logística incondicional, ajustado por sexo, edad, consumo de AINES y tipo de lesión en el reclutamiento, modelo aditivo.

2.2. Estudio caso-control de cáncer gástrico anidado en la cohorte EPIC

El análisis de asociación de polimorfismos ($n=42$), y haplotipos inferidos, en genes de protección de la mucosa gástrica (*TFFs* y *MUCs*) con cáncer gástrico se realizó en un grupo de 1649 muestras procedentes de un estudio caso-control anidado en la cohorte EPIC (*European Prospective Investigation into Cancer and Nutrition*, APARTADO 1.2 de materiales y métodos) para poder establecer su posible implicación en el desarrollo de este tipo de cáncer.

2.2.1. Descriptiva de la población y asociación de covariables con cáncer gástrico

De las 2005 muestras inicialmente genotipadas, 64 (10 casos y 54 controles) fueron excluidas del análisis porque no pasaron los controles de calidad correspondientes: 10 casos y 49 controles no amplificaron correctamente, 4 controles tuvieron un porcentaje de genotipado de SNPs menor del 80% y 1 control, de las 101 muestras duplicadas, no concordó con su réplica. Además, 292 muestras fueron excluidas a posteriori por criterios de clasificación de los casos y controles (no tumores primarios, tumores prevalentes o controles cuyos casos se habían eliminado del estudio). Finalmente el estudio caso-control incluyó 365 casos (334 de cáncer gástrico y 31 de la unión gastroesofágica) y 1284 controles.

Se realizó una estadística descriptiva de las principales características de la población caso-control disponibles para este estudio y se analizó su posible asociación con cáncer gástrico mediante regresión logística. En la TABLA R-19 se muestra la descriptiva de la población a nivel global y por grupo caso o control.

Los casos y controles (alrededor de cuatro por cada caso), se aparearon por edad, sexo, país de procedencia y fecha de extracción de sangre, por lo que su distribución según estas variables es prácticamente idéntica en ambos grupos. La edad de reclutamiento de los participantes estuvo comprendida entre los 28 y 78 años, con una media de 58.4 ± 7.74 . Los individuos procedieron de centros de hasta 10 países diferentes en porcentajes muy diversos (0.30%-17.2%) y la presencia de hombres (59%) en el estudio fue ligeramente superior a la de mujeres (41%), en concordancia con la mayor incidencia de este tipo de cáncer en el sexo masculino.

Globalmente, una gran parte de la población fue positiva para la infección de *H.pylori* (65.7%). pero se observó un porcentaje mucho mayor de infectados entre los casos, siendo la diferencia entre grupos altamente significativa (test χ^2 , $p=1.84 \times 10^{-12}$, TABLA R-19). Estos resultados se confirmaron en el análisis mediante regresión logística, ya que la infección por la bacteria se asoció, muy significativamente, con un mayor riesgo de cáncer gástrico con una OR de 2.73 (TABLA R-20), en concordancia con su papel ya establecido como uno de los factores de riesgo más importantes en la patogénesis del este tipo de cáncer. Todos los individuos del estudio infectados por *H.pylori* se clasificaron según la cepa de virulencia CagA y, globalmente, la mayoría de estos individuos (78%) fueron positivos para esta cepa (TABLA R-19). El porcentaje de los portadores aumentó ligeramente entre los casos de cáncer gástrico, siendo las diferencias entre los dos grupos de nuevo significativas. De hecho, los individuos CagA + se asociaron a un mayor riesgo de cáncer gástrico (TABLA R-20). En cuanto al consumo de tabaco, un 40.6% de la

población se declaró como no fumadora, un 32.9% como no fumadora y un 24.9% como ex fumadora (TABLA R-19). Estos porcentajes variaron sensiblemente entre casos y controles, mostrando los casos un menor porcentaje de no fumadores y mayor porcentaje de fumadores en comparación con la población control, siendo significativa la diferencia observada entre estos grupos. Estas diferencias se confirmaron en el análisis de regresión logística donde el consumo de tabaco se asoció significativamente como factor de riesgo para cáncer gástrico (TABLA R-18). En el presente estudio se utilizó el nivel educacional para caracterizar la posición socioeconómica de los participantes, dividiéndose en cinco categorías que se detallan en la TABLA R-19. Aunque se observa una menor proporción de casos con niveles educativos superiores en comparación con los controles, en general, no existen diferencias significativas entre los dos grupos.

Tabla R-19. Características principales de la población caso-control de cáncer gástrico.

Características	Total		Controles		Casos		Valor p ^a
	N	%	N	%	N	%	
Sexo							
Hombre	973	(59.0)	759	(59.1)	214	(58.6)	
Mujer	676	(41.0)	525	(40.9)	151	(41.4)	Apareado
Pais							
Francia	5	(0.3)	3	(0.2)	2	(0.6)	
Italia	262	(15.9)	206	(16.0)	56	(15.3)	
España	175	(10.6)	134	(10.4)	41	(11.2)	
Reino Unido	176	(10.7)	135	(10.5)	41	(11.2)	
Holanda	125	(7.6)	99	(7.7)	26	(7.1)	
Grecia	112	(6.8)	88	(6.8)	24	(6.6)	
Alemania	234	(14.2)	186	(14.5)	48	(13.2)	
Suecia	284	(17.2)	220	(17.1)	64	(17.5)	
Dinamarca	266	(16.1)	205	(16.0)	61	(16.7)	
Noruega	10	(0.6)	8	(0.6)	2	(0.6)	Apareado
Infección <i>H.pylori</i>							
No	560	(34.3)	492	(38.7)	68	(18.8)	
Si	1073	(65.7)	779	(61.3)	294	(81.2)	1.84x10 ⁻¹²
Cepa CagA							
Negativo	236	(22.0)	195	(25.0)	41	(13.9)	
Positivo	837	(78.0)	584	(75.0)	253	(86.1)	9.21x10 ⁻⁶
Consumo tabaco							
Nunca	670	(41.3)	542	(43.0)	128	(35.5)	
Ex fumador	542	(33.4)	427	(33.9)	115	(31.8)	
Habitual	410	(25.3)	292	(23.1)	118	(32.7)	0.0008
Educación							
Ninguna	103	(6.4)	82	(6.5)	21	(5.9)	
Primaria	664	(41.3)	507	(40.5)	157	(44.0)	
Esc.técnica	396	(24.6)	305	(24.4)	91	(25.5)	
Secundaria	180	(11.2)	140	(11.2)	40	(11.2)	
Niveles superiores	266	(16.5)	218	(17.4)	48	(13.4)	0.4341

a, Valor p de las diferencias entre los grupos de evolución para cada variable basado en el test χ^2 .

Tabla R-20. Asociación de los factores de riesgo con cáncer gástrico

Variables	Controles/ Casos	OR	(CI 95%) ^a	Valor p
Infección <i>H.pylori</i>				
No	492/68	1.00		
Si	779/294	2.73	(2.05-3.64)	6.74x10⁻¹²
CagA				
Negativo	195/41	1.00		
Positivo	584/253	2.06	(1.43-2.98)	1.17x10⁻⁴
Consumo tabaco				
Nunca	542/128	1.00		
Ex fumador	427/115	1.14	(0.86-1.51)	0.3611
Habitual	292/118	1.71	(1.28-2.82)	0.0003
Nivel educacional				
Ninguna	82/21	1.00		
Primaria	507/157	1.21	(0.72-2.02)	0.4669
Esc.	305/91	1.16	(0.68-1.97)	0.5747
Secundaria	140/40	1.12	(0.62-2.02)	0.7818
Niveles superiores	218/48	0.86	(0.48-1.52)	0.6048

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia.

Además, se dispone información sobre la dieta y consumo de alcohol de los participantes cuya descriptiva se resume en la TABLA A-8, conjuntamente con la edad. Se observaron diferencias significativas entre casos y controles en cuanto al consumo de frutas, que disminuyó entre los casos, y consumo de carnes procesadas, que por el contrario aumentó entre los casos. El estudio de las variables relacionadas con los hábitos de vida, como la dieta, la posición socioeconómica o el consumo de tabaco, requieren de un análisis más exhaustivo para determinar e interpretar su posible asociación con cáncer gástrico, que no es objeto de esta tesis y ya está descrito en relación a la cohorte EPIC (107,114).

Todos los casos se clasificaron según su localización anatómica (29.3% en el cardias o unión gastroesofágica, 49.6% no cardias y 21.1% no especificado) y tipo histológico (34% intestinal, 35.1% difuso y 30.4% no especificado, TABLA R-21), clasificación que fue validada por un experto grupo de patólogos (368). Los casos no especificados para las clasificaciones anteriores no se tuvieron en cuenta para las descriptivas y análisis de asociación estratificados por los diferentes subtipos de cáncer gástrico. La clasificación final del subtipo histológico constó de un porcentaje casi equitativo de casos del tipo intestinal (49.6%, n=126) y difuso (51.4%, n=128), y en cuanto al subtipo anatómico, se clasificó un porcentaje bastante superior de casos no cardias (62.8%, n=181) que en el cardias (37.2%, n=107). De los 200 casos para los cuales se dispuso de ambas clasificaciones, 109 casos del tipo intestinal se distribuyeron de manera similar entre los subtipos anatómicos (66% del no cardia y 44% del cardias). Sin embargo, en el tipo difuso (n=91) se observó un porcentaje mucho mayor de casos no cardias (80.2%) que del cardias (19.8%), siendo las diferencias entre subgrupos significativa (test χ^2 , $p=2.8 \times 10^{-4}$).

En la TABLA R-21 se muestra la descriptiva de las principales características de los casos según la localización anatómica y tipo histológico.

Tabla R-21. Descriptiva de las principales características de los casos de cáncer gástrico según la localización anatómica y tipo histológico.

Características	Subtipo anatómico			Subtipo histológico		
	Cardia (N=107)	No cardia (N=181)	Valor p ^a	Intestinal (N=126)	Difuso (N=128)	Valor p ^a
Sexo						
Hombre	83 (77.6)	85 (47.0)		83 (65.9)	59 (46.1)	
Mujer	24 (22.4)	96 (53.0)	3.56x10⁻⁷	43 (34.1)	69 (34.1)	0.0015
Infección <i>H.pylori</i>						
No	42 (39.3)	15 (8.4)		25 (20.0)	16 (12.6)	
Si	65 (60.7)	163 (91.6)	2.98x10⁻¹⁰	100 (80.0)	111 (87.4)	0.1115
Cepa CagA						
Negativo	22 (33.8)	12 (7.4)		12 (12.0)	15 (13.5)	
Positivo	43 (66.2)	151 (92.6)	4.01x10⁻⁷	88 (88.0)	96 (86.5)	0.7424
Consumo tabaco						
Nunca	28 (26.4)	73 (40.6)		43 (34.3)	52 (41.6)	
Ex fumador	40 (37.7)	47 (26.1)		42 (33.6)	32 (25.6)	
Habitual	38 (35.8)	60 (33.3)	0.0323	40 (32.0)	41 (32.8)	0.3302
Educación						
Ninguna	2 (1.9)	15 (8.4)		11 (8.9)	7 (5.6)	
Primaria	45 (43.3)	86 (48.0)		59 (48.0)	55 (44.0)	
Esc.técnica	33 (31.7)	38 (21.2)		28 (22.8)	31 (24.8)	
Secundaria	10 (9.6)	19 (10.6)		15 (12.2)	14 (11.2)	
Nivel superior	14 (13.5)	21 (11.7)	0.0914	10 (8.1)	18 (14.4)	0.4800

a, Valor p de las diferencias entre los grupos de evolución para cada variable basado en el test χ^2 .

Aunque ambos sexos presentan una incidencia relativamente similar por lo que respecta a los casos no cardias y del tipo difuso, se observa una proporción mucho mayor de hombres entre los casos del cardias y del tipo intestinal, siendo las diferencias de la distribución de los subtipos anatómico e histológico entre sexos muy significativas. La presencia de la infección por *H.pylori* es muy superior en los casos del no cardias, alcanzando prácticamente la totalidad de los casos de este tipo (91.6%), mientras que los casos del cardias presentan un porcentaje de infección de la bacteria muy similar a la que presentaron los controles (TABLA R-19). De hecho, el análisis de regresión indica que este factor se asocia a un riesgo muy elevado para cáncer gástrico no cardias (OR=6.86, 95%IC=4.00-11.79; $p=2.92 \times 10^{-12}$) mientras que su efecto en el cardias no es nada significativo (OR=0.98, 95%IC=0.65-1.46; $p=0.912$). De la misma manera, casi la totalidad de casos del no cardias (92.6%), entre los infectados por *H.pylori*, son portadores de la cepa de virulencia CagA, proporción mucho mayor que la observada en los casos localizados en el cardias (66.2%). En cuanto a los tipos histológicos, intestinal y difuso, no se observaron diferencias significativas. Aunque se observaron ligeras diferencias en el consumo de tabaco entre los subtipos anatómicos e histológicos de cáncer gástrico, siendo el porcentaje de fumadores y ex fumadores mayor en el tipo intestinal y el cardias, las diferencias fueron muy poco significativas.

2.2.2. Genotipado de los polimorfismos

En la TABLA R-6 se resumen las principales características de todos los polimorfismos en *TFFs* y *MUCs* incluidos en este estudio y las técnicas de genotipación utilizadas en cada caso; y en la TABLA A-9 del apéndice II se detallan las frecuencias alélicas obtenidas en el genotipado de cada uno de los polimorfismos en la totalidad de la población, así como por grupos casos y controles de cáncer gástrico.

De los 58 SNPs inicialmente seleccionados en *TFFs* (n=33) y *MUCs* (n=25), el rs7396030 de *MUC2* se eliminó del estudio porque no cumplió los requisitos necesarios para su genotipado en Illumina. Además, dos tagSNPs de *TFF1* de la selección inicial, rs225357 y rs4920094, se sustituyeron por SNPs marcadores de las mismas características, rs225356 y rs2839488, respectivamente, para facilitar su análisis en la plataforma de alto rendimiento (TABLA R-6). Por otra parte, los requisitos exigidos por la metodología de Illumina restringieron considerablemente la posibilidad de genotipar la mayoría de los polimorfismos potencialmente funcionales de *TFFs*. La inserción de 3 pb, c.*67_68insCTT, y el VNTR de *TFF2* no se incluyeron inicialmente por no ser polimorfismos del tipo SNP. Además, se excluyeron otros 7 SNPs potencialmente funcionales previamente validados o identificados por no tener un diseño apropiado para su genotipado: 3 en *TFF1* (c.*62, rs34795821, rs2156310), 3 en *TFF2* (rs4920084, Q7R, rs13052596) y 1 en *TFF3* (c.*41). Uno de estos SNPs no incluidos, el c.*62 de *TFF1*, resultó ser monomórfico en la población de pacientes con lesiones gástricas. De los polimorfismos excluidos de la plataforma Illumina, se genotiparon en la población caso-control aquellos para los cuales disponíamos de una técnica de genotipado alternativa de rendimiento medio: c.*67_68insCTT de *TFF2* mediante HRM, conjuntamente con el rs225334, y el c.*41G>A de *TFF3*, mediante sondas de hibridación FRET. Finalmente se incluyeron únicamente 9 SNPs potencialmente funcionales (2 en *TFF1*, 4 en *TFF2* y 3 en *TFF3*, TABLA R-6). Otros 6 SNPs se excluyeron por problemas de genotipación (no amplificación y señal o diferenciación de los clústeres de genotipos defectuosa): 1 en *MUC6*, 2 en *MUC2*, 2 en *TFF3* y 2 en *TFF1* (TABLA A-9). Además se excluyó el rs7277409 de *TFF2* por encontrarse únicamente un heterocigoto en toda la población (frecuencia de 0.0004). Globalmente, de los 1536 SNPs genotipados con la metodología de Illumina Golden Gate, se excluyeron un total de 214 por las mismas razones. El total de SNPs genotipados fue, por lo tanto, de 1287 (83.8%) en 249 genes/regiones genómicas diferentes (165). El promedio de muestras genotipadas en Illumina fue del 97.1% para los 42 SNPs de *TFFs* (n=21) y *MUCs* (n=21) finalmente analizados en esa plataforma. Para los polimorfismos c.*67_68insCTT y rs225334 de *TFF2*, analizados simultáneamente, el promedio fue del 96.2% y para el SNP c.*41 de *TFF3* fue tan sólo del 77.6%. Globalmente, la concordancia de genotipos de las muestras control duplicadas incluidas fue del 99.2%, para los más de 1000 SNPs analizados, y no se observaron errores de genotipado en ninguno de los SNPs de *MUCs* y *TFFs* analizados. Todos los SNPs genotipados con éxito cumplieron el equilibrio de Hardy-Weinberg (Fisher's exact test $p > 10^{-4}$) en la población control (TABLA A-9), por lo que no se excluyó ninguno del análisis de asociación.

Aunque las frecuencias alélicas son muy variables, desde 0.01, o incluso algo menores, hasta 0.49, son pocos SNPs los que presentan MAF muy bajas y, en su totalidad, son SNPs potencialmente funcionales identificados en este estudio. La mayoría de los SNPs analizados son tagSNPs con frecuencias mayores de 0.05, un criterio indispensable para su selección. Tan sólo un tagSNP presentó una MAF inferior a 0.05, el rs11203197 (MAF=0.03), que se seleccionó como tagSNP a partir de los datos de HapMap, donde aparecía con una frecuencia de 0.075.

2.2.3. Análisis de LD y determinación de haplotipos

Para el análisis de LD, se calcularon las medidas D' y r^2 , entre pares de polimorfismos de cada uno de los genes *TFFs* y *MUCs* genotipados en la población caso-control de cáncer gástrico. En el caso de *TFFs*, este análisis también se realizó a nivel del clúster ya que los genes se encuentran próximos en la región cromosómica 21q22.3. Con los datos obtenidos previamente en la genotipación y en el análisis de LD se infirieron los haplotipos para cada uno de los genes, seleccionando únicamente aquellos con frecuencia superior al 0.01. Los haplotipos inferidos se muestran en las TABLAS R-30 (*TFFs*) y R-31 (*MUCs*) del APARTADO 2.2.4, conjuntamente con los resultados de asociación con cáncer gástrico.

Como se había observado anteriormente, tanto en la población CEU de HapMap como en la población de pacientes con lesiones gástricas, no se observan bloques de LD compactos ni en la región cromosómica 21q22.3 donde se localiza el clúster de *TFFs* (FIGURA A-3A y A-3B del apéndice III) ni en cada uno de los locus que lo forman (FIGURA R-27). Teniendo en cuenta la complejidad y el poco LD del clúster, el análisis detallado de LD y la inferencia de haplotipos se realizó individualmente para cada gen *TFF* y región adyacente (FIGURA R-27).

La estructura de LD en el gen *TFF1* en la población caso-control es idéntica a la ya identificada en la población de pacientes con lesiones gástricas. Se observa claramente un bloque de LD compacto localizado a 5' del gen y formado por tres SNPs (rs9976977, rs424694, rs130478838). La región génica de *TFF1* no está abarcada por un único bloque, a pesar de su pequeño tamaño (4.2 kb), sino que se distinguen dos posibles bloques. Uno formado por los 3 SNPs localizados en el intrón 1 del gen y otro formado por un SNP en el intrón 2 y un SNP en la región 3'UTR. El rs4920094 se sustituyó por otro tagSNP de las mismas características, el rs2839488, ambos localizados en el intrón 1. Otro de los tagSNPs inicialmente seleccionados, el rs225358 en el intrón 1 también, se excluyó por problemas de genotipado. No obstante, este SNP no era marcador de ningún otro polimorfismo por lo que se pierde información de una única variante en la región. De esta manera, si se toman en cuenta los bloques adyacentes al rs225358, la reducción de variabilidad cubierta en la zona respecto a la selección inicial es únicamente del 3%. Los 8 polimorfismos analizados en *TFF1* y región adyacente forman un total de 15 haplotipos diferentes con una frecuencia superior de 0.01 (TABLA R-30). El gran número de haplotipos observados y su frecuencia, únicamente dos haplotipos superan una frecuencia del 10%, evidencian de nuevo el poco LD existente en esta región cromosómica. En el caso de *TFF2*, la estructura observada es muy similar a la de *TFF1*, en que no existe un bloque de LD claro que englobe totalmente el gen.

En *TFF2* se observan dos zonas de LD, una formada por los SNPs localizados en el extremo 3' del gen y otra formada por los SNPs en la región 5' y región genómica adyacente (FIGURA R-27). A pesar de los pocos SNPs analizados en *TFF2*, algunos de los tagSNPs analizados en *TFF1* cubren también la región genómica a 5' e incluso regiones génicas de *TFF2* debido a la proximidad de ambos genes en el clúster, (bloque 3 de la FIGURA R-15 y TABLA R-2). Por esta razón, SNPs puntuales de ambos genes presentan valores D' bastante elevados (0.8-1) entre ellos aunque se localicen en zonas de poco LD.

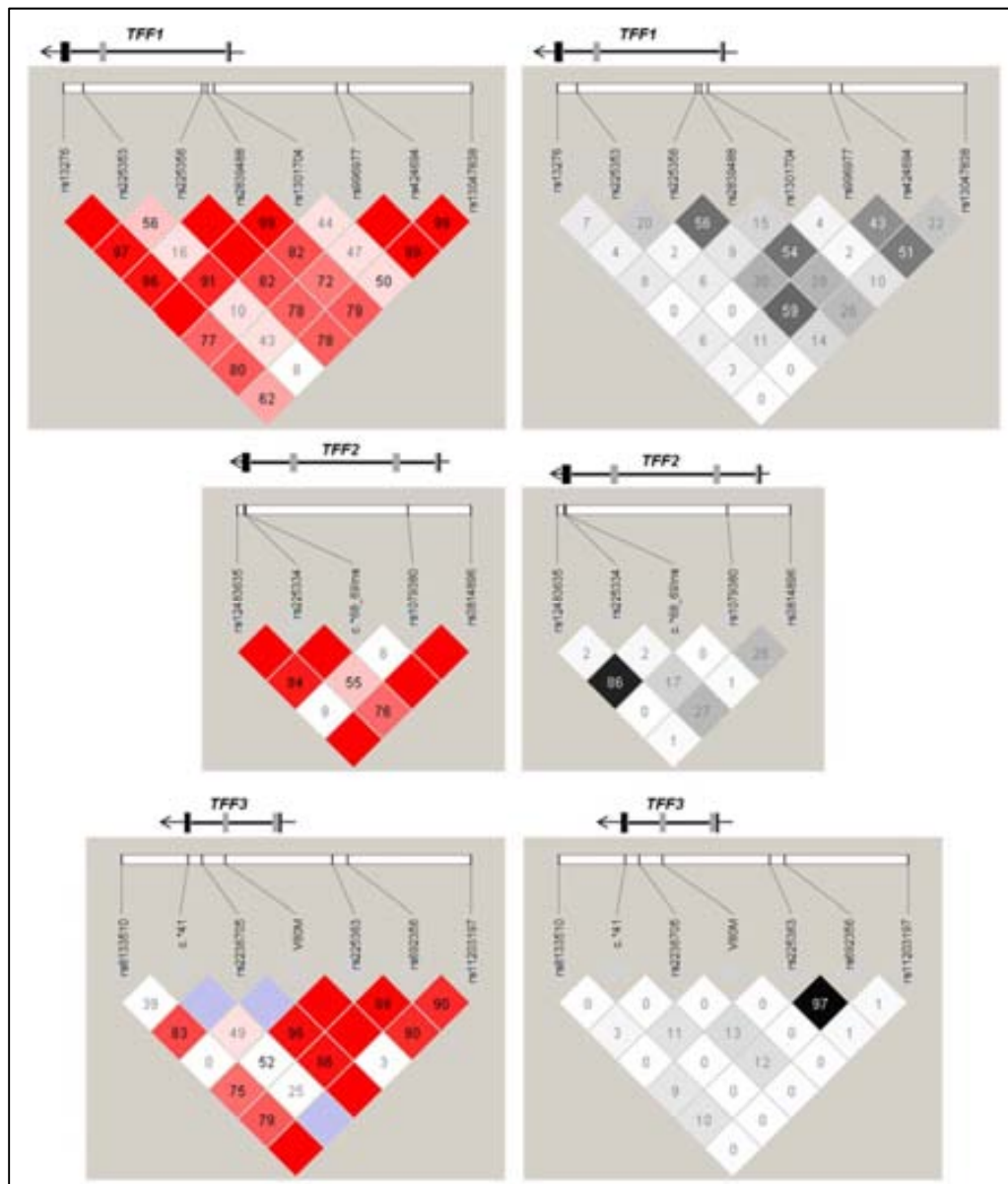


Figura R-27. Análisis de desequilibrio de ligamiento entre polimorfismos de las regiones en 21q22.3 donde se localizan *TFF1*, *TFF2* y *TFF3*, a partir de los datos de genotipado en la población caso-control de cáncer gástrico. En todos los casos se indica la estructura de cada uno de los genes Izquierda, representación gráfica del coeficiente D' . Los valores D' de 1 no se muestran. Esquema de color: cuadrados en blanco, $D' < 1$ y $\text{LOD} < 2$; en azul, $D' = 1$ y $\text{LOD} < 2$; en tonos rosa y rojo, $D' < 1$ y $\text{LOD} \geq 2$ y en rojo fuerte, $D' = 1$ y $\text{LOD} \geq 2$. Derecha, representación gráfica del coeficiente r^2 . Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

Además, todos los tagSNPs localizados en *TFF2* y región adyacente se genotiparon con éxito en este estudio por lo que se puede concluir que la cobertura de la variabilidad de este gen fue muy elevada. No obstante, el polimorfismo V66L se excluyó del análisis de LD y determinación de haplotipos ya que su frecuencia alélica en la población caso-control fue de 0.001. Con los 5 SNPs finalmente analizados en *TFF2* se infirieron 8 haplotipos en esta región, presentando frecuencias comprendidas entre 0.022 y 0.388 (TABLAS R-30).

TFF3 presenta un bloque compacto de LD formado por los SNPs situados en la región genómica a 5' (rs22563, rs692356 y rs11203197), de manera similar a lo observado en *TFF1* (FIGURA R-27). Estos polimorfismos están, además, en elevado LD ($D' > 0.8$) con todos los otros SNPs frecuentes analizados en la región. Sin embargo, los SNPs potencialmente funcionales y menos frecuentes (c.*41G>A y V80M, MAF-0.01) rompen claramente el LD en la zona por lo que el resultado final no es un bloque de LD tan consistente como el que se había observado previamente en la población de pacientes con lesiones gástricas (FIGURA R-25) y que englobaba el gen *TFF3* en su totalidad. De hecho, a causa de su baja frecuencia, uno de estos SNPs (V80M) no queda representado en ninguno de los haplotipos inferidos con frecuencia superior al 0.01. Aunque uno de los cinco tagSNPs (rs11701143) que marcaban el bloque que incluía todo el gen *TFF3* (bloque 1 en TABLA R-2), se excluyó debido a problemas en la genotipado, la cobertura de la variabilidad genética en la región permaneció inalterada ya que está marcado por otro tagSNP incluido, rs11203197, marcador del bloque localizado en la región genómica a 5' del gen (bloque 2 en TABLA R-2). Sin embargo, la exclusión del rs9981660, SNP que tampoco se pudo genotipar con éxito en la población de pacientes con lesiones precursoras, resultó en una reducción de la cobertura de la variabilidad genética en el gen *TFF3* del 12% o de únicamente el 6%, si se tienen en cuenta los SNPs en región genómica a 5' del gen. En este caso, el mayor LD observado en este gen también se refleja en la formación de haplotipos en que los 8 SNPs genotipados forman únicamente 6 haplotipos diferentes con frecuencias entre 0.015 y 0.403 (TABLAS R-30).

En cuanto al coeficiente de correlación r^2 , que depende de la frecuencia alélica de los polimorfismos, únicamente dos valores fueron mayores de 0.8: entre el rs12483635 y la inserción de *TFF2*, un polimorfismo potencialmente funcional del cual desconocíamos con antelación su relación de LD, y entre el rs225363 y rs692356 de *TFF3*, seleccionados inicialmente en dos bloques diferentes (derecha de la FIGURA R-27).

En la población caso-control los genes *MUC* presentan, en general, un mayor grado de LD en comparación con los *TFFs*, confirmando los resultados obtenidos previamente a partir de los datos de HapMap y en la población de pacientes con lesiones gástricas.

Los cuatro tagSNPs inicialmente seleccionados en el gen *MUC1*, y que marcaban una región de -55 kb, se genotiparon con éxito proporcionándonos una cobertura de la variabilidad genética en la región del 100%. La estructura de LD observada, constituida por dos bloques diferenciados (FIGURA R-28), es idéntica a la obtenida anteriormente en las dos poblaciones. El SNP marcador y funcional rs4072037, que es el único localizado propiamente en el gen *MUC1*, se encuentra en elevado LD ($D > 0.95$) tanto con los SNPs situados en la región genómica a 3' del gen, localizados en el gen adyacente *TRIM46*, como con el SNP situado en la región genómica a 5'.

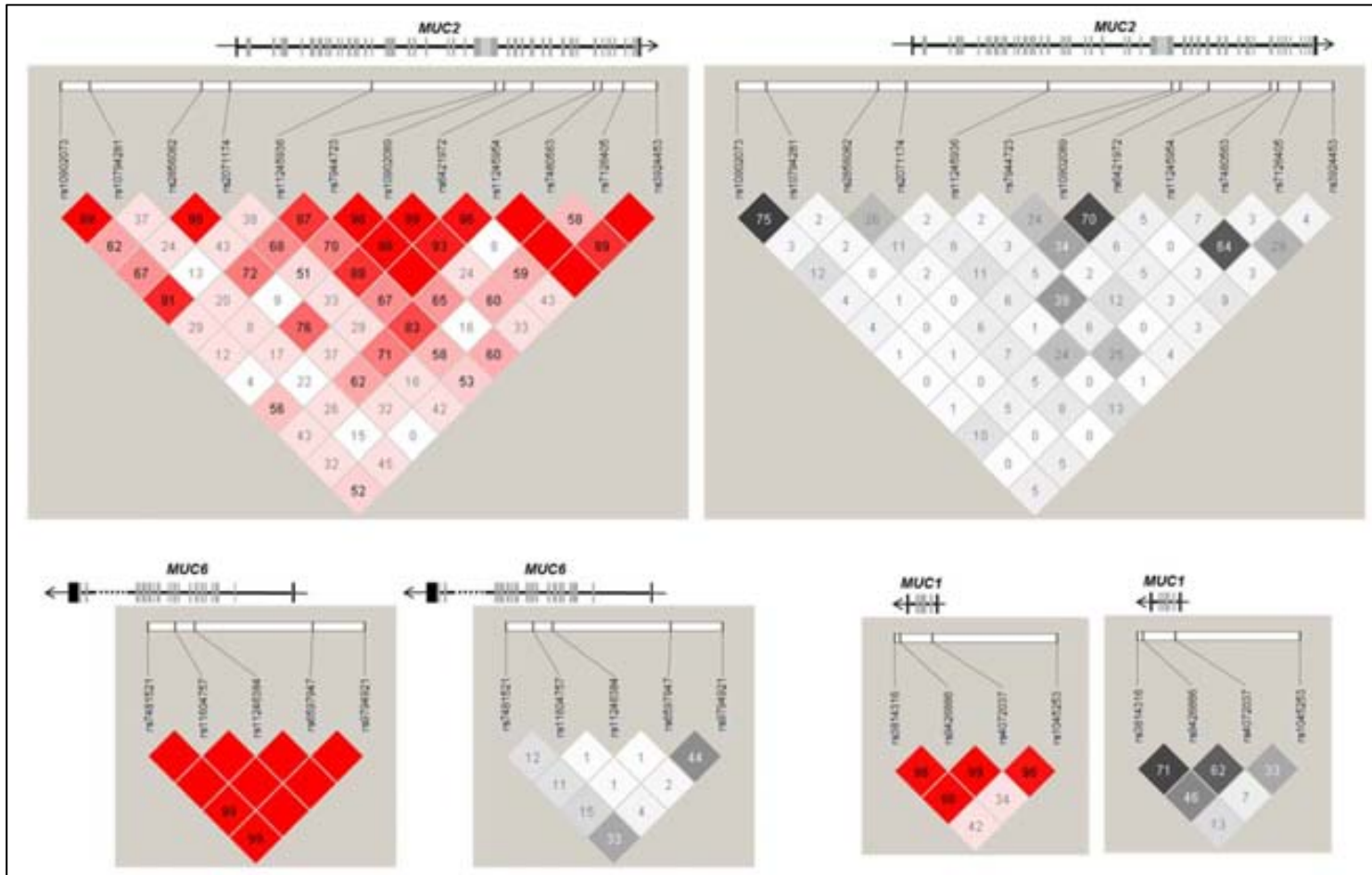


Figura R-28. Análisis de desequilibrio de ligamiento entre polimorfismos de las regiones donde se localizan *MUC1*, en 1q21, y *MUC2* y *MUC6*, en 11p15, a partir de los datos de genotipado la población caso-control de cáncer gástrico. En todos los casos se indica la estructura de cada uno de los genes. Derecha, representación gráfica del coeficiente D' . Los valores D' de 1 no se muestran. Esquema de color: cuadrados en blanco, $D' < 1$ y $\text{LOD} < 2$; en azul, $D' = 1$ y $\text{LOD} < 2$; en tonos rosa y rojo, $D' < 1$ y $\text{LOD} \geq 2$ y en rojo fuerte, $D' = 1$ y $\text{LOD} \geq 2$. Izquierda, representación gráfica del coeficiente r^2 . Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

A partir de los resultados de los 4 SNPs genotipados, se infirieron un total de 6 haplotipos con frecuencias de entre 0.011 y 0.463. Como ya se observó en la población de pacientes con lesiones gástricas, el haplotipo CAGC, que porta los alelos más comunes de cada uno de los SNPs que lo forman, es el haplotipo más frecuente con diferencia (TABLA R-31).

El gen *MUC2*, que es el de mayor tamaño en este estudio, presentó un bajo grado de LD como en las poblaciones anteriores (HapMap y pacientes con lesiones gástricas o LPCG). Se observaron diversos puntos de recombinación a lo largo de todo el gen y numerosos bloques de LD (FIGURA R-28), de manera similar a la población CEU en la que se seleccionaron hasta 7 bloques diferentes. Los tags correspondientes a cada uno de los bloques seleccionados previamente, generalmente 2-3 por bloque (FIGURA R-18), presentan un fuerte LD entre ellos, en valores de D' , aunque un valor de r^2 muy bajos, ninguno superior a 0.8 y en muchos casos inferiores al 0.1. Se excluyeron del análisis dos de los 14 tagSNPs inicialmente seleccionados por problemas de genotipado, el rs10794293 y el rs40777592. El rs10794293, en el intrón 37, es uno de los marcadores del bloque que engloba la región génica adyacente al exón central, a 3', y el rs40777592 es el tagSNP seleccionado situado más a 3' de toda la región y uno de los marcadores del bloque que abarca el extremo 3' del gen y región adyacente. Si se tienen en cuenta los 4 últimos bloques de LD, que engloban toda la región 3' del gen definida por el exón central (FIGURA R-18), sólo se pierde la información de estos dos SNPs, con una reducción de la cobertura de la variabilidad del 15% en esta zona. Siguiendo la pauta que se llevó a cabo en la población de pacientes con lesiones gástricas, y que obedece al poco LD existente en *MUC2*, el análisis de haplotipos y posterior análisis de asociación se realizó por separado en las regiones 5' y 3' del gen. En la región 5' se observaron 9 haplotipos, conformados por los 5 SNPs analizados en esa zona, con unas frecuencias de 0.014 hasta 0.313. Con los otros 7 SNPs genotipados en la región 3' se infirieron un total de 13 haplotipos con frecuencias comprendidas entre 0.010 y 0.292 (TABLAS R-28).

Finalmente, el gen *MUC6* es el gen del estudio que presenta claramente el mayor LD (FIGURA R-28). En este caso, además, todos los tagSNPs analizados en la región presentan un fuerte LD, en valores D' , observándose un único bloque que, a diferencia de la población de pacientes con lesiones gástricas (FIGURA R-26), no se rompe por el SNP rs11246384 que tiene una relativa baja frecuencia (MAF=0.08). El bloque de LD engloba tanto a *MUC6* (~24 kb) como la región genómica a 5' del gen (~5 kb). Uno de los seis tagSNPs del único bloque seleccionado inicialmente para *MUC6*, rs4076950 en el intrón 30 cerca del extremo 3', se excluyó del estudio, resultando únicamente en un 5% menos de cobertura de la variabilidad genética de toda la región, ya que este SNP se marcaba únicamente a sí mismo. Los 5 SNPs finalmente analizados configuraron un total de 6 haplotipos con frecuencias de entre 0.082 y 0.428 (TABLAS R-31).

En cuanto al coeficiente de r^2 , en ningún caso se encontraron valores superiores a 0.8 y en muchas ocasiones, especialmente entre SNPs de *MUC2*, no alcanzaron ni un valor de 0.1. En general, la estructura y los valores de LD (D' y r^2) para cada uno de los genes observados en el grupo de pacientes con lesiones gástricas está en concordancia con los obtenidos a partir de los datos de HapMap tal y como se ejemplifica en el caso de *MUC2* (FIGURA A-4 del apéndice III).

2.2.4. Asociación polimorfismos con cáncer gástrico

El análisis de asociación de polimorfismos en genes *MUC* (n=21) y *TFF* (n=21) con cáncer gástrico, y los subtipos anatómicos (cardias y no cardias) e histológico (intestinal y difuso), se realizó mediante regresión logística no condicional incluyéndose como variables de ajuste la edad, el sexo y el país de procedencia, que son las utilizadas en el apareamiento de los casos con las muestras control. Las diferencias entre los subtipos anatómicos e histológicos se midieron mediante un test de heterogeneidad. El análisis de asociación se estratificó por la infección de *H.pylori* en los casos del no cardias, presencia de la cepa de virulencia CagA y sexo para determinar la posible interacción de los genes analizados con estas tres variables. En las TABLAS A-9 y A-10 del apéndice II se indica la significación determinada por el valor p de todos los resultados obtenidos en el análisis de asociación con cáncer gástrico, así como por subtipos anatómico e histológico.

2.2.4.1 TFFs

Se analizó la asociación de un total de 21 polimorfismos en genes *TFF* (8 en *TFF1*, 6 en *TFF2* y 7 en *TFF3* con cáncer gástrico y sus subtipos (TABLAS A-10 y A-11 del apéndice II). En este grupo de genes se incluyeron SNPs potencialmente funcionales con bajas frecuencias alélicas que, en algunos casos, dificultaron la realización e interpretación de los análisis de asociación.

En cuanto a *TFF1*, dos SNPs, el rs225356 en el intrón 1 y el rs424694 en la región genómica a 5' del gen, se asociaron a un menor riesgo de cáncer gástrico en el modelo codominante y sobredominante (TABLA R-22). El rs225356 es tagSNP de un bloque de LD que abarca parte del gen *TFF1* (la mitad 5') y el rs424694 también es tagSNP pero, en este caso, del bloque de LD que engloba la región genómica a 5' del gen (FIGURA R-27 y TABLA R-2 bloques 4 y 6). Aunque estos dos polimorfismos presentan el mayor valor de r^2 (0.59) observado en *TFF1*, indicando cierto grado de desequilibrio de ligamiento, el valor de D' (0.78) no es tan elevado como el que presentan con SNPs adyacentes de su propio bloque de LD.

En el análisis estratificado en función del sexo, el rs225356 mantuvo una asociación significativa, como factor protector, en las mujeres en los modelos codominante y sobredominante sin observarse una interacción significativa con esta variable (TABLA A-12). Además, el rs225356 se asoció a un mayor riesgo de cáncer gástrico en el modelo recesivo en mujeres, por lo que se deduce que el genotipo protector de este último SNP es el heterocigoto GA y, sin embargo, la presencia de las dos copias del alelo A conferiría un mayor riesgo de cáncer gástrico en este grupo. El genotipo AA ya muestra una tendencia como factor de riesgo, aunque no significativa, globalmente en cáncer gástrico. Otro de los tagSNP de *TFF1*, el rs13047838, que marca el bloque de LD en la región genómica a 5' del gen, juntamente con el rs424694, se asoció como factor de riesgo de cáncer gástrico en mujeres (modelo recesivo, TABLA A-12), mostrando una tendencia muy diferente a la de los hombres, aunque no llegó a ser significativa (p interacción=0.0874).

En referencia a los subtipos histológicos, los SNPs rs225356 y rs424694 se asociaron únicamente al tipo intestinal, aunque la tendencia observada en el tipo difuso fue similar,

especialmente para el rs424694 , y las diferencias entre ambos subtipos no fueron significativas (TABLA R-22). Los dos SNPs se asociaron con un menor riesgo del tipo intestinal en los modelos codominante, dominante y sobredominante y, una vez más, el rs225356 se asoció también a un mayor riesgo en el modelo recesivo. Otros tres SNPs de *TFF1* se asociaron significativamente como factores de riesgo sólo para cáncer gástrico intestinal: el rs13276, rs2839488, rs9976977, siendo el test de heterogeneidad significativo para los dos primeros. El rs13276, un SNP potencialmente funcional localizado en la región 3'UTR, se asoció a un mayor riesgo en el modelo dominante, sobredominante y log-aditivo (TABLA R-22). En este caso se trata de un SNP muy poco frecuente (MAF=0.06), con sólo 8 individuos homocigotos AA en la población (7 controles y un caso). Los otros dos SNPs asociados a un mayor riesgo, rs2839488 y rs9976977 (modelos codominante, recesivo y log-aditivo TABLA R-22), son tagSNPs, localizados en el intrón 1 y la región genómica a 5' del gen, respectivamente, que marcan a dos bloques diferentes (bloques 4 y 6 de la TABLA R-2, el rs2839488 sustituye al rs4920094), siguiendo un esquema similar a lo que observado con los SNPs protectores rs225356 y rs424694. Todos los SNPs asociados al tipo intestinal presentan valores D' superiores a 0.7 entre ellos, aunque valores r^2 inferiores a 0.6 en todos los casos y, exceptuando el rs13276 (c.*41G>A), son tagSNPs frecuentes, con MAFs entre 0.35 y 0.46.

Cuando se analizó el efecto de los SNPs de *TFF1* en cáncer gástrico intestinal en función del sexo, el rs13276 se asoció a un mayor riesgo únicamente en mujeres, sin ser la interacción significativa (TABLA A-13 del apéndice II). El rs2839488 también mantuvo su asociación como factor de riesgo únicamente en mujeres en el modelo recesivo. Además, su genotipo heterocigoto GC (modelo sobredominante) se asoció como factor protector del tipo intestinal en este mismo grupo. El modelo codominante, en este caso, refleja las dos tendencias significativas. El mismo genotipo heterocigoto mostró una tendencia inversa en hombres en el modelo sobredominante siendo la interacción significativa (TABLA A-13). Por otra parte, el tagSNP rs13047838 de la región genómica a 5' del gen, se asoció fuertemente al tipo intestinal como factor de riesgo, y no al difuso, únicamente en mujeres en los modelos codominante, recesivo y aditivo siendo la interacción significativa en los tres casos (TABLA A-13). Esta tendencia en mujeres ya se había observado previamente en cáncer gástrico a nivel global. Por otra parte, el rs424694 se asoció a un menor riesgo de cáncer gástrico intestinal sólo en los hombres (TABLA A-13) aunque, en este caso, no se halló interacción.

Aunque, en el análisis general, ninguna de las variantes de *TFF1* se asoció significativamente al cáncer gástrico de tipo difuso (TABLA A-10), cuando se estratificó por el sexo el rs13051704 , un tagSNP en el intrón 1, mostró una asociación significativa con un mayor riesgo para cáncer gástrico difuso en los hombres en el modelo aditivo y, simultáneamente, se asoció como factor protector, es decir, en sentido contrario, en mujeres, siendo la interacción significativa en prácticamente todos los modelos de herencia (TABLA A-14). Además, el rs424694, se asoció también como factor protector del tipo difuso pero, en este caso, únicamente en mujeres (TABLA A-14). En este caso, como ya ocurrió en el intestinal, la interacción no fue significativa.

Tabla R-22. SNPs en el gen *TFF1* que mostraron asociación con cáncer gástrico y/o con alguno de sus subtipos histológicos

SNP	Modelo	Genotipo	Cáncer gástrico				Intestinal				Difuso				P-Heterog
			Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	
rs13276	Cod	GG	1150/323	1.00			1150/105	1.00			1150/115	1.00			
		GA	126/40	1.14	(0.78-1.67)		126/20	1.89	(1.12-3.19)		126/12	0.92	(0.49-1.73)		
		AA	7/2	1.03	(0.21-5.02)	0.791	7/1	1.91	(0.23-15.89)	0.068	7/1	1.37	(0.16-11.44)	0.926	0.0988
	Dom	GG	1150/323	1.00			1150/105	1.00			1150/115	1.00			
		GA-AA	133/42	1.14	(0.78-1.65)	0.500	133/21	1.89	(1.13-3.16)	0.020	133/13	0.94	(0.51-1.73)	0.848	0.0315
	Rec	GG-GA	1276/363	1.00			1276/125	1.00			1276/127	1.00			
		AA	7/2	1.02	(0.21-4.94)	0.983	7/1	1.73	(0.21-14.39)	0.633	7/1	1.38	(0.17-11.54)	0.774	0.6778
	Sob	GG-AA	1157/325	1.00			1157/106	1.00			1157/116	1.00			
		GA	126-40	1.14	(0.78-1.67)	0.494	126/20	1.88	(1.11-3.17)	0.024	126/12	0.92	(0.49-1.72)	0.784	0.0359
	Adi	GG,GA,AA	1283/365	1.12	(0.79-1.58)	0.529	1283/126	1.77	(1.11-2.82)	0.024	1283/128	0.97	(0.55-1.70)	0.915	0.0380
rs225356	Cod	GG	306/99	1.00			306/41	1.00			306/36	1.00			
		GA	601/144	0.74	(0.55-0.99)		601/38	0.47	(0.29-0.75)		601/56	0.82	(0.52-1.28)		
		AA	217/79	1.11	(0.79-1.57)	0.020	217/31	1.04	(0.62-1.72)	0.000	217/24	1.03	(0.59-1.79)	0.567	0.2894
	Dom	GG	306/99	1.00			306/41	1.00			306/36	1.00			
		GA-AA	818/223	0.84	(0.64-1.10)	0.207	818/69	0.62	(0.41-0.94)	0.027	818/80	0.87	(0.57-1.33)	0.523	0.3432
	Rec	GG-GA	907/243	1.00			907/79	1.00			907/92	1.00			
		AA	217/79	1.35	(1.00-1.81)	0.052	217/31	1.60	(1.02-2.51)	0.044	217/24	1.17	(0.72-1.89)	0.537	0.4525
	Sob	GG-AA	523/178	1.00			523/72	1.00			523/60	1.00			
		GA	601/144	0.71	(0.55-0.91)	0.006	601/38	0.46	(0.31-0.70)	0.000	601/56	0.81	(0.55-1.19)	0.288	0.1153
	Adi	GG,GA,AA	1124/322	1.03	(0.86-1.23)	0.743	1124/110	0.96	(0.72-1.28)	0.799	1124/116	0.99	(0.74-1.32)	0.943	0.8653
rs2839488	Cod	GG	467/129	1.00			467/41	1.00			467/45	1.00			
		GC	603/172	1.03	(0.80-1.34)		603/51	0.97	(0.63-1.50)		603/61	1.02	(0.68-1.54)		
		CC	194/58	1.09	(0.76-1.55)	0.902	194/32	1.78	(1.13-3.08)	0.022	194/20	1.03	(0.59-1.81)	0.992	0.1280
	Dom	GG	467/129	1.00			467/41	1.00			467/45	1.00			
		GC-CC	797/230	1.05	(0.82-1.34)	0.723	797/83	1.19	(0.80-1.77)	0.385	797/81	1.02	(0.69-1.51)	0.907	0.6985
	Rec	GG-GC	1070/301	1.00			1070/92	1.00			1070/106	1.00			
		CC	194/58	1.06	(0.77-1.47)	0.702	194/92	1.89	(1.22-2.93)	0.005	194/20	1.02	(0.69-1.69)	0.950	0.0472
	Sob	GG-CC	661/187	1.00			661/73	1.00			661/65	1.00			
		GC	603/172	1.01	(0.80-1.27)	0.949	603/51	0.77	(0.53-1.13)	0.182	603/61	1.01	(0.70-1.47)	0.947	0.2015
	Adi	GG,GC,CC	1264/359	1.04	(0.88-1.23)	0.653	1264/124	1.33	(1.02-1.73)	0.033	1264/126	1.02	(0.77-1.33)	0.908	0.1766

SNP	Modelo	Genotipo	Cáncer gástrico				Intestinal				Difuso				P-Heterog
			Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	
rs9976977	Cod	GG	563/153	1.00			563/45	1.00			563/54	1.00			
		GA	573-161	1.03	(0.80-1.33)		573/55	1.19	(0.79-1.80)		573/55	1.01	(0.68-1.50)		
		AA	145/51	1.30	(0.90-1.88)	0.382	145/26	2.26	(1.34-3.81)	0.012	145/19	1.34	(0.76-2.36)	0.578	0.2824
	Dom	GG	563/153	1.00			536/45	1.00			563/54	1.00			
		GA-AA	718/212	1.09	(0.86-1.38)	0.487	718/81	1.40	(0.95-2.06)	0.081	718/74	1.07	(0.74-1.56)	0.711	0.1615
	Rec	GG-GA	1136/314	1.00			1136/100	1.00			1136/109	1.00			
		AA	145/51	1.28	(0.90-1.80)	0.173	145/26	2.06	(1.28-3.30)	0.004	145/19	1.33	(0.79-2.26)	0.296	0.2150
	Sob	GG-AA	708/204	1.00			708/71	1.00			708/73	1.00			
		GA	573/161	0.98	(0.77-1.23)	0.834	573/55	0.95	(0.65-1.38)	0.778	573/55	0.94	(0.65-1.36)	0.741	0.6730
	Adi	GG,GA,AA	1281/365	1.11	(0.93-1.32)	0.240	1281/126	1.45	(1.11-1.89)	0.006	1281/128	1.12	(0.85-1.46)	0.432	0.1118
rs424694	Cod	CC	377/125	1.00			377/48	1.00			377/46	1.00			
		CT	639/155	0.73	(0.55-0.95)		639/47	0.58	(0.38-0.88)		639/52	0.66	(0.43-1.01)		
		TT	267/85	0.95	(0.69-1.31)	0.041	267/31	0.92	(0.57-1.50)	0.024	267/30	0.89	(0.54-1.47)	0.144	0.9802
	Dom	CC	377/125	1.00			377/48	1.00			377/46	1.00			
		CT-TT	906/240	0.79	(0.62-1.02)	0.067	906/78	0.68	(0.46-0.99)	0.048	906/82	0.73	(0.50-1.08)	0.119	0.8960
	Rec	CC-CT	1016/280	1.00			1016/95	1.00			1016/98	1.00			
		TT	267/85	1.15	(0.87-1.52)	0.325	267/31	1.26	(0.82-1.95)	0.298	267/30	1.14	(0.73-1.76)	0.567	0.8469
	Sob	CC-TT	644/210	1.00			644/79	1.00			644/76	1.00			
		CT	639/155	0.74	(0.59-0.94)	0.012	639/47	0.60	(0.41-0.87)	0.006	639/52	0.70	(0.48-1.01)	0.055	0.9672
	Adi	CC,CT,TT	1283/365	0.81	(0.81-1.12)	0.537	1283/126	0.91	(0.70-1.19)	0.487	1283/128	0.91	(0.70-1.19)	0.495	0.8503

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

En cuanto a los subtipos anatómicos el rs424694, se asoció ligeramente a un menor riesgo de cáncer gástrico del cardias en el modelo sobredominante (CT, OR=0.65, 95%IC=0.43-0.98; $p=0.0358$), aunque las diferencias entre los dos subtipos (cardias y no cardias) no fueron significativas (p heterogeneidad= 0.3769). Cuando se analizó la asociación con los subtipos anatómicos en función del sexo, este genotipo sólo mantuvo el efecto protector en hombres (CT, OR=0.65, 95%IC=0.43-0.98; $p=0.0358$) sin observarse una interacción significativa (p interacción=0.6896). Esta tendencia en los hombres, no significativa, también se observó en el tipo intestinal como se ha comentado anteriormente. Por otro lado, el rs23839488 se asoció a un mayor riesgo de cáncer gástrico no cardias en hombres (modelos codominante, dominante y sobredominante, TABLA A-15), opuestamente a su asociación al tipo intestinal que se mantuvo únicamente en mujeres, y mostrando una interacción significativa con el sexo. Otro SNP de *TFF1*, el rs225356 asociado previamente a cáncer gástrico globalmente y al subtipo intestinal, se asoció también con un menor riesgo de cáncer gástrico no cardias pero, en este caso, únicamente en mujeres, siendo la interacción significativa (modelo codominante y sobredominante TABLA A-15). El genotipo AA, por el contrario, mostró una tendencia de riesgo en este grupo aunque no fue significativa.

En cuanto a *TFF2*, ninguno de los 7 polimorfismos analizados en el gen se asoció, en general, con cáncer gástrico o con alguno de sus subtipos. Sin embargo, el tagSNP rs3814896, en la región genómica a 5' de *TFF2*, se asoció significativamente con un mayor riesgo de cáncer gástrico globalmente en los hombres (TABLA A-12) y mostró una interacción significativa con el sexo. Esta misma asociación se repitió en el tipo difuso (modelos dominante y aditivo, TABLA A-14), donde se puede observar claramente la diferencia de la tendencia entre ambos sexos, incluso en aquellos modelos de herencia en que la asociación, por poco, no llega a ser significativa en hombres, como el codominante, y que resulta en una interacción significativa en prácticamente todos los modelos de herencia. En cuanto a los subtipos anatómicos, de nuevo este polimorfismo se asoció a un mayor riesgo de cáncer gástrico no cardias únicamente en hombres en todos los modelos excepto el recesivo (TABLA A-15), siendo la interacción significativa y mostrando un patrón muy similar en este subtipo al observado con el SNP de *TFF1* rs2839488 ($D'=0.72$, $r^2=0.21$). Aunque el polimorfismo potencialmente funcional c.*67_68insCTT mostró una ligera asociación con el cáncer gástrico localizado en el cardias en mujeres ($p=0.0460$), al tratarse de un polimorfismo muy poco frecuente (MAF=0.04), no se pudieron calcular los OR ya que la asociación se basa en la presencia de 51 heterocigotos en controles (n=511) y ningún heterocigoto en los casos (n=22).

En cuanto a *TFF3*, dos SNPs mostraron asociación con cáncer gástrico globalmente, el rs8133510, en la región genómica a 3' del gen, y el rs2236705, en el intrón 2 (TABLA R-23). Ambos polimorfismos son tagSNPs que se seleccionaron en el mismo bloque de LD (bloque 1, TABLA R-2) que englobaba todo el gen. En ambos casos, la asociación se observa con el genotipo heterocigoto; el genotipo GA del rs8133510 se asoció como factor protector (modelos codominante, dominante y sobredominante) mientras que el genotipo CA del rs2236705 se asoció como factor de riesgo (modelos codominante y sobredominante). Cuando se analizó la efecto de los SNPs de *TFF3* en función del sexo, estos dos mismos polimorfismos (rs8133510 y rs2236705) mantuvieron la asociación con cáncer gástrico únicamente en hombres, aunque no se observó interacción (TABLA A-13). A pesar de que uno de los SNPs potencialmente funcionales de *TFF3*, V116M, mostró una débil asociación con cáncer gástrico en mujeres ($p=0.0434$), su baja frecuencia (MAF=0.01) no permitió calcular los OR ya que no se encontró ningún homocigoto para el alelo menos frecuente en la población caso-control). En todo caso, la interacción con el sexo tampoco fue significativa (p interacción=0.105).

Referente a los subtipos histológicos, el rs2236705 se asoció únicamente a un mayor riesgo del tipo intestinal (modelos codominante y sobredominante) y específicamente en hombres (modelo sobredominante, AC, OR=1.66, 95%IC=1.03-2.69, $p=0.0423$), aunque no se observaron ni diferencias significativas entre los dos tipos histológicos (TABLA R-23) ni interacción con el sexo (modelo sobredominante, p interacción=0.7371). Sin embargo, dos de los tagSNPs que se localizan y marcan un bloque de LD en la región genómica a 5' de *TFF3*, rs225363 y rs692356, sí que se asociaron específicamente como factores protectores del tipo intestinal, mostrando diferencias significativas en comparación con el tipo difuso (modelos codominante, dominante y sobredominante, TABLA R-23). Por otro lado, el SNP de la región 3'UTR de *TFF3* c.*41G>A, mostró una asociación significativa con cáncer gástrico difuso (modelo codominante, GA, OR=9.25, 95%IC=2.41-35.53, $p=0.0037$). No hay presencia de homocigotos para el alelo menos frecuente de este SNP muy poco común en la población caso-control y el elevado OR observado se debe a que se basa en la presencia de 6 individuos heterocigotos en la población control y 4 heterocigotos del tipo difuso. Además el porcentaje de muestras genotipadas para este SNP fue el más bajo de todo el estudio caso-control (~77%) por lo que los resultados son poco concluyentes. Aunque en el análisis general, el rs8133510 no se asoció al tipo difuso, sí que lo hizo específicamente en hombres (TABLA A-14) pero sin ser la interacción significativa. Este SNP que se asoció previamente con cáncer gástrico, mostró también la misma tendencia en hombres globalmente, como se comentó anteriormente.

Tabla R-23. SNPs en el gen *TFF3* que mostraron asociación con cáncer gástrico y/o con sus subtipos histológicos.

SNP	Modelo	Genotipo	Cáncer gástrico				Intestinal				Difuso				P-Heterog
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs8133510	Cod	GG	811/249	1.00			811/88	1.00			811/86	1.00			
		GA	387/84	0.70	(0.53 - 0.92)		387/29	0.66	(0.43 - 1.03)		387/29	0.69	(0.45 - 1.08)		
		AA	45/17	1.23	(0.69 - 2.20)	.0216	45/6	1.18	(0.49 - 2.89)	.1442	45/6	1.13	(0.46 - 2.76)	.2259	.8379
	Dom	GG	811/249	1.00			811/88	1.00			811/86	1.00			
		GA-AA	432/101	0.76	(0.58 - 0.98)	.0328	432/35	0.72	(0.47 - 1.09)	.1093	432/35	0.74	(0.49 - 1.12)	.1539	.5524
	Rec	GG-GA	1198/333	1.00			1198/117	1.00			1198/115	1.00			
		AA	45/17	1.37	(0.77 - 2.43)	.2947	45/6	1.34	(0.55 - 3.23)	.5336	45/6	1.25	(0.52 - 3.04)	.6256	.8199
	Sob	GG-AA	856/266	1.00			856/94	1.00			856/92	1.00			
		GA	387/84	0.69	(0.53 - 0.91)	.0073	387/29	0.66	(0.42 - 1.02)	.0531	387/29	0.69	(0.44 - 1.07)	.0883	.6076
Adi	GG,GA,AA	1243/350	0.85	(0.68 - 1.06)	.1527	1243/123	0.82	(0.57 - 1.16)	.2537	1243/121	0.83	(0.59 - 1.19)	.3049	.5726	
rs2236705	Cod	CC	914/245	1.00			914/79	1.00			914/91	1.00			
		CA	326/114	1.30	(1.01 - 1.68)		326/45	1.57	(1.06 - 2.32)		326/34	1.07	(0.70 - 1.62)		
		AA	39/6	0.57	(0.24 - 1.36)	.0405	39/2	0.58	(0.14 - 2.47)	.0507	39/3	0.75	(0.23 - 2.51)	.8389	.2876
	Dom	CC	914/245	1.00			914/79	1.00			365/91	1.00			
		CA-AA	365/120	1.22	(0.95 - 1.57)	.1175	365/47	1.46	(1.00 - 2.15)	.0557	1240/37	1.03	(0.69 - 1.55)	.8813	.2687
	Rec	CC-CA	1240/359	1.00			1240/124	1.00			39/125	1.00			
		AA	39/6	0.53	(0.22 - 1.26)	.1198	39/2	0.51	(0.12 - 2.13)	.3053	1240/3	0.74	(0.22 - 2.46)	.6083	.3818
	Sob	CC-AA	953/251	1.00			953/81	1.00			953/94	1.00			
		CA	326/114	1.33	(1.03 - 1.71)	.0321	326/45	1.60	(1.08 - 2.35)	.0208	326/34	1.08	(0.71 - 1.63)	.7275	.1658
Adi	CC,CA,AA	1279/365	1.11	(0.89 - 1.38)	.3630	1279/126	1.26	(0.91 - 1.75)	.1752	1279/128	0.99	(0.70 - 1.41)	.9734	.4372	
rs225363	Cod	GG	421/122	1.00			421/53	1.00			421/40	1.00			
		GC	630/166	0.91	(0.69 - 1.18)		630/46	0.59	(0.39 - 0.89)		630/63	1.08	(0.71 - 1.65)		
		CC	224/73	1.12	(0.80 - 1.56)	.4052	224/26	0.92	(0.56 - 1.53)	.0296	224/22	1.09	(0.63 - 1.90)	.9244	.0564
	Dom	GG	421/122	1.00			421/53	1.00			421/40	1.00			
		GC-CC	854/239	0.96	(0.75 - 1.23)	.7547	854/72	0.68	(0.46 - 0.99)	.0438	854/85	1.08	(0.73 - 1.62)	.6929	.0247
	Rec	GG-GC	1051/288	1.00			1051/99	1.00			1051/103	1.00			
		CC	224/73	1.19	(0.88 - 1.60)	.2599	224/26	1.23	(0.78 - 1.95)	.3821	224/22	1.04	(0.64 - 1.70)	.8730	.9576
	Sob	GG-CC	645/195	1.00			645/79	1.00			645/62	1.00			
		GC	630/166	0.87	(0.69 - 1.10)	.2415	630/46	0.60	(0.41 - 0.88)	.0084	630/63	1.05	(0.72 - 1.52)	.8029	.0328
Adi	GG,GC,CC	1275/361	1.04	(0.88 - 1.23)	.6765	1275/125	0.89	(0.68 - 1.16)	.3716	1275/125	1.05	(0.80 - 1.37)	.7223	.1298	
rs692356	Cod	AA	422/126	1.00			422/54	1.00			422/42	1.00			
		AG	472/128	0.90	(0.68 - 1.19)		472/34	0.57	(0.36 - 0.90)		472/51	1.13	(0.73 - 1.74)		
		GG	222/76	1.13	(0.81 - 1.57)	.3942	222/28	0.99	(0.61 - 1.62)	.0274	222/25	1.18	(0.70 - 2.01)	.7896	.0319
	Dom	AA	422/126	1.00			422/54	1.00			422/42	1.00			
		AG-GG	694/204	0.97	(0.75 - 1.26)	.8434	694/62	0.71	(0.48 - 1.04)	.0807	694/76	1.14	(0.76 - 1.71)	.5101	.0188
	Rec	AA-AG	894/254	1.00			894/88	1.00			894/93	1.00			
		GG	222/76	1.19	(0.89 - 1.61)	.2475	222/28	1.28	(0.82 - 2.02)	.2885	222/25	1.11	(0.69 - 1.78)	.6633	.9939
	Sob	AA-GG	644/202	1.00			644/82	1.00			644/67	1.00			
		AG	472/128	0.86	(0.67 - 1.11)	.2480	472/34	0.57	(0.37 - 0.87)	.0073	472/51	1.06	(0.72 - 1.56)	.7718	.0173
Adi	AA,AG,GG	1116/330	1.04	(0.88 - 1.23)	.6123	1116/116	0.93	(0.71 - 1.20)	.5734	1116/118	1.09	(0.84 - 1.42)	.5048	.1392	

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

En cuanto a los subtipos anatómicos, el genotipo heterocigoto CA del rs2236705 se asoció como factor de riesgo al cáncer gástrico no cardias mostrando una diferencia significativa en comparación con el localizado en el cardias (TABLA R-24). Este efecto, en el análisis estratificado por sexos, sólo se observó en los hombres (TABLA A-15), aunque la interacción no fue significativa, tal y como había ocurrido anteriormente con su asociación con cáncer gástrico globalmente y con el subtipo intestinal en este mismo sexo. Por otra parte, los genotipos GG de los tagSNPs a 5' de *TFF3*, rs225363 y rs692356, se asociaron a un mayor riesgo de cáncer gástrico del cardias, siendo las diferencias entre los dos subtipos anatómicos significativas (modelo recesivo, TABLA R-24). Cuando se analizó la asociación de estos SNPs en función del sexo, ambos mantuvieron la asociación únicamente en los hombres (TABLA A-16), aunque la interacción no fue significativa, observándose la misma tendencia de riesgo en las mujeres. El genotipo heterocigoto GA del rs8133510, inicialmente asociado como factor protector de cáncer gástrico globalmente, se asoció con un menor riesgo del subtipo anatómico del cardias en hombres en el análisis estratificado por sexo (modelos codominante, dominante y sobredominante, TABLA A-16), siendo la interacción significativa. Finalmente, el SNP de cambio de aminoácido V80M mostró una tendencia como factor protector para el cáncer gástrico localizado en el cardias ($p=0.050$) pero su baja frecuencia no permitió calcular los OR. Las diferencias con el otro subtipo anatómico, no cardias, fueron significativas (p heterogeneidad=0.0412) aunque esta comparación se basa únicamente en dos casos heterocigotos no cardias y ninguno en el cardias, por lo que no se puede interpretar fiablemente esta asociación.

Referente a la estratificación por la infección por *H.pylori*, se realizó, en primer lugar, para los casos de cáncer gástrico no cardias, debido a que la infección por la bacteria es un factor de riesgo únicamente para este subtipo. Además, en este subtipo, se analizó la posible interacción de los polimorfismos de *TFFs* y las cepa de virulencia CagA.

Tres de los SNPs analizados en *TFF1*, rs2839488, rs13051704 y rs9976977, se asociaron con un mayor riesgo de cáncer gástrico del no cardias únicamente en los individuos no infectados por *H.pylori* (TABLA A-17). El tamaño de la muestra en los casos negativos para *H.pylori* es muy pequeño, y, por tanto, en algunos casos específicos los OR son muy elevados. Aun así se observa una clara diferencia en la tendencia de los dos grupos (infectados y no infectados) y se observa una interacción significativa en todos los casos (rs2839488 y rs9976977 modelo sobredominante, y rs13051704 modelo codominante, dominante y aditivo, TABLA A-17). El rs2839488, asociado a mayor riesgo de cáncer gástrico intestinal y el rs13051704, asociado previamente a cáncer gástrico difuso en hombres y mujeres pero en sentido opuesto, son dos SNPs intrónicos muy próximos marcadores del mismo bloque y con cierto LD entre ellos ($D'=0.99$, $r^2=0.56$). El SNP a 5' del gen asociado también a cáncer gástrico intestinal, rs9976977, marca un bloque diferente y presenta un menor LD con los SNPs anteriores (TABLA R-2 y FIGURA R-27).

Tabla R-24. SNPs en el gen *TFF3* que mostraron asociación con los subtipos anatómicos de cáncer gástrico.

SNP	Modelo	Genotipo	Cáncer gástrico				Cardias				No cardias ^a				P-Heterog
			Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%)	Valor p	
rs2236705	Cod	CC	914/245	1.00			914/76	1.00			914/117	1.00			
		CA	326/114	1.30	(1.01 - 1.68)		326/29	1.06	(0.67 - 1.66)		326/61	1.47	(1.05 - 2.06)		
		AA	39/6	0.57	(0.24 - 1.36)	.0405	39/2	0.62	(0.15 - 2.67)	.7575	39/3	0.56	(0.17 - 1.85)	.0417	.0916
	Dom	CC	914/245	1.00			914/76	1.00			914/117	1.00			
		CA-AA	365/120	1.22	(0.95 - 1.57)	.1175	365/31	1.01	(0.65 - 1.57)	.9600	365/64	1.36	(0.98 - 1.90)	.0699	.0336
	Rec	CC-CA	1240/359	1.00			1240/105	1.00			1240/178	1.00			
		AA	39/6	0.53	(0.22 - 1.26)	.1198	39/2	0.61	(0.14 - 2.62)	.4801	39/3	0.50	(0.15 - 1.64)	.2081	.9215
	Sob	CC-AA	953/251	1.00			953/78	1.00			953/120	1.00			
		CA	326/114	1.33	(1.03 - 1.71)	.0321	326/29	1.07	(0.68 - 1.69)	.7600	326/61	1.50	(1.07 - 2.09)	.0213	.0290
	Adi	CC,CA,AA	1279/365	1.11	(0.89 - 1.38)	.3630	1279/107	0.97	(0.66 - 1.42)	.8647	1279/181	1.19	(0.90 - 1.59)	.2339	.0541
rs225363	Cod	GG	421/ 122	1.00			421/34	1.00			421/63	1.00			
		CG	630/ 166	0.91	(0.69 - 1.18)		630/43	0.80	(0.50 - 1.29)		630/90	0.98	(0.69 - 1.39)		
		CC	224/ 73	1.12	(0.80 - 1.56)	.4052	224/29	1.45	(0.85 - 2.47)	.0775	224/26	0.80	(0.49 - 1.31)	.6375	.1171
	Dom	GG	421/ 122	1.00			421/34	1.00			421/63	1.00			
		GC-CC	854/ 239	0.96	(0.75 - 1.23)	.7547	854/72	0.98	(0.63 - 1.51)	.9125	854/116	0.93	(0.67 - 1.30)	.6882	.6443
	Rec	GG-GC	1051/ 288	1.00			1051/77	1.00			1051/153	1.00			
		CC	224/ 73	1.19	(0.88 - 1.60)	.2599	224/29	1.65	(1.04 - 2.62)	.0385	224/26	0.81	(0.52 - 1.26)	.3458	.0406
	Sob	GG-CC	645/ 195	1.00			645/63	1.00			645/89	1.00			
		GC	630/ 166	0.87	(0.69 - 1.10)	.2415	630/43	0.69	(0.46 - 1.03)	.0701	630/90	1.05	(0.77 - 1.44)	.7518	.2149
	Adi	GG,GC,C	1275/ 361	1.04	(0.88 - 1.23)	.6765	1275/106	1.18	(0.89 - 1.57)	.2565	1275/179	0.91	(0.73 - 1.15)	.4357	.1466
rs692356	Cod	AA	422/126	1.00			422/34	1.00			422/66	1.00			
		AG	472/128	0.90	(0.68 - 1.19)		472/31	0.76	(0.45 - 1.27)		472/70	0.98	(0.68 - 1.41)		
		GG	222/76	1.13	(0.81 - 1.57)	.3942	222/30	1.51	(0.89 - 2.57)	.0433	222/26	0.77	(0.47 - 1.25)	.5364	.0614
	Dom	AA	422/126	1.00			422/34	1.00			422/66	1.00			
		AG-GG	694/204	0.97	(0.75 - 1.26)	.8434	694/61	1.00	(0.64 - 1.57)	.9846	694/96	0.91	(0.65 - 1.28)	.5900	.6047
	Rec	AA-AG	894/254	1.00			894/65	1.00			894/136	1.00			
		GG	222/76	1.19	(0.89 - 1.61)	.2475	222/30	1.75	(1.09 - 2.79)	.0232	222/26	0.78	(0.50 - 1.22)	.2674	.0213
	Sob	AA-GG	644/202	1.00			644/64	1.00			644/92	1.00			
		AG	472/128	0.86	(0.67 - 1.11)	.2480	472/31	0.64	(0.40 - 1.00)	.0466	472/70	1.06	(0.76 - 1.49)	.7359	.1351
	Adi	AA,AG,GG	1116/330	1.04	(0.88 - 1.23)	.6123	1116/95	1.21	(0.91 - 1.61)	.1883	1116/162	0.90	(0.71 - 1.13)	.3453	.1088

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Otros dos SNPs en *TFF2*, rs3814896 (modelo aditivo) y rs225334 (modelo recesivo), y uno en *TFF3*, rs2236705 (modelos codominante, dominante y sobredominante), se asociaron como factores de riesgo de cáncer gástrico no cardias en los individuos no infectados por la bacteria (TABLA A-17). En el caso del rs3814896, asociado previamente a cáncer gástrico en hombres, no se halló interacción. Sin embargo, la interacción si fue significativa para el rs225334, un tagSNP potencialmente funcional en la región 3'UTR de *TFF2*, y para el rs2236705, asociado previamente al tipo intestinal y al no cardias. Ningún polimorfismo de *TFFs* se asoció con los casos no cardias en los sujetos infectados. Además, cuando se analizó el efecto de las variantes en casos del no cardias positivos para la infección de *H.pylori* en función de CagA, no se observó ninguna interacción con la cepa de virulencia.

Después de aplicar el test de permutaciones para comparaciones múltiples en cáncer gástrico y en los subtipos histológicos y anatómicos globalmente para cada gen, tan sólo *TFF1* se acercó a la significación estadística en el subtipo intestinal ($p= 0.0524$).

2.2.4.2 MUCs

En la tabla TABLAS A-11 y A-12 del apéndice II se resumen los valores p de la significación obtenida en el análisis de asociación de un total de 21 tagSNPs en genes *MUC* (4 en *MUC1*, 12 en *MUC2* y 5 en *MUC6*) con cáncer gástrico y sus subtipos. Además, como en el caso de *TFFs*, se realizaron análisis estratificados en función del sexo, infección por *H.pylori* y presencia de la cepa de virulencia CagA en los individuos infectados.

Referente a *MUC1*, tres de los cuatro SNPs analizados en este gen se asociaron significativamente con cáncer gástrico (TABLA R-25). Los SNPs rs3814316 y rs9426886, en la región genómica a 3' de *MUC1* y muy próximos al gen, se asociaron a un mayor riesgo de cáncer gástrico (modelos codominante, recesivo y aditivo para el rs3814316 y modelo recesivo para el rs9426886). Por otra parte, el rs4072037, un SNP sinónimo localizado en el exón 2 de *MUC1*, se asoció como factor protector de cáncer gástrico en los modelos codominante y dominante. Aunque el rs3814316 se asoció únicamente en mujeres (OR=1.69, $p=0.0483$), la tendencia en hombres fue también muy similar y casi significativa (OR=1.47, $p=0.0930$). En el caso del rs4072037, la asociación se mantuvo sólo en hombres, sin embargo, tampoco se halló interacción (TABLA A-18). Cuando se analizó la asociación de estos SNPs con el subtipo histológico, los tres mantuvieron la asociación en el tipo intestinal, siendo el test de heterogeneidad significativo para el rs9426886 (modelo recesivo, TABLA R-25). Además, en la estratificación por sexo del tipo intestinal, los tres SNPs se asociaron en los hombres, aunque la interacción no fue significativa (TABLA A-19). Referente a los subtipos anatómicos, los tres SNPs anteriores mostraron una tendencia similar tanto en el cardias como en el no cardias, aunque para el rs3814316 y rs4072037 fue significativa únicamente en el cardias. Aun así no se observan diferencias significativas entre los dos subtipos en ninguno de los casos (test de heterogeneidad, TABLA R-26). El único SNP de *MUC1* que no se asoció a cáncer gástrico ni a ninguno de sus subtipos fue el rs1045253, localizado a unas 40 kb a 5' del gen.

Tabla R-25. SNPs en *MUC1* que mostraron asociación significativa con cáncer gástrico y sus subtipos histológicos.

SNP	Modelo	Genotipo	Cáncer gástrico				Intestinal				Difuso				P Heterog
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs3814316	Cod	CC	566/143	1.00			566/46	1.00			566/54	1.00			
		CT	573/162	1.13	(0.87-1.46)		573/57	1.28	(0.85-1.94)		573/54	1.03	(0.69-1.53)		
		TT	142/59	1.66	(1.16-2.38)	.0246	142/23	2.17	(1.26-3.75)	.0261	142/19	1.40	(0.79-2.48)	.5040	.3513
	Dom	CC	566/143	1.00			566/46	1.00			566/54	1.00			
		CT-TT	715/221	1.23	(0.97-1.57)	.0894	715/80	1.45	(0.98-2.13)	.0583	715/73	1.10	(0.75-1.61)	.6191	.3711
	Rec	CC-CT	1139/305	1.00			1139/103	1.00			1139/108	1.00			
		TT	142/59	1.56	(1.12-2.17)	.0104	142/23	1.90	(1.16-3.11)	.0150	142/19	1.38	(0.81-2.35)	.2443	.1659
Adi	CC,CT,TT	1281/364	1.25	(1.05-1.48)	.0120	1281/126	1.44	(1.10-1.88)	.0092	1281/127	1.14	(0.87-1.50)	.3484	.1834	
rs9426886	Cod	AA	440/117	1.00			440/41	1.00			440/41	1.00			
		AT	631/172	1.03	(0.79-1.34)		631/55	0.95	(0.62-1.46)		631/66	1.12	(0.74-1.70)		
		TT	199/75	1.41	(1.01-1.98)	.1026	199/30	1.72	(1.04-2.87)	.0503	199/20	1.04	(0.59-1.85)	.8546	.0795
	Dom	AA	440/117	1.00			440/41	1.00			440/41	1.00			
		AT-TT	830/247	1.12	(0.87-1.44)	.3830	830/85	1.13	(0.76-1.68)	.5419	830/86	1.10	(0.74-1.64)	.6223	.8074
	Rec	AA-AT	1071/289	1.00			1071/96	1.00			1071/107	1.00			
		TT	199/75	1.39	(1.03-1.87)	.0336	199/30	1.77	(1.14-2.76)	.0149	199/20	0.97	(0.58-1.62)	.9155	.0444
Adi	AA,AT,TT	1270/364	1.16	(0.98-1.38)	.0775	1270/126	1.27	(0.98-1.67)	.0753	1270/127	1.04	(0.79-1.37)	.7758	.3407	
rs4072037	Cod	AA	337/118	1.00			337/43	1.00			337/37	1.00			
		AG	660/157	0.68	(0.51-0.89)		660/51	0.60	(0.39-0.92)		660/62	0.86	(0.56-1.32)		
		GG	276/86	0.98	(0.65-1.23)	.0141	276/32	0.90	(0.55-1.47)	.0419	276/27	0.86	(0.51-1.46)	.7718	.2998
	Dom	AA	337/118	1.00			337/43	1.00			337/37	1.00			
		AG-GG	936/243	0.74	(0.57-0.95)	.0211	936/83	0.68	(0.46-1.01)	.0620	936/89	0.86	(0.57-1.29)	.4719	.4521
	Rec	AA-AG	997/275	1.00			997/94	1.00			997/99	1.00			
		GG	276/86	1.14	(0.86-1.50)	.3733	276/32	1.23	(0.80-1.88)	.3512	276/27	0.95	(0.61-1.50)	.8326	.3402
Adi	AA,AG,GG	1273/361	0.92	(0.78-1.09)	.3382	1273/126	0.92	(0.70-1.19)	.5142	1273/126	0.92	(0.71-1.21)	.5573	.9436	

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla R-26. SNPs en *MUC1* que mostraron asociación significativa con los subtipos anatómicos de cáncer gástrico.

SNP	Modelo	Genotipo	Cáncer gástrico				Cardias				No cardias				P Heterog
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs3814316	Cod	CC	566/143	1.00			566/34	1.00			566/78	1.00			
		CT	573/162	1.13	(0.87 - 1.46)		573/53	1.41	(0.89 - 2.23)		573/77	1.03	(0.73 - 1.46)		
		TT	142/59	1.66	(1.16 - 2.38)	.0246	142/19	1.96	(1.06 - 3.61)	.0859	142/26	1.40	(0.86 - 2.29)	.4044	.5409
	Dom	CC	566/143	1.00			566/34	1.00			566/78	1.00			
		CT-TT	715/221	1.23	(0.97 - 1.57)	.0894	715/72	1.52	(0.98 - 2.35)	.0554	715/103	1.11	(0.80 - 1.52)	.5423	.2870
	Rec	CC-CT	1139/305	1.00			1139/87	1.00			1139/155	1.00			
		TT	142/59	1.56	(1.12 - 2.17)	.0104	142/19	1.60	(0.93 - 2.75)	.0992	142/26	1.38	(0.87 - 2.18)	.1827	.5151
	Adi	CC,CT,TT	1281/364	1.25	(1.05 - 1.48)	.0120	1281/106	1.40	(1.04 - 1.89)	.0267	1281/181	1.14	(0.90 - 1.44)	.2713	
	rs9426886	Cod	AA	440/117	1.00			440/30	1.00			440/62	1.00		
AT			631/172	1.03	(0.79 - 1.34)		631/53	1.25	(0.78 - 2.01)		631/84	0.95	(0.67 - 1.36)		
TT			199/75	1.41	(1.01 - 1.98)	.1026	199/24	1.61	(0.90 - 2.87)	.2724	199/34	1.26	(0.80 - 2.00)	.4543	.7889
Dom		AA	440/117	1.00			440/30	1.00			440/62	1.00			
		AT-TT	830/247	1.12	(0.87 - 1.44)	.3830	830/77	1.34	(0.86 - 2.10)	.1902	830/118	1.03	(0.73 - 1.43)	.8829	.4918
Rec		AA-AT	1071/289	1.00			1071/83	1.00			1071/146	1.00			
		TT	199/75	1.39	(1.03 - 1.87)	.0336	199/24	1.40	(0.86 - 2.29)	.1870	199/34	1.30	(0.86 - 1.96)	.2190	.8472
Adi		AA,AT,TT	1270/364	1.16	(0.98 - 1.38)	.0775	1270/107	1.27	(0.95 - 1.69)	.1071	1270/180	1.10	(0.87 - 1.38)	.4408	.5806
rs4072037		Cod	AA	337/118	1.00			337/40	1.00			337/53	1.00		
	AG		660/157	0.68	(0.51 - 0.89)		660/44	0.59	(0.37 - 0.93)		660/79	0.75	(0.51 - 1.09)		
	GG		276/86	0.98	(0.65 - 1.23)	.0141	276/22	0.75	(0.43 - 1.30)	.0764	276/46	1.01	(0.66 - 1.55)	.1837	.6666
	Dom	AA	337/118	1.00			337/40	1.00			337/53	1.00			
		AG-GG	936/243	0.74	(0.57 - 0.95)	.0211	936/66	0.63	(0.41 - 0.96)	.0357	936/125	0.82	(0.58 - 1.17)	.2812	.4142
	Rec	AA-AG	997/275	1.00			997/84	1.00			997/132	1.00			
		GG	276/86	1.14	(0.86 - 1.50)	.3733	276/22	1.02	(0.62 - 1.68)	.9383	276/46	1.22	(0.84 - 1.75)	.2986	.4928
	Adi	AA,AG,GG	1273/361	0.92	(0.78 - 1.09)	.3382	1273/106	0.82	(0.61 - 1.10)	.1756	1273/178	0.99	(0.79 - 1.25)	.9540	.3696

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Referente a *MUC2*, un único SNP se asoció a cáncer gástrico a nivel global. El rs2071174, localizado en la región genómica a 5' de *MUC2*, y que marca un bloque de haplotipos que contiene la región promotora y los primeros exones del gen (FIGURAS R-18 y R-28), se asoció a un mayor riesgo de cáncer gástrico (TABLA R-27). Aunque el efecto del rs2071174 se mantuvo, e incluso aumentó, únicamente en las mujeres en el análisis estratificado, la tendencia es muy similar a la de los hombres y la interacción no fue significativa (TABLA A-18). Este SNP no se asoció significativamente a ninguno de los subtipos, anatómicos o histológicos, de cáncer gástrico a nivel general. En cuanto a los subtipos histológicos intestinal y difuso, ninguno de los SNPs analizados en *MUC2* se asoció específicamente al tipo difuso, pero tres de ellos se asociaron al tipo intestinal. Los SNPs rs10902073 (modelo codominante) y rs7480563 (modelo aditivo) se asociaron a un mayor riesgo de tipo intestinal, pero las diferencias con el tipo difuso no fueron significativas (p heterogeneidad > 0.05 , TABLA R-27). Por otro lado, el rs11245936 se asoció como factor protector de cáncer gástrico intestinal en los modelos dominante y aditivo, (TABLA R-27). La relativa baja frecuencia de este SNP ($MAF=0.08$), resultó en la ausencia de homocigotos para el alelo menos frecuente entre los casos del tipo intestinal y no se pudo calcular los OR ni intervalos de confianza para este genotipo. En este caso sí que fue significativo el test de heterogeneidad y se observa una clara diferencia en la tendencia entre los subtipos histológicos a pesar de la baja frecuencia alélica del SNP. Aunque los SNPs rs10902073 y rs11245936 mantuvieron la asociación con el tipo intestinal en hombres, la interacción no fue significativa (TABLA A-19). Referente a los subtipos anatómicos, los SNPs rs2856082 y rs6421972 se asociaron significativamente al no cardias. El rs2856082, que juntamente con el SNP rs2071174 marca un bloque que engloba la región promotora y los primeros exones de *MUC2*, se asoció a un mayor riesgo de cáncer gástrico no cardias en el modelo dominante (TABLA R-28) pero, a pesar de las diferencias observadas con el cardias, el test de heterogeneidad no fue significativo. El rs6421972 se asoció también el modelo dominante pero en sentido opuesto, como factor protector del no cardias, y en este caso el test de heterogeneidad fue significativo (TABLA R-28). Este es un tagSNP de cambio de aminoácido (I2154T) localizado en el exón 34 que marca, juntamente con los SNPs rs7944723 y rs10902089, un bloque de LD que abarca parte del exón central y región génica adyacente a 3'. Estos SNPs, que se encuentran en relativo LD con el rs6421972 (FIGURA R-28), mostraron también una tendencia en el no cardias, y con el mismo modelo dominante, como factor protector (rs7944723, CC-CG, $OR=0.74$, $95\%IC=0.53-1.04$, $p=0.0755$) o de riesgo (rs10902089, AG-AA, $OR=1.41$, $95\%IC=0.96-2.06$, $p=0.0733$) no significativa. Aunque globalmente ningún SNP de *MUC2* se asoció con el cáncer gástrico del cardias, tres SNP se asociaron específicamente en hombres en este subtipo anatómico (TABLA A-20). El homocigoto para el alelo menos frecuente CC del rs2071174 se asoció con un menor riesgo de cáncer gástrico del cardias en hombres en los modelos codominante y recesivo, siendo la interacción con el sexo significativa. Además, los SNPs rs11245936 y rs7480563, asociados significativamente con el tipo intestinal, se asociaron en el mismo sentido con el cáncer gástrico de cardias en los hombres (TABLA A-20) siendo la interacción significativa para rs11245936, que ya había mostrado la misma tendencia de interacción, aunque no significativa, en hombres en el tipo intestinal.

Tabla R-27. SNPs en *MUC2* que mostraron asociación significativa con cáncer gástrico y sus subtipos histológicos.

SNP	Modelo	Genotipo	Controles/ casos	Cáncer gástrico			Controles/ casos	Intestinal			Controles/ casos	Difuso			P Heterog
				OR	(CI 95%) ^a	Valor p		OR	(CI 95%) ^a	Valor p		OR	(CI 95%) ^a	Valor p	
rs2071174	Cod	TT	559/139	1.00			559/45	1.00			559/52	1.00			
		CT	521/169	1.32	(1.02 - 1.71)		521/56	1.34	(0.89 - 2.03)		521/57	1.20	(0.81 - 1.80)		
		CC	151/33	0.89	(0.58 - 1.35)	.0403	151/16	1.32	(0.72 - 2.42)	.3417	151/12	0.91	(0.47 - 1.76)	.5497	.3681
	Dom	TT	559/139	1.00			559/45	1.00			559/52	1.00			
		TC-CC	672/202	1.23	(0.96 - 1.57)	.1029	672/72	1.34	(0.90 - 1.99)	.1430	672/69	1.14	(0.78 - 1.67)	.5060	.2109
	Rec	TT-TC	1080/308	1.00			1080/101	1.00			1080/109	1.00			
		TC	521/169	1.36	(1.07 - 1.73)	.0135	151/16	1.13	(0.65 - 1.98)	.6651	151/12	0.82	(0.44 - 1.55)	.5389	.3176
	Adi	TT,CT,CC	1231/341	1.05	(0.88 - 1.26)	.5674	1231/117	1.20	(0.91 - 1.57)	.2066	1231/121	1.03	(0.78 - 1.36)	.8390	.1574
rs10902073	Cod	CC	532/140	1.00			532/46	1.00			532/50	1.00			
		CA	578/186	1.22	(0.95 - 1.57)		578/70	1.40	(0.95 - 2.08)		578/60	1.08	(0.73 - 1.61)		
		AA	168/38	0.86	(0.58 - 1.28)	.1072	168/10	0.69	(0.34 - 1.40)	.0480	168/18	1.11	(0.63 - 1.97)	.8979	.3091
	Dom	CC	532/140	1.00			532/46	1.00			532/50	1.00			
		CA-AA	746/224	1.14	(0.90 - 1.45)	.2808	746/80	1.24	(0.85 - 1.82)	.2613	746/78	1.09	(0.75 - 1.59)	.6496	.9749
	Rec	CC-CA	1110/326	1.00			1110/116	1.00			1110/110	1.00			
		AA	168/38	0.77	(0.53 - 1.12)	.1601	168/10	0.57	(0.29 - 1.11)	.0756	168/18	1.07	(0.63 - 1.81)	.8118	.1444
	Adi	CC,CA,AA	1278/364	1.01	(0.85 - 1.20)	.9144	1278/126	1.00	(0.76 - 1.31)	.9903	1278/128	1.06	(0.81 - 1.39)	.6567	.5045
rs11245936	Cod	GG	1065/304	1.00			1065/110	1.00			1065/102	1.00			
		GA	190/51	0.93	(0.67 - 1.31)		190/12	0.58	(0.31 - 1.07)		190/21	1.18	(0.72 - 1.96)		
		AA	12/3	0.88	(0.25 - 3.15)	.9066	12/0	Na	Na	.0559	12/3	3.41	(0.91 - 12.74)	.2270	.0429
	Dom	GG	1065/304	1.00			1065/122	1.00			1065/102	1.00			
		GA-AA	202/54	0.93	(0.67 - 1.29)	.6642	202/12	0.54	(0.29 - 1.01)	.0379	202/24	1.29	(0.80 - 2.08)	.3063	.0780
	Rec	GG-GA	1255/355	1.00			1255/122	1.00			1255/123	1.00			
		AA	12/3	0.89	(0.25 - 3.19)	.8570	12/0	Na	Na	.1252	12/3	3.30	(0.89 - 12.31)	.1107	.0311
	Adi	GG,GA,AA	1267/358	0.93	(0.69 - 1.26)	.6581	1267/122	0.54	(0.30 - 0.97)	.0261	1267/126	1.36	(0.89 - 2.08)	.1723	.0362
rs7480563	Cod	GG	341/88	1.00			341/26	1.00			341/32	1.00			
		GA	617/180	1.13	(0.85 - 1.51)		617/56	1.21	(0.75 - 1.97)		617/64	1.02	(0.65 - 1.61)		
		AA	311/91	1.13	(0.81 - 1.57)	.6827	311/40	1.70	(1.01 - 2.87)	.1139	311/31	0.98	(0.58 - 1.65)	.9777	.5169
	Dom	GG	341/88	1.00			341/26	1.00			341/32	1.00			
		GA-AA	928/271	1.13	(0.86 - 1.48)	.3825	928/96	1.38	(0.87 - 2.17)	.1566	928/95	1.01	(0.66 - 1.54)	.9718	.7887
	Rec	GG-GA	958/268	1.00			958/82	1.00			958/96	1.00			
		AA	311/91	1.04	(0.79 - 1.36)	.7851	311/40	1.50	(1.00 - 2.24)	.0535	311/31	0.96	(0.62 - 1.48)	.8527	
	Adi	GG,GA,AA	1269/359	1.06	(0.90 - 1.25)	.4839	1269/122	1.31	(1.01 - 1.71)	.0404	1269/127	0.99	(0.76 - 1.28)	.9270	.4317

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla R-28. Asociación de los SNPs en *MUC2* que mostraron asociación significativa con los subtipos anatómicos de cáncer gástrico.

SNP	Modelo	Genotipo	Cáncer gástrico				Cardias				No cardias				P Heterog
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs2856082	Cod	CC	925/253	1.00			925/76	1.00			925/116	1.00			
		CG	241/76	1.16	(0.86 - 1.55)		241/17	0.83	(0.48 - 1.44)		241/45	1.53	(1.05 - 2.23)		
		GG	24/5	0.76	(0.29 - 2.01)	.5129	24/2	1.00	(0.23 - 4.46)	.7929	24/3	1.07	(0.31 - 3.63)	.0970	.3626
	Dom	CC	925/253	1.00			925/76	1.00			925/116	1.00			
		CG-GG	265/81	1.12	(0.84 - 1.49)	.4391	265/19	0.84	(0.50 - 1.43)	.5233	265/48	1.49	(1.03 - 2.15)	.0376	.2640
	Rec	CC-CG	1166/329	1.00			1166/93	1.00			1166/161	1.00			
		GG	24/5	0.73	(0.28 - 1.94)	.5213	24/2	1.04	(0.24 - 4.62)	.9560	24/3	0.96	(0.28 - 3.26)	.9484	.5639
	Adi	CC,CG,GG	1190/334	1.07	(0.83 - 1.38)	.6128	1190/95	0.88	(0.55 - 1.40)	.5854	1190/164	1.36	(0.98 - 1.87)	.0692	.4067
	rs6421972	Cod	GG	372/107	1.00			372/26	1.00			372/66	1.00		
GA			589/166	0.97	(0.74 - 1.28)		589/54	1.23	(0.75 - 2.03)		589/69	0.67	(0.46 - 0.97)		
AA			255/67	0.91	(0.64 - 1.28)	.8557	255/19	1.01	(0.54 - 1.88)	.6240	255/31	0.68	(0.43 - 1.07)	.0763	.1313
Dom		GG	372/107	1.00			372/26	1.00			372/66	1.00			
		GA-AA	844/233	0.95	(0.73 - 1.24)	.7140	844/73	1.17	(0.72 - 1.87)	.5239	844/100	0.67	(0.48 - 0.94)	.0233	.04705
Rec		GG-GA	961/273	1.00			961/80	1.00			961/135	1.00			
		AA	255/67	0.92	(0.68 - 1.25)	.6031	255/19	0.88	(0.52 - 1.49)	.6221	255/31	0.85	(0.56 - 1.29)	.4364	.7169
Adi		GG,GA,AA	1216/340	0.95	(0.80 - 1.13)	.5926	1216/99	1.02	(0.76 - 1.37)	.9019	1216/166	0.80	(0.63 - 1.00)	.0526	.1306

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Referente a *MUC6*, un único SNP se asoció a cáncer gástrico globalmente. El rs11246384, en el intrón 7 del gen, se asoció a un mayor riesgo de cáncer gástrico en el modelo recesivo (TABLA R-29) y mantuvo su efecto sólo en hombres en el análisis estratificado por sexo, sin ser la interacción significativa. Aunque la tendencia fue claramente diferente en mujeres se trata de un SNP no muy común (MAF=0.07) con pocos homocigotos para el alelo menos frecuente (TABLA R-29). Además el rs7481521, se asoció como factor de riesgo en mujeres en el modelo aditivo aunque tampoco se halló interacción. Este SNP localizado en el exón 16 (V619M), ya mostró una tendencia, no significativa, como factor de riesgo con cáncer gástrico globalmente (TABLA R-29). No se encontraron asociaciones significativas en relación a los subtipos anatómicos e histológicos y los SNPs de *MUC6*. Aun así el alelo G del rs7481521 y el alelo T del rs9794921 mostraron una tendencia no significativa como factores de riesgo para cáncer gástrico del no cardias en el modelo aditivo (OR=1.24, 95%IC=0.99-1.55, $p=0.0661$, p heterogeneidad=0.8000 y OR=1.27, 95%IC=0.97-1.65, $p=0.0820$, p heterogeneidad=0.01749, respectivamente).

Cuando se analizó el efecto de los SNPs en función de la infección por *H.pylori* en los casos del no cardias, los dos polimorfismos de *MUC2* que se asociaron ya anteriormente a este subtipo, rs2856082 y rs6421972, mantuvieron su asociación únicamente en los individuos infectados (TABLA A-21). La tendencia como factor protector del rs6421972 fue muy similar en los no infectados, aunque en este subgrupo la asociación no fue significativa y no hubo una interacción positiva. Por el contrario, el rs2856082, asociado como factor de riesgo en los infectados y a cáncer gástrico no cardias en general, mostró una tendencia claramente diferente en los casos *H.pylori* negativo resultando en una interacción significativa en prácticamente todos los modelos de herencia. No obstante, cabe destacar que el SNP es relativamente poco frecuente (MAF=0.12) y el tamaño de la muestra es muy pequeño en los casos no cardias, no infectados ($n=14$) con la presencia de un único heterocigoto CG y ningún homocigoto para el alelo menos frecuente GG. Por otra parte, el rs2071174 de *MUC2* se asoció al cáncer gástrico no cardias únicamente en los individuos no infectados en el modelo recesivo siendo la interacción significativa (TABLA A-21). Debido a la ausencia de homocigotos para el alelo menos frecuente CC en los casos no cardias no se pudieron calcular los OR ni intervalos de confianza, aunque se deduce del modelo que este genotipo podría conferir un menor riesgo. De hecho, si se realiza el análisis estratificado en cáncer gástrico globalmente, el rs2071174 se asocia también únicamente en los no infectados por *H.pylori* como factor protector en los modelos codominante (CT, OR=1.31, 95%IC=0.76-2.27, TT, OR=0.19, 95%IC=0.04-0.84, $p=0.0048$) y recesivo (TT, OR=0.17, 95%IC=0.04-0.71, $p=0.0018$), siendo la interacción muy significativa ($p=0.0183$ y $p=0.0047$, respectivamente). Aunque el rs11246384, un SNP poco común de *MUC6*, se asoció con un mayor riesgo de cáncer gástrico del cardias en los no infectados, la tendencia en los sujetos positivos para la bacteria fue la misma aunque no significativa y con un intervalo de confianza mucho menos amplio (TABLA A-21).

Tabla R-29. Asociación de los SNPs en *MUC6* que mostraron asociación significativa con cáncer gástrico en la estratificación por sexo.

SNP	Modelo	Genotipo	Cáncer gástrico			Hombres				Mujeres				P Interac.	
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a		Valor p
rs7481521	Cod	AA	408/100	1.00			244/64	1.00			164/36	1.00			
		AG	582/164	1.15	(0.87 - 1.52)		346/95	1.05	(0.73 - 1.50)		236/69	1.34	(0.85 - 2.12)		
		GG	240/79	1.35	(0.96 - 1.89)	.2204	138/42	1.15	(0.74 - 1.79)	.8296	102/37	1.71	(1.01 - 2.90)	.1322	.5067
	Dom	AA	408/100	1.00			244/64	1.00			164/36	1.00			
		AG-GG	822/243	1.21	(0.93 - 1.57)	.1538	484/137	1.08	(0.77 - 1.51)	.6635	338/106	1.45	(0.95 - 2.22)	.0815	.2713
	Rec	AA-AG	990/264	1.00			590/159	1.00			400/105	1.00			
		GG	240/79	1.24	(0.93 - 1.66)	.1537	138/42	1.12	(0.76 - 1.65)	.5798	102/37	1.42	(0.91 - 2.20)	.1230	.4470
Adi	AA,AG,GG	1230/343	1.16	(0.98 - 1.37)	.0822	728/201	1.07	(0.86 - 1.33)	.5512	502/142	1.31	(1.01 - 1.70)	.0448	.2537	
rs11246384	Cod	CC	1035/280	1.00			616/168	1.00			419/112	1.00			
		CT	178/54	1.12	(0.80 - 1.57)		100/26	0.94	(0.59 - 1.50)		78/28	1.38	(0.85 - 2.25)		
		TT	6/6	3.70	(1.18 - 11.64)	.0808	2/5	8.77	(1.67 - 46.12)	.0219	4/1	1.01	(0.11 - 9.22)	.4336	.1146
	Dom	CC	1035/280	1.00			616/168	1.00			419/112	1.00			
		CT-TT	184/60	1.21	(0.87 - 1.66)	.2614	102/31	1.09	(0.70 - 1.69)	.7008	82/29	1.37	(0.85 - 2.20)	.2073	.5428
	Rec	CC-CT	1213/334	1.00			716/194	1.00			497/140	1.00			
		TT	6/6	3.62	(1.15 - 11.38)	.0325	2/5	8.86	(1.69 - 46.53)	.0059	4/1	0.94	(0.10 - 8.59)	.9574	.0750
Adi	CC,CT,TT	1219/340	1.27	(0.94 - 1.70)	.1191	718/199	1.24	(0.84 - 1.84)	.2909	501/141	1.31	(0.84 - 2.05)	.2425	.9197	

a, Análisis de regresión logística no condicional ajustado edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

En el análisis estratificado por la presencia de la cepa de virulencia CagA en los casos del no cardias infectados por la bacteria, dos SNPs de *MUC6*, el rs7481521 y el rs9794921, se asociaron a un mayor riesgo de cáncer gástrico no cardias en los CagA positivos. La interacción en este caso no fue significativa y estos dos SNPs ya mostraron una tendencia similar, aunque no significativa, como factores de riesgo para el no cardias globalmente. Un caso similar se observó para otros tres SNPs de *MUC2* (rs7944723, rs6421972 y rs2856082, TABLA A-22) que ya mostraron una tendencia o se asociaron significativamente como factores protectores (rs7944723, rs6421972) o factores de riesgo (rs2856082) del no cardias en el análisis previo. En este último caso, el rs2856082 se asoció en el análisis estratificado tanto a los positivos como los negativos para CagA. El único SNP asociado únicamente a los CagA positivos y para el cual se halló interacción fue el rs7126405 de *MUC2* (TABLA A-22). Este SNP intrónico, en la región 3' del gen, se asoció como factor protector en el modelo dominante y no se había asociado previamente a cáncer gástrico ni ninguno de sus subtipos aunque mostró la misma tendencia para cáncer gástrico globalmente en el modelo aditivo (OR=0.77, 95%IC=0.58-1.04, $p=0.08551$). Finalmente, un único SNP, rs2071174 de *MUC2*, se asoció sólo a los sujetos no portadores de la cepa de virulencia. La tendencia observada es similar en los portadores de CagA pero las OR son mucho menos elevadas en este grupo y la interacción es significativa en todos los modelos de herencia (TABLA A-22). Tal y como se ha comentado anteriormente, este SNP se asoció a cáncer gástrico en general y la asociación era mucho más significativa en los individuos no infectados por la bacteria. Es importante destacar las diferencias en el tamaño de la muestra entre los grupos CagA positivo y CagA negativo, ya que este último grupo sólo representa el 7.4% de todos los casos del no cardias infectados por *H.pylori* y el tamaño de muestra es muy pequeño en este grupo.

Cuando se aplicó el test de permutaciones para comparaciones múltiples en cáncer gástrico y los subtipos histológicos para cada gen *MUC*, *MUC1* se asoció significativamente al tipo intestinal ($p=0.0454$) y se acercó a la significación estadística en cáncer gástrico globalmente ($p=0.0517$) y en el cardias ($p=0.0998$).

2.2.5. Asociación de haplotipos con cáncer gástrico

Se realizó el análisis de asociación de los haplotipos para los cada uno de los genes *MUC* y *TFF* (APARTADO 2.2.3 de resultados) con cáncer gástrico y los subtipos histológicos y anatómicos mediante regresión logística considerando un modelo aditivo.

En cuanto a los genes *TFF*, solamente un haplotipo muy poco común (frecuencia del 0.017) de *TFF1*, CTGGCGGG se asoció como factor protector de cáncer gástrico globalmente. En comparación con el haplotipo más común, CTGGGACG, el haplotipo protector contiene el alelo G, del rs225356, cuyo genotipo heterocigoto se asoció a un menor riesgo de cáncer gástrico (TABLA R-30). Cuando se realizó el análisis en función del subtipo histológico, tres haplotipos se asociaron con el cáncer gástrico intestinal (TABLA R-30). El segundo haplotipo más frecuente de *TFF1* con una frecuencia de 0.126, CCGGGGGG, se asoció a un menor riesgo del cáncer gástrico del tipo intestinal y nuevamente, este haplotipo contiene el alelo comun G del rs225356, cuya

variante A se había asociado como factor de riesgo en el modelo recesivo, y también del alelo común C del rs424694, cuyo genotipo heterocigoto se asoció a un menor riesgo de cáncer gástrico del tipo intestinal en el modelo codominante y sobredominante. Este haplotipo, como el de referencia, porta los alelos comunes de los SNPs cuyos alelos menos frecuentes se asociaron a un mayor riesgo del cáncer gástrico del tipo intestinal, en el modelo aditivo (rs9976977, rs2839488 y rs13276, TABLA R-19).

Tabla R-30. Haplotipos en los genes *TFF* y asociación con cáncer gástrico y sus subtipos histológicos.

Gen	Haplotipo ^a	Frec.	Gástrico			Intestinal			Difuso		
			OR	(IC 95%) ^d	Valor p	OR	(IC 95%) ^d	Valor p	OR	(IC 95%) ^d	Valor p
<i>TFF1</i> ^a	CTGGGACG	0.320	1.00			1.00			1.00		
	CCGGGGGG	0.126	0.91	(0.68-1.20)	.4846	0.52	(0.30-0.92)	.0236	1.17	(0.77-1.79)	.4609
	TCAGCGGG	0.091	0.92	(0.67-1.27)	.6101	1.08	(0.66-1.75)	.7687	0.98	(0.58-1.63)	.9241
	CTGGGAGG	0.079	1.06	(0.74-1.50)	.7637	1.17	(0.68-2.02)	.5622	1.36	(0.81-2.28)	.2440
	CCAGCGCG	0.060	1.35	(0.94-1.96)	.1059	1.60	(0.92-2.80)	.0966	1.34	(0.74-2.42)	.3332
	TCACCGCG	0.057	0.95	(0.64-1.35)	.7788	1.38	(0.80-2.36)	.2470	1.18	(0.66-2.09)	.5815
	CCAGCGGA	0.046	1.23	(0.83-1.83)	.3070	1.91	(1.10-3.31)	.0210	1.16	(0.60-2.23)	.6560
	TCAGCGCG	0.032	1.13	(0.68-1.87)	.6325	1.14	(0.52-2.51)	.7369	1.77	(0.90-3.49)	.0980
	CCGGGACG	0.029	0.86	(0.49-1.52)	.6127	0.93	(0.37-2.34)	.8717	0.75	(0.28-2.02)	.5687
	CCGGCGGG	0.022	1.38	(0.78-2.44)	.2635	2.35	(1.13-4.87)	.0217	0.73	(0.22-2.41)	.6068
	CTGCCGCG	0.021	0.83	(0.42-1.64)	.5962	1.10	(0.41-2.91)	.8504	0.71	(0.21-2.46)	.5942
	TCAGGACG	0.018	0.83	(0.41-1.67)	.6056	1.81	(0.80-4.10)	.1548	0.44	(0.10-1.92)	.2734
	CTGGCGGG	0.017	0.25	(0.07-0.95)	.0413	0.23	(0.03-1.76)	.1577	0.35	(0.05-2.30)	.2729
	CCAGCGGG	0.011	0.98	(0.36-2.65)	.9629	0.80	(0.12-5.20)	.8156	1.08	(0.22-5.21)	.9245
CCGCCGCG	0.011	0.80	(0.31-2.08)	.6524	Na	Na	Na	1.17	(0.34-3.97)	.8045	
<i>TFF2</i> ^b	AG1TA	0.388	1.00			1.00			1.00		
	GA1CA	0.182	1.00	(0.79-1.26)	.9909	1.11	(0.78-1.58)	.5699	1.05	(0.73-1.52)	.7942
	AA1TA	0.170	0.86	(0.65-1.13)	.2774	0.97	(0.63-1.49)	.8929	0.89	(0.58-1.38)	.6152
	AA1CA	0.104	0.84	(0.60-1.16)	.2932	0.60	(0.33-1.10)	.1015	1.03	(0.64-1.67)	.8956
	AG1CA	0.080	0.94	(0.65-1.36)	.7306	0.61	(0.33-1.23)	.1650	0.93	(0.51-1.72)	.8217
	GA1TA	0.030	0.90	(0.52-1.57)	.7238	1.18	(0.54-2.58)	.6726	1.16	(0.51-2.64)	.7162
	AG2TG	0.022	0.74	(0.36-1.54)	.4191	1.22	(0.44-3.42)	.6996	0.54	(0.15-1.98)	.3553
	AA2TG	0.022	1.29	(0.70-2.38)	.4087	1.14	(0.38-3.48)	.8137	1.58	(0.67-3.69)	.2958
<i>TFF3</i> ^c	AGCGCGG	0.403	1.00								
	AAGGCGG	0.212	0.93	(0.74-1.16)	.5070	1.07	(0.75-1.52)	.7157	0.97	(0.68-1.37)	.8536
	AAGCGGA	0.166	0.86	(0.68-1.11)	.2468	0.91	(0.61-1.36)	.6391	0.80	(0.54-1.19)	.2761
	AAGGAGG	0.156	1.07	(0.84-1.36)	.5854	1.32	(0.91-1.91)	.1393	0.96	(0.66-1.42)	.8540
	GAGGCGG	0.027	0.91	(0.53-1.56)	.7334	1.70	(0.84-3.43)	.1392	0.64	(0.25-1.63)	.3497
	AGCGCGA	0.015	0.61	(0.25-1.54)	.2992	0.75	(0.19-3.04)	.6923	0.83	(0.22-3.09)	.3497

a, Los haplotipos de *TFF1* incluyen los SNPs en el siguiente orden: rs13047838, rs424694, rs9976977, rs13051704, rs2839488, rs225356; rs225353, rs13276; b, Los haplotipos de *TFF2* incluyen los SNPs en el siguiente orden: rs3814896, rs1079380, c.*67_68insCTT (alelo sin inserción=1, InsCTT=2), rs225334 y rs12483635; c, Los haplotipos de *TFF3* incluyen los SNPs en el siguiente orden: rs8133510, c.*41G>A, rs2236705, V80M, rs225363, rs692356 y rs11203197; d, Análisis de regresión logística ajustado por sexo, edad y país de referencia.

Los resultados que mejor concuerdan con el análisis individual de SNPs son los obtenidos con los haplotipos CCAGCGGA y CCGGCGGG, que se asociaron a un mayor riesgo de cáncer gástrico intestinal (TABLA R-30). El haplotipo CCAGCGGA, en comparación con el de referencia, contiene los alelos (A, C y A del rs9976977, rs2839488 y rs13276, respectivamente) que se asociaron con un mayor riesgo de cáncer gástrico del tipo intestinal en el modelo aditivo del análisis individual de SNPs. Este mismo haplotipo mostró una tendencia de riesgo para el cáncer

gástrico del cardias (OR=1.74, 95%IC=0.95-3.18) aunque no llegó a ser significativa ($p=0.0727$, TABLA A-23 del apéndice II). El otro haplotipo asociado como factor de riesgo, CCGGCGGG, mantiene el alelo de riesgo 'C' de rs2839488, que es el que, además, lo diferencia del haplotipo protector CCGGGGGG. Como se comentó anteriormente en la asociación de haplotipos con la evolución de lesiones gástricas, muchas de las asociaciones con SNPs individuales y cáncer gástrico o sus subtipos se observaron en los modelos codominante o dominante.

Aunque se observaron asociaciones puntuales para SNPs de *TFF2* y *TFF3* ningún haplotipo de estos dos genes se asoció significativamente con el cáncer gástrico o sus subtipos, en el modelo aditivo (TABLAS R-30 y A-23). Un haplotipo de *TFF2*, AA2TG (TABLA A-23), mostró una tendencia como factor de riesgo para el cáncer gástrico no cardias (OR=2.05, 95%IC=0.96-4.34, $p=0.0622$). A pesar de la tendencia observada, el haplotipo AA2TG en comparación con el de referencia AG1TA, porta el alelo común del rs1079380, la inserción de 3pb del c.*67_68insCTT y el alelo menos frecuente del rs12483635, ninguno de ellos asociados en el análisis de SNP individual al cáncer gástrico del no cardias.

En cuanto los haplotipos de genes *MUC*, un haplotipo poco frecuente (0.022) de la región 3' de *MUC2*, GGAAGAG, se asoció globalmente a cáncer gástrico como factor protector (TABLA R-31). En comparación con el haplotipo de referencia CAGAAAG, este haplotipo incluye el alelo A del SNP rs6421972 que se asoció como factor protector del no cardias, así como el alelo común G del rs7480563, cuya variante A se asoció a un mayor riesgo del tipo intestinal. El haplotipo de *MUC1* TTAT mostró una tendencia como factor de riesgo para cáncer gástrico, aunque no llegó a ser significativa. En comparación con el haplotipo de referencia CAGC, porta los dos alelos T de los rs3814316y rs9426886 asociados con mayor riesgo de cáncer gástrico globalmente, y el alelo A del rs4072037 cuya variante menos frecuente G se asoció como factor protector. En cuanto a los subtipos anatómicos, un haplotipo similar de *MUC1*, TTAC se asoció significativamente a cáncer gástrico del cardias como factor de riesgo (TABLA R-31). Este haplotipo porta, de nuevo, los alelos de riesgo para los 3 SNPs de *MUC1* que se asociaron a cáncer gástrico y que dos de ellos, el rs3814316 y rs4072037, también lo hicieron significativamente en el cardias. Además, el haplotipo CCGCG de la región 5' de *MUC2* y el GCTGG de *MUC6* se asociaron a un mayor riesgo, en este caso, de cáncer gástrico del no cardias (TABLA R-31). El haplotipo de *MUC2*, en comparación con el de referencia, es portador del alelo G del rs2856082 asociado a un mayor riesgo del cáncer gástrico del no cardias en el análisis individual y el alelo C del rs2071174 que mostró una tendencia, aunque no significativa, con un mayor riesgo de este subtipo en el modelo dominante. El haplotipo de *MUC6* es portador del alelo G del rs7481521 también asociado a cáncer gástrico del no cardias. Un único haplotipo se asoció con los subtipos histológicos de cáncer gástrico (TABLA A-24). El haplotipo CTCTA de la región 5' de *MUC2* se asoció con un menor riesgo del tipo intestinal. Este haplotipo, en comparación al de referencia, ACCTG, porta el alelo común del rs10902073 cuya variante A se asoció a un mayor riesgo de cáncer intestinal y el alelo A del rs11245936, también asociado a este subtipo histológico, con un efecto protector.

Tabla R-31. Haplotipos en los genes *MUC* y asociación con cáncer gástrico y sus subtipos anatómicos.

Gen	Haplotipo	Frec.	Gástrico			Cardias			No cardias		
			OR	(IC 95%) ^d	Valor p	OR	(IC 95%) ^d	Valor p	OR	(IC 95%) ^d	Valor p
<i>MUC1</i> ^a	CAGC	0.463	1.00								
	TTAT	0.178	1.22	(0.97-1.53)	.0859	1.02	(0.67-1.56)	.9130	1.25	(0.92-1.69)	.1536
	TTAC	0.164	1.17	(0.92-1.49)	.1925	1.80	(1.24-2.61)	.0022	0.91	(0.64-1.28)	.5748
	CAAT	0.102	0.84	(0.62-1.14)	.2691	1.05	(0.64-1.73)	.8314	0.74	(0.48-1.13)	.1630
	CTAC	0.068	0.90	(0.62-1.29)	.5579	0.79	(0.38-1.62)	.5149	1.03	(0.65-1.63)	.8904
	CAAC	0.011	1.21	(0.51-2.88)	.6600	0.68	(0.10-4.79)	.7003	1.45	(0.46-4.58)	.5229
<i>MUC2 5'</i> ^b	ACCTG	0.313	1.00			1.00			1.00		
	CTCTG	0.307	1.01	(0.81-1.26)	.9316	0.94	(0.64-1.40)	.7726	1.04	(0.78-1.40)	.7794
	CTCCG	0.183	1.02	(0.80-1.31)	.8637	0.92	(0.60-1.42)	.7027	1.05	(0.75-1.47)	.7551
	CTGCG	0.042	0.70	(0.40-1.22)	.2079	0.80	(0.32-2.01)	.6354	0.64	(0.28-1.45)	.2857
	CCGCA	0.038	1.18	(0.78-1.80)	.4285	0.85	(0.39-1.85)	.6863	1.25	(0.71-2.21)	.4365
	CTCTA	0.035	0.65	(0.37-1.14)	.1339	0.42	(0.11-1.57)	.1977	0.85	(0.42-1.75)	.6649
	ACCCG	0.027	0.98	(0.49-1.95)	.9550	1.26	(0.45-3.55)	.6601	0.78	(0.29-2.10)	.6291
	CCGCG	0.026	1.25	(0.75-2.09)	.3826	0.70	(0.22-2.25)	.5492	2.12	(1.19-3.78)	.0103
	ACGCG	0.014	1.15	(0.43-3.13)	.7797	1.06	(0.13-8.48)	.9586	1.78	(0.63-5.04)	.2752
	<i>MUC2 3'</i> ^c	CAGAAAG	0.292	1.00			1.00			1.00	
CGAAGAA		0.168	0.98	(0.76-1.27)	.8804	0.87	(0.55-1.38)	.5523	0.85	(0.60-1.21)	.3686
GGAAAAG		0.154	0.93	(0.71-1.23)	.6265	1.18	(0.76-1.83)	.4623	0.75	(0.51-1.08)	.1229
CAGAGAA		0.078	0.90	(0.63-1.30)	.5804	1.25	(0.71-2.20)	.4325	0.70	(0.42-1.17)	.1747
CAGAGAG		0.072	1.13	(0.79-1.61)	.5154	1.00	(0.53-1.90)	.9972	1.14	(0.73-1.79)	.5681
CGGGGGG		0.067	0.81	(0.56-1.17)	.2574	0.81	(0.43-1.52)	.5044	0.88	(0.55-1.40)	.5808
CGAAGAG		0.048	0.83	(0.53-1.29)	.4022	1.05	(0.51-2.14)	.8974	0.61	(0.31-1.19)	.1445
GGAAGAG		0.022	0.29	(0.09-0.93)	.0379	Na	Na	Na	0.37	(0.09-1.51)	.1648
GGAAAGG		0.017	0.74	(0.37-1.50)	.4035	0.57	(0.14-2.38)	.4384	0.65	(0.23-1.82)	.4137
CAGAGGG		0.016	0.46	(0.19-1.10)	.0826	Na	Na	Na	0.43	(0.13-1.41)	.1647
CGGAGAG		0.014	1.89	(0.96-3.71)	.0641	2.01	(0.63-6.47)	.2414	1.80	(0.75-4.35)	.1905
GGAAGAA		0.014	1.17	(0.51-2.69)	.7080	0.39	(0.05-3.04)	.3701	1.38	(0.52-3.65)	.5172
GGAAAAA		0.010	0.73	(0.26-2.03)	.5482	1.03	(0.18-5.71)	.9770	0.67	(0.18-2.50)	.5499
<i>MUC6</i> ^d	ACCGG	0.428	1.00			1.00			1.00		
	GCCGG	0.146	1.12	(0.86-1.45)	.3920	1.34	(0.89-2.01)	.1615	1.04	(0.72-1.49)	0.8384
	ATCGG	0.132	1.04	(0.78-1.38)	.8006	0.79	(0.47-1.34)	.3871	1.11	(0.76-1.62)	.5890
	GCCGT	0.106	1.16	(0.87-1.55)	.3018	0.71	(0.41-1.23)	.2272	1.44	(0.99-2.09)	.0535
	GCCTT	0.105	1.18	(0.89-1.55)	.2556	0.82	(0.50-1.35)	.4406	1.41	(0.97-2.03)	.0694
	GCTGG	0.082	1.33	(0.98-1.82)	.0702	0.89	(0.49-1.63)	.7134	1.58	(1.05-2.19)	.0285

a, Los haplotipos de *MUC1* incluyen los SNPs en el siguiente orden: rs3814316, rs9426886, rs4072037, rs1045253, rs2839488, rs225356; rs225353, rs13276; b, Los haplotipos de la región 5' *MUC2* incluyen los SNPs en el siguiente orden: rs10902073, rs10794281, rs2856082, rs2071174 y rs11245936 ; c, Los haplotipos la región 3' *MUC2* incluyen los SNPs en el siguiente orden: rs7944723, rs10902089, rs6421972, rs11245954, rs7480563, rs7126405, rs3924453; d, Los haplotipos la región 3' *MUC6* incluyen los SNPs en el siguiente orden: 7481521, rs11604757, rs11246384, rs6597947 y rs9794921; e, Análisis de regresión logística ajustado por sexo, edad y país de referencia.

Discusión

El objetivo principal de esta tesis ha sido el estudio de la variabilidad genética de los genes *TFF* y *MUCy* su relación con la carcinogénesis gástrica. A continuación se discutirán los resultados más relevantes en relación a los objetivos propuestos. La discusión relativa a los resultados de asociación de los estudios de evolución de lesiones gástricas y caso-control de cáncer gástrico se realiza conjuntamente para cada uno de los genes analizados.

La identificación de polimorfismos potencialmente funcionales en los genes *TFF1*, *TFF2* y *TFF3* se realizó mediante una selección de polimorfismos a partir de las bases públicas y un rastreo sistemático en todos los exones y regiones adyacentes de cada uno de estos genes. En el momento de iniciar este trabajo, en el año 2006, la versión activa de dbSNP era la 126, y el proyecto HapMap se encontraba en su primera fase. La mayoría de SNPs descritos para *TFFs* eran intrónicos y muy pocos estaban validados. Además, para alguno de los genes la cantidad de SNPs reportados era realmente baja, en el caso de *TFF3* un total de siete. Aun así, dbSNP demostró ser, con diferencia, la base más completa en la recopilación de este tipo de variabilidad, en comparación con Ensembl o SNP500Cancer, ya que, además, tiene en cuenta las secuencias de referencia más actuales de los genes para localizar y referenciar los SNPs. De los 9 SNPs seleccionados, 3 de ellos estaban validados y se identificaron también en nuestra población control y de los 6 SNPs que no estaban validados, la mitad no se pudo validar en nuestra población control (eran monomórficos). Además, el análisis mediante SSCPS de las regiones funcionales de estos 3 genes nos permitió identificar 9 polimorfismos no descritos en dbSNP 126. Teniendo estos datos en cuenta, se puede concluir que fue imprescindible tanto la validación en nuestra población de los SNPs seleccionados de dbSNP, ya que el 50% de SNPs no validados no se identificaron en nuestra población, como el rastreo sistemático de las regiones de interés para la búsqueda de nuevas variantes no descritas, que constituyeron más del 50% (n=7) del total de SNPs potencialmente funcionales seleccionados para su análisis en este trabajo (n=13). La mayoría de los polimorfismos seleccionados presentaron frecuencias bajas (<0.1), especialmente los de cambio de aminoácido, ya que los 3 identificados de este tipo presentaron frecuencias menores a 0.05. La versión actual de dbSNP, la 138, ha aumentado considerablemente tanto el número de SNPs publicados como los validados. Actualmente cuenta con más de 40 millones de SNPs validados, mientras que la versión que utilizamos, la 126, sólo incluía de 5 millones de este tipo de SNPs. Además, actualmente integra la información de la tercera fase del HapMap (124) y del proyecto 1000 genomas (379), un catálogo muy detallado de la variación genética humana realizado a partir de la secuenciación del genoma de más de 1000 individuos. Los nueve polimorfismos identificados en esta tesis están ya descritos actualmente en dbSNP 138, la mayoría de ellos sin validar o con frecuencias inferiores a 0.1. Así, para *TFF3*, el gen para el cual había menos SNPs en dbSNP 126, en la actualidad hay descritas unas 150 variantes. No obstante, tan sólo 12 de ellas, 10 en intrones y 2 en regiones 3' o 5'UTR, están validadas con frecuencias superiores a 0.01 y el único SNP de cambio de aminoácido validado actualmente en dbSNP en todo el gen es el V80M (rs118095917, con MAF=0.004 según dbSNP 138), identificado en nuestra población mediante SSCPS antes de que fuera descrito. Situaciones similares se dan para los otros *TFFs*. En *TFF2*, de las aproximadamente 200 variantes descritas, tan sólo hay dos

polimorfismos en exones y regiones adyacentes validados y con frecuencia mayor a 0.01, que corresponden a los SNPs de la región 3'UTR c.*82C>T (rs225334), que ya estaba validado en la versión dbSNP 126, y a la inserción que identificamos mediante SSCPs c.*68_69insCTT (rs372488289). De los aproximadamente 200 SNPs actualmente descritos en *TFF1*, 3 SNPs están validados y con una frecuencia superior a 0.01 en regiones funcionales: dos en regiones 5'UTR y 3'UTR, ambos seleccionados por nosotros a partir de la versión dbSNP 126 (rs2156310 y rs13276), y un SNP sinónimo en el exón 2, I54I (rs61735845), con una frecuencia de 0.028 que no identificamos en nuestra población. De los 3 SNPs que no pudimos validar de dbSNP 126, todos ellos de cambio de aminoácido, dos de ellos continúan sin estarlo actualmente y uno sí que se ha validado pero con frecuencia de 0.005. Por tanto, a pesar del gran incremento de información disponible hoy en dbSNP, nuestra estrategia de validación y búsqueda de SNPs en *TFFs* con frecuencias superiores a 0.01, que es la que nos permitió el análisis mediante de SSCPs de unas 50 muestras, continua siendo válida ya que detectamos todas las variantes que actualmente están descritas con esa frecuencia, a excepción del cambio sinónimo de *TFF1* (I54I, rs61735845) y un SNP de la región 3'UTR de *TFF3* (c.*15C>T, rs192423444), ambas variantes con una frecuencia inferior a 0.1.

Aunque el rastreo mediante SSCPs permitió identificar la mayoría de variantes actualmente descritas en las regiones analizadas, en la actualidad existen otras estrategias de mucho mayor rendimiento para el rastreo y/o genotipado de variantes como son las basadas en las nuevas técnicas de secuenciación masiva (*Next Generation Sequencing*, NGS) . Por otro lado, mientras realizábamos esta trabajo se describió el HRM (*High Resolution Melting*) como otra técnica menos laboriosa y de mayor rendimiento que los SSCPs. Respecto a los SSCPs, esta técnica permite el análisis simultáneo de un mayor número de muestras y en un único paso, ya que el producto de PCR no requiere de un procesamiento posterior. Para comparar la sensibilidad de las dos técnicas, ambas se utilizaron en la búsqueda de polimorfismos potencialmente funcionales en *MUC1*. De los 10 polimorfismos seleccionados únicamente se pudo validar el SNP sinónimo c.93A>G (T31T, rs4072037), un polimorfismo muy frecuente (MAF=0.42) cuya funcionalidad se ha demostrado ampliamente en diversos estudios (166,167). De los 9 polimorfismos que no pudimos validar en nuestra población control, siete estaban descritos en dbSNP con una frecuencia de 0.01 o menor, uno no estaba validado y uno estaba reportado con una frecuencia de 0.024, por lo que es posible que nuestro análisis en aproximadamente 50 muestras, cuyo objetivo era capturar las variantes con frecuencia superior al 0.01, no pudiera identificar estas variantes tan poco frecuentes. No obstante, en la actual versión de dbSNP únicamente están descritos dos polimorfismos validados y con MAF superior a 0.01 en todo el gen, incluyendo regiones intrónicas, el rs4072037 y el rs12743084 en el intrón 2. El SNP rs4072037 se validó tanto por SSCPs como por HRM, pero estos resultados son insuficientes para poder realizar una comparativa exhaustiva de las dos técnicas. Sin embargo sí que se pudo evaluar el HRM como técnica de genotipado, demostrando la misma eficiencia que las otras técnicas utilizadas para detectar cambios de un único nucleótido y, además, permitiendo el genotipado de pequeñas inserciones/deleciones.

Debido a los criterios usados para su búsqueda, todas las variantes seleccionadas en esta fase previa cumplieron el requisito de encontrarse en regiones genómicas que pueden estar sujetas a una alteración funcional por causa de un cambio en la secuencia como puede ser una sustitución nucleotídica (FIGURA D-1) (380) y este posible efecto funcional se analizó de una manera más detallada para alguna de las variantes.

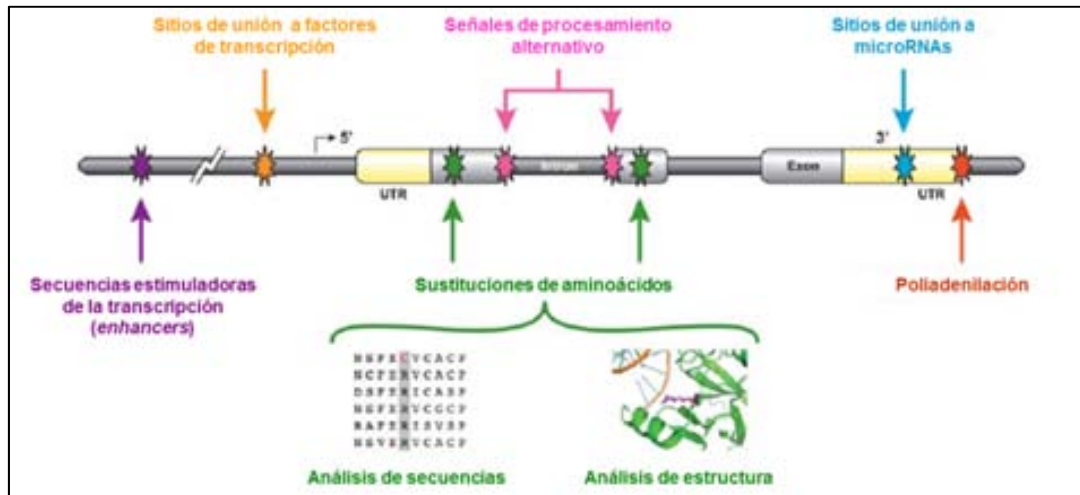


Figura D-1. Regiones genómicas que están sujetas a una alteración funcional por causa de una sustitución nucleotídica (modificado de Lee et al., 2009, (380)).

El hecho de encontrar tan pocos SNPs en regiones codificantes, y los identificados presentar unas frecuencias muy bajas, concuerda con la estructura y funcionalidad de estos péptidos. Los TFFs son péptidos muy pequeños (59-106 aminoácidos) y su estructura característica en forma de 3 bucles es extremadamente importante en poder mantener su integridad en un ambiente tan desfavorable como es el lumen del tacto gastrointestinal (172). Por tanto, cualquier cambio en la secuencia aminoacídica puede tener un gran impacto en la funcionalidad del péptido. Este punto se evidencia en el caso del SNP V80M, el único polimorfismo de cambio de aminoácido identificado en *TFF3* hasta ahora, que se encuentra en una posición altamente conservada y que nuestro análisis *in silico* identificó como seguramente perjudicial. Esta variante forma parte de la proteína madura, en el residuo 45 próximo a una de las cisteínas que forman los enlaces disulfuro, y podría alterar la correcta formación de los bucles o disminuir la resistencia del péptido a digestión proteolítica y degradación por ácidos. Además de los posibles efectos en la estructura y funcionalidad, en *TFF2* un cambio de aminoácido podría modificar también su patrón de glicosilación, ya que este péptido se encuentra mayoritariamente glicosilado *in vivo*, en su residuo 15 del péptido maduro (381,382). Sin embargo, no se encontraron efectos perjudiciales claros para los SNPs no sinónimos de *TFF2*. Uno de estos SNPs, Q7P, se encuentra en una zona muy poco conservada, forma parte del precursor del péptido y no está incluido en el péptido maduro. Para la variante de *TFF2*, V66L, no hubo concordancia entre las herramientas consultadas en predecir su posible efecto perjudicial. Este residuo se localiza en el primer dominio P de *TFF2*, aunque en una zona poco

conservada, y no es susceptible de ser glicosilado. Para *TFF1* no se encontraron variantes de cambio de aminoácido y la única que se encuentra actualmente descrita en las bases de datos con frecuencia superior al 0.01 es sinónima. Es de destacar que este péptido se encuentra generalmente formando homodímeros o heterodímeros con la proteína TFIZ1 (176) y que cualquier cambio en su estructura, además, podría alterar esta unión que es imprescindible para su funcionalidad.

No sólo las regiones codificantes son importantes para la correcta funcionalidad de un gen, se ha descrito que las regiones 3'UTR están involucradas en regular la expresión génica en múltiples niveles. Estas regiones pueden tener efectos en el pre-mRNA, están implicadas en la formación del extremo 3' del mRNA y la poliadenilación y en el mRNA maduro determinan algunas propiedades como la estabilidad, la degradación, la exportación nuclear, la localización subcelular y la eficiencia de traducción (383). El análisis funcional de las regiones 3'UTR de los genes *TFF* indicó que cambios en estas secuencias pueden afectar la expresión génica. El mayor efecto se observó para el SNP c.*41G>A de *TFF1* que disminuyó significativamente la expresión de la luciferasa en las dos líneas celulares analizadas (AGS y HEK-293T). Un efecto similar se observó para la variante que se encuentra en la misma posición de *TFF3*, descrita por nosotros en este trabajo. En el caso de *TFF2*, con dos polimorfismos en su región 3'UTR, el cambio de expresión fue opuesto y el mayor efecto se observa para el cambio c.*82C>T. Aunque todo indica que, efectivamente, estos cambios podrían tener consecuencias en la expresión del gen, hay que tener en cuenta que las regiones reguladoras del RNA habitualmente dependen de la combinación de la estructura primaria y secundaria y de la secuencia adyacente a estos elementos (384). Por esa razón, este posible efecto, observado mediante ensayos luciferasa, debería ser confirmado en ensayos que permitieran analizar la expresión diferencial del mRNA completo de *TFFs* en función de cada uno de los alelos de los SNPs de 3'UTR.

Los genes *TFF* están altamente conservados entre especies y, además, los tres miembros presentan muchas similitudes en su secuencia, especialmente en elementos promotores proximales y regiones codificantes, sugiriendo que podrían haber evolucionado de un ancestro común (177). Las regiones 3'UTR de los tres genes *TFF* también presentaron cierta similitud en nuestro análisis, con una identidad nucleotídica del ~50%, y la región de 3'UTR de *TFF1* parece estar altamente conservada entre especies de mamíferos. Diversos estudios comparativos han mostrado que algunas regiones 3'UTR son de las más conservadas en vertebrados y han identificado 106 elementos altamente conservados entre especies la mayoría de los cuales eran, además, dianas de microRNAs (385,386). En nuestro estudio identificamos numerosos microRNAs predichos a interactuar con las regiones 3'UTR de *TFFs*, aunque los diferentes programas utilizados dieron resultados muy dispares ya que utilizan distintos parámetros para su predicción en relación al apareamiento de las secuencias, la termodinámica de la interacción y la conservación entre secuencias (387). En el momento de la búsqueda, y hasta la actualidad, no se ha confirmado experimentalmente ninguna interacción de microRNAs y *TFFs*. Cuando la búsqueda se restringió a la secuencia donde se localizaban los polimorfismos de 3'UTR, tan sólo 3 programas (Target Scan, MicroInspector y MicroCosm, anteriormente miRBase) identificaron

microRNAs que tuvieran estas secuencias como potenciales dianas: miR-524-3p, miR-671, miR-615, miR-652 y miR-760. Aunque ninguna de las posibles dianas de miRNAs se encontraron en sitios conservados, se ha estimado que entre el 30%-50% de los sitios no conservados son funcionales cuando se expresan en el mismo tejido (388). En el momento de la selección de estos miRNAs existía muy poco conocimiento sobre su función o posible relación con patologías aunque recientemente se han hecho avances al respecto. Cuando se testaron los miRNAs predichos con sus correspondientes dianas, mediante ensayos luciferasa no se observaron, en general, diferencias significativas en la expresión. No obstante, estos son resultados de experimentos preliminares que requieren de la confirmación de la expresión de los miRNAs clonados en el vector pSIL. En el caso de *TFF2*, los miRNAs miR-524-3p y miR-671 parecen aumentar ligeramente la expresión aunque si hubiera una interacción real entre este tipo de reguladores y la región 3'UTR de *TFF2*, sería de esperar encontrar una disminución de la expresión de la luciferasa en este ensayo. No obstante, en la mayoría de los casos no se encontraron diferencias significativas, por lo que no se puede concluir que exista una expresión diferencial en función de la presencia de los microRNAs analizados. Aun así, no se puede descartar que estos microRNAs estén reprimiendo algún gen que regule negativamente a *TFF2* y que se exprese en células HEK-293T. Estas células son originarias de cultivos de riñón embrionarias y se asemejan a las células epiteliales de este órgano, aunque también pueden presentar características de otros tipos celulares debido a su transformación con adenovirus (389). Se ha descrito la expresión del mRNA de *TFF2* en riñón aunque no de proteína (390), por lo que podría haber algún tipo de regulación negativa en esta localización. Recientemente, se ha reportado que el miR-671 tiene como diana transcritos de sentido reverso (NAT, del inglés *natural antisense transcripts*) y que puede influir en la expresión del gen rompiendo este tipo de mRNA reguladores (391), que generalmente se asocian con el silenciamiento genético (392). En relación a la funcionalidad de estos miRNAs, recientemente se ha propuesto a miR-524-5p, que se origina a partir del mismo pre-miRNA que el -3p, como un candidato a supresor de tumores de glioma y se ha demostrado que miR-671 interacciona con la región 3'UTR del gen *CD44*, un gen asociado a la matriz extracelular, y está involucrado en migración, invasión y desprendimiento celular (393). En cuanto *TFF1*, no se observó interacción con el miR-652 pero si se observó interacción y expresión diferencial entre los dos alelos del SNP c.*41G>A en presencia del miR-760. Recientemente se ha descrito que la expresión del miR-760 en plasma desciende en cáncer colorrectal, donde *TFF1* se encuentra sobreexpresado (394), y podría actuar como marcador para el diagnóstico temprano de este tipo de cáncer (395). Además miR-760 está regulado por estrógenos y, junto a otros microRNAs, podría tener como diana numerosos transcritos pertenecientes a clústeres de genes que responden a estrógeno. (396). *TFF1* ha sido ampliamente estudiado como modelo de gen de respuesta a estrógeno en mama (189,397) aunque su papel biológico principal es en tejido gástrico. También se observó una posible interacción del miRNA-615 con la región 3'UTR que llevaba el alelo común del SNP c.*41G>A de *TFF3*, disminuyendo la expresión de la luciferasa, efecto que se perdería en el alelo A, aunque las diferencias no fueron significativas. En líneas celulares de cáncer de colon, localización donde se expresa normalmente *TFF3*, el miRNA-615

disminuye su expresión en respuesta a NGX6, un nuevo potencial gen supresor de este tipo de tumores (398). Las posibles interacciones entre los miRNAs y algunas de las regiones 3'UTR de *TFFs* deberían confirmarse en el tejido donde se expresan estos genes normalmente y cuantificar la expresión relativa tanto de los mRNAs de *TFFs* como de los miRNAs candidatos.

Por otra parte es cada vez más evidente la importancia de analizar no tanto el efecto individual de los distintos polimorfismos funcionales de un gen, como su efecto combinado en los haplotipos y bloques de desequilibrio de ligamiento en que se encuentran sus alelos (399). La selección de tagSNPs en estudios de asociación es muy útil porque permite reducir el número de SNPs analizados manteniendo la capacidad para detectar variantes asociadas. Esta selección es complementaria a la selección de polimorfismos potencialmente funcionales, ya que el polimorfismo causativo puede no haber sido identificado en el momento de la selección, o puede encontrarse en una región que no esté considerada como funcional. De esta manera, la selección de tagSNPs nos proporciona una cobertura más completa de la región de interés. El proyecto internacional HapMap (122,124,400) ha proporcionado un conocimiento global de la variabilidad en el genoma y de su estructura en bloques de desequilibrio de ligamiento que permite la selección de tagSNPs. Diversos estudios indicaron que la información sobre frecuencias alélicas y distribución de haplotipos en el genoma humano de población CEU, obtenida en el proyecto internacional HapMap, podría ser de gran utilidad en el estudio de genes asociados al cáncer en población española (401) y era transferible a población europea (402-404) aunque la densidad de SNPs, especialmente en zonas de bajo desequilibrio de ligamiento, y el sesgo existente hacia SNPs frecuentes tenía que tenerse en cuenta en el diseño de estudios de asociación.

El método escogido para realizar la selección de tagSNPs fue la comparación por pares (*pairwise*). Existen otras opciones alternativas que se basan en la selección de tagSNPs multimarcadores. Esta estrategia se basa en testar cada SNP con una combinación de 2-3 SNPs y su uso puede mejorar la eficiencia de la selección reduciendo el número SNPs seleccionados, aunque este algoritmo requiere de una mejor calidad de genotipado y puede resultar en una pérdida de potencia estadística. Además se debería analizar como este tipo de test multimarcador es transferible entre poblaciones, es decir, si los mismos haplotipos siempre predicen el mismo proxie (402,405). Otros estudios, sin embargo, no han observado una disminución de la potencia estadística en el uso de estos predictores multimarcador y recomiendan esta estrategia para evitar un genotipado adicional de SNPs (406). No obstante, la selección de tagSNPs en este estudio también se realizó en función de los haplotipos que marcaban, estableciendo como criterio frecuencias superiores al 1-5%, optimizando así la relación entre número de tagSNPs seleccionados y la cobertura de la variabilidad. Al rebajar el umbral de una correlación perfecta ($r^2=1$) a un menor nivel, como el que aplicamos ($r^2=0.8$), se disminuye sustancialmente el número de tagsSNPs requerido manteniendo un eficiencia casi completa (407). Además se fijó un límite de SNPs con MAFs superiores al 5% y se ha estimado que, con estos parámetros se podría capturar un 80% de los SNPs presentes, con esta frecuencia, en el genoma (408). El análisis del desequilibrio de ligamiento en los genes de protección de la mucosa gástrica a partir de los datos de genotipado de HapMap para población CEU, mostró que

el clúster de 55 kb en 21q22.3 donde se localizan los tres genes *TFF* es una zona de muy poco LD. Además, ninguno de estos genes presentó bloques de LD compactos que englobaran la totalidad del gen y región adyacente a pesar de ser unos genes muy pequeños. Se ha sugerido que estos genes puedan derivar de un ancestro común por duplicación (179), y este hecho concordaría con la elevada recombinación observada en la región. Inevitablemente, las zonas de poco LD conllevan un incremento considerable en la densidad de SNPs a seleccionar y genotipar para poder capturar el máximo de variabilidad genética existente. Este hecho se hace evidente al comparar en el clúster de *TFFs*, donde se seleccionaron 19 tagSNPs en 7 bloques diferentes con longitudes comprendidas entre 1 y 18 kb que abarcaban en total una región de aproximadamente 35 kb, con la región donde se localiza *MUC6*, en la que con únicamente seis tagSNP se consiguió capturar prácticamente el 90% de los haplotipos de un único bloque que abarcaba 47 kb. *MUC2*, a pesar de encontrarse en el mismo clúster que *MUC6* pero separado por 38.2 kb (290), presentó poco LD y diversos puntos de recombinación en la región de 47 kb donde se localiza, por lo que el número de tagSNPs seleccionados, un total de 15, fue relativamente elevado. Ambos genes están localizados en un cluster de 400 kb del extremo telómero en 11p15.5, una región que ya se ha descrito como rica en recombinación y con un elevado porcentaje de G y C que, además, contiene otros dos genes de mucinas, *MUC5AC* y *MUC5B* (280,289). Se considera que estos genes también comparten un ancestro común y han evolucionado por duplicación (291). La complejidad de esta región y los genes que contiene se hace evidente en el hecho de que actualmente HapMap no dispone de datos de genotipado para la región donde se localiza *MUC5AC*. La estructura de este gen es muy compleja y presenta cuatro dominios diferentes de repeticiones en tándem (TRs) de 24 pb que, además, están interrumpidos por las secuencias que codifican dominios ricos en cisteína. Estas secuencias repetitivas dificultaron en su momento la dilucidación de su estructura genómica y, aunque fue posible determinar la organización de su región central, no se pudo determinar con exactitud su secuencia (279). Es posible que estas mismas secuencias repetitivas dificulten el genotipado en la región y sean el motivo de la falta de información de SNPs en este gen. El gen *MUC1*, el más pequeño de los genes *MUC*, se localiza en una región de 1q21 que presenta un elevado LD, permitiendo una gran optimización en la selección de tagSNPs ya que en una región de aproximadamente 50 kb se seleccionaron únicamente 4 tagSNPs. El gran desequilibrio de ligamiento observado y el hecho de identificar pocas variantes frecuentes en las regiones funcionales de *MUC1* parecen indicar que este gen se encuentra en una región de poca recombinación y altamente conservada. Con la selección de tagSNPs realizada, se capturó en prácticamente todos los casos más de un 90% de los haplotipos de cada uno de los bloques identificados.

La mayor parte de los polimorfismos potencialmente funcionales que se validaron y/o identificaron previamente no habían sido analizados en el proyecto HapMap y, por tanto, se desconocía si podían estar incluidos en los bloques de desequilibrio de ligamiento identificados. La segunda fase del proyecto HapMap identificó gran parte de los SNPs con una MAF superior al 5%, pero las variantes raras o menos frecuentes están poco representadas (408). La inclusión de los polimorfismos potencialmente funcionales menos frecuentes en el mapa de haplotipos, que

posiblemente no estén representadas por los tagSNP seleccionados con MAFs superiores al 0.05, puede ser de gran utilidad para su futuro estudio y para una mayor comprensión de las asociaciones entre variantes de genes candidatos y enfermedad (409) y, por esa razón, se utilizó una estrategia combinada para la selección final. Actualmente, HapMap con la publicación de su tercera fase donde se han secuenciado algunas regiones del genoma para identificar e incluir también variantes raras (124), intenta dar mayor representación de estas variantes poco frecuentes.

Con la finalidad de mejorar la comprensión del papel de los genes *MUC* y *TFF* en la carcinogénesis gástrica se realizaron análisis de asociación entre los polimorfismos seleccionados como marcadores de su variabilidad genética y la evolución de lesiones gástricas precursoras de cáncer gástrico y el cáncer gástrico en si mismo.. Estos genes tienen un papel importante en el mantenimiento y protección de la mucosa gástrica (173,177), por lo que su posible asociación con la carcinogénesis gástrica podría observarse ya en las primeras fases de la progresión de las lesiones precursoras.

Es conocido que la potencia de un estudio para identificar asociaciones significativas depende de la prevalencia/incidencia de la enfermedad, de la frecuencia del polimorfismo, de los modelos de herencia, del efecto de los polimorfismos (*odds ratio*) y del tamaño de muestra (410), siendo este último el único parámetro modificable al diseñar el estudio, para aumentar la potencia del mismo. La potencia de los estudios realizados para identificar asociaciones significativas se ha estimado con el programa Quanto (411,412), para dar una mejor interpretación de los resultados.

En el caso de la población caso control, con 365 casos y 1284 controles (de 3 a 4 controles por caso), y considerando una potencia del estudio óptima en aproximadamente un 80%, es posible detectar ORs a partir de 1.3 para polimorfismos frecuentes (MAF=0.1-0.5), suponiendo un modelo aditivo. La potencia del estudio se reduce cuando se consideran polimorfismos con MAFs inferiores al 1%. En este caso, el estudio caso-control tendría potencia para detectar ORs a partir de 1.4-2.5 (para SNPs con MAFs de 0.1 a 0.01). Si se considerara un modelo recesivo la potencia del estudio se reduce considerablemente para aquellos SNPs poco frecuentes ya que sólo se podrían detectar ORs a partir de 3 (SNPs con MAFs de 0.1) y la potencia sería prácticamente nula para SNPs con MAFs inferiores a 0.01.

En la población de 453 pacientes con lesiones gástricas la variable respuesta se definió en función de la evolución de las lesiones a los casi 13 años de seguimiento, determinándose dos tipos de respuesta: aquellas lesiones que progresaron (28.5%), frente a las que se mantuvieron estables o revertieron, y aquellas que revertieron (29.3%), frente a las que se mantuvieron estables o progresaron. De esta manera, el estudio tiene una potencia óptima (~80%) para estimar efectos moderados, OR~2.0, de SNPs relativamente frecuentes (MAFs de 0.1 a 0.5) considerando un modelo aditivo. Si se considerara un modelo recesivo, disminuye considerablemente la potencia a detectar el mismo efecto para los SNPs frecuentes y es prácticamente nula para SNPs con frecuencias de 0.05 o menores. Se debe tener en cuenta, además, que cuando los análisis de asociación se estratifican en subgrupos, ya sea por variables

de riesgo o por subtipo de cáncer gástrico, en el caso de la población caso-control, el tamaño de muestra disminuye y, en concordancia, también la potencia estadística. Esta limitación en la potencia dificulta la interpretación de los resultados para estos SNPs poco comunes, generalmente potencialmente funcionales, pero sin homocigotos, o muy pocos, para el alelo menos frecuente, los cuales deberían replicarse o abordarse en estudios con una población de pacientes mucho más amplia, y distinta de la usada en este estudio, para su confirmación. Aunque la potencia estadística del estudio de evolución de lesiones gástricas es limitada hay que tener en cuenta que debido a las características de su diseño, con un largo seguimiento y toma de muestras biológicas, biopsias y extracción de sangre, puede considerarse que posee un tamaño muestral relativamente grande (45,366). Además, hasta la fecha no nos consta que ningún otro estudio haya evaluado la asociación entre variantes en genes de protección de la mucosa y la evolución de las lesiones gástricas. Y muy pocos estudios han abordado el papel que las variantes genéticas en genes candidatos puedan tener sobre el riesgo de lesiones gástricas precursoras de cáncer gástrico y tan sólo uno de ellos analiza los genes *MUC* y ninguno los *TFF* (365); en relación a la evolución de las lesiones solamente nos consta un estudio que analiza el efecto de variantes en genes de los mecanismos de reparación del DNA (413). Por tanto, a pesar de las limitaciones del estudio, su diseño es idóneo para poder estudiar y comprender la relación entre variabilidad genética y la evolución de las lesiones durante la carcinogénesis gástrica.

La infección por *H.pylori*, y su cepa de virulencia *cagA*, se asociaron con un mayor riesgo de cáncer gástrico en la población caso-control de cáncer gástrico, en concordancia con su papel ya establecido como uno de los factores de riesgo más importantes en la patogénesis de este tipo de cáncer. Además este efecto se observó únicamente, y en mayor medida, para el cáncer gástrico del no cardias. De hecho el 93% de casos de este tipo en el estaban infectados por la bacteria y se ha propuesto que la infección por *H.pylori* pueda ser una condición prácticamente necesaria para el cáncer gástrico del no cardias (30,414). El consumo de tabaco también se asoció a un mayor riesgo de cáncer gástrico en la población caso-control. Además se observó un mayor porcentaje significativo de fumadores y ex fumadores en el tipo intestinal y el cardias. El consumo de tabaco se ha asociado consistentemente con las dos localizaciones anatómicas del cáncer gástrico (108,415,416) y algunos estudios han observado un mayor efecto de este factor en el riesgo del cáncer gástrico del cardias y/o del tipo intestinal (107,417).

Sin embargo, no se encontró asociación con algunos de los principales factores de riesgo de cáncer gástrico (infección por *H.pylori*, consumo de tabaco y antecedentes familiares) y la evolución de las lesiones gástricas. A pesar de no haber encontrado asociación entre evolución de las lesiones e infección por *H.pylori*, un estudio exhaustivo de los factores de virulencia *cagA* (positivo/negativo) y *vacA* (s1/s2 y m1/m2) en la misma población de pacientes con lesiones gástricas infectados por la bacteria, demostró que la

infección con las cepas positivas para *cagA*, *vacA* s1 y *vacA* m1 estaba asociada a un mayor riesgo de progresión (366). Aunque el tabaco se ha asociado al cáncer gástrico, especialmente del cardias (107,108,418), su asociación con las lesiones gástricas precursoras, ha sido discrepante en función del tipo y localización de la lesión y la población estudiada (419-422) y en algunos casos se ha asociado a un mayor riesgo en combinación con la infección por *H.pylori* (423,424). Además, ninguno de los estudios realizados analiza el efecto del consumo de tabaco sobre la evolución de las lesiones. Por otro lado, el consumo de AINES se asoció significativamente con un menor riesgo de progresión en nuestra población. Estudios previos indican que los AINES, especialmente la aspirina, podrían ser factores protectores de cáncer gástrico así como de otros tipos de cáncer como el de colon (425-427). El consumo de estos fármacos se ha asociado con un menor riesgo de cáncer gástrico con un efecto dosis dependiente y se ha especulado con la posibilidad de que los AINES inhiban la replicación y proliferación de *H.pylori* (428,429). Por otra parte, estos medicamentos se han asociado a la aparición de úlceras pépticas, por lo que el uso a largo de plazo de AINES podría estar asociado con un mayor de riesgo de úlceras gastroduodenales y, consecuentemente, a un menor riesgo de cáncer gástrico (430). El posible efecto protector de los fármacos AINES se ha atribuido a la inhibición de COX-2, la enzima responsable de la síntesis de la prostaglandina, que incrementa su expresión en respuesta a la infección por *H.pylori*. Se ha reportado la expresión anormal de COX, en particular la isoforma COX-2, en diversas líneas de cáncer y se ha implicado en procesos de carcinogénesis, crecimiento tumoral, apoptosis y angiogénesis. Otros mecanismos adiciones de el posible efecto anticarcinogénico de los AINES podrían incluir la inducción de la apoptosis mediante vías independientes de COX, la inhibición de NF- κ B y la estimulación de genes supresores de tumores (431-433). El efecto protector de los AINES se haría relevante ya en las etapas iniciales de la carcinogénesis gástrica que cursan con inflamación debido a la infección por *H.pylori* y, por tanto, podrían ser determinantes en la progresión de las lesiones precursoras de cáncer gástrico como hemos observado en nuestro estudio.

La asociación con las diferentes variables genéticas se realizó considerando diversos modelos de herencia. No es fácil establecer un criterio para determinar el modelo de herencia más adecuado para un polimorfismo concreto. Habitualmente, se suelen considerar los modelo codominante o aditivo ya que son los más generales, aunque los modelos dominantes y recesivos suelen ser complementarios. El enfoque más adecuado sería obtener una validación experimental de los distintos modelos asociados y que fueran replicados por otros estudios epidemiológicos. El criterio de información Akaike (AIC), que es una medida de la calidad relativa de un modelo, proporciona un medio para su selección. Aunque el modelo sobredominante suele ser difícil de explicar y en muchos estudios no se considera, en el caso de los genes *TFF* podría ser plausible debido a la formación de homodímeros por parte estos péptidos. Los polimorfismos asociados con la evolución de lesiones en el modelo sobredominante, aunque estén localizados en intrones o

zonas intergénicas, podrían estar marcando otros SNPs, de cambio de aminoácido, que afectaran la correcta homodimerización de estos péptidos. La diversidad de modelos de herencia, aunque aporta información adicional al estudio de asociación, en algunos casos también dificulta su interpretación y su relación con los haplotipos, que se analizan considerando un modelo aditivo.

El resultado más relevante en relación a la evolución de lesiones gástricas, ha sido su asociación con la variabilidad genética de *MUC2*, cuyo efecto es mucho más evidente en los pacientes infectados por *H.pylori* y aquellos que no son consumidores de AINES. Hasta ahora un único estudio había abordado la asociación entre SNPs candidatos de *MUC1*, *MUC2* y *MUC6* y el riesgo de gastritis crónica atrófica, una lesión gástrica específica, pero no identificó ninguna asociación (365). *MUC2* es el principal constituyente de la barrera de mucus secretado en el intestino delgado y grueso, donde se expresa principalmente en células caliciformes (434,435). Existen numerosas evidencias que confirman el papel de *MUC2* en el mantenimiento y la integridad de la capa mucosa protectora del intestino y se ha relacionado una disminución de su expresión con inflamación y diversas enfermedades intestinales, incluyendo cáncer (274,328,350,435). Por el contrario, *MUC2*, que normalmente no se expresa en el estómago, aumenta su expresión en metaplasia intestinal y cáncer gástrico (42,346,349,436). Se ha propuesto que el aumento de expresión de *MUC2* pueda ser la consecuencia de la activación de citosinas proinflamatorias en respuesta a la infección por *H.pylori* (437). Este aumento de expresión también puede reflejar la transformación del estómago, durante la metaplasia intestinal, a un epitelio del tipo intestinal, en el que las células gástricas glandulares son reemplazadas por células caliciformes. En relación al papel de *MUC2* en la carcinogénesis gástrica se ha sugerido que pueda contribuir a un fenotipo maligno evitando el reconocimiento por parte de efectores inmunes antitumorales (350).

La región 5' de *MUC2*, de 29 exones, está claramente diferenciada de la región 3', que contiene 19 exones, por su largo exón central de repeticiones en tándem (TRs), a 5' del cual parece haber una ruptura del LD (280,290). Significativamente, los polimorfismos asociados con una menor (rs10902073, rs10794281, rs794472) o mayor (rs2071174) probabilidad de regresión estaban localizados en la región central y la región 5' o región genómica promotora, mientras que los polimorfismos asociados negativamente con la progresión (rs10794293, rs3924453, rs4077759) se localizaron en la región 3' del gen o región adyacente. La compleja organización del gen y la presencia de un bajo grado de LD en la región dificultan la interpretación de la variabilidad genética en *MUC2* y la evolución de las lesiones gástricas precursoras de cáncer gástrico. En este contexto, el análisis de asociación de los haplotipos de *MUC2* con la progresión y regresión nos proporcionó información adicional. Cuando se realizó el análisis de haplotipos, el haplotipo CTCCCG en la región 5' del gen se asoció con una mayor probabilidad de regresión y, en concordancia, también se identificó un efecto protector de este haplotipo contra la progresión de lesiones. Además, el haplotipo CATAGAAC de la región central y 3' se asoció también significativamente con un menor riesgo de progresión. Finalmente, cuando el análisis se realizó con todos los SNPs de *MUC2*, uno de los haplotipos CTCCCGCATAGAAC, que combinaba los haplotipos protectores de 5' y 3', también mostró una tendencia, aunque no significativa, como

factor protector de la progresión. Todos estos resultados sugieren que la combinación de diferentes variantes genéticas en el gen o el efecto de una variante no identificada en LD con los SNPs analizados en cada uno de las regiones 5' y 3' del gen, reduce la capacidad de progresión de las lesiones hacia el carcinoma gástrico. Estos resultados no informan de la naturaleza de las variantes funcionales responsables de las asociaciones observadas. Sin embargo, es de suponer que las variantes causativas están en elevado LD con los SNPs significativos ya que hemos utilizado una estrategia de selección de tagSNPs. Aunque deben llevarse a cabo estudios funcionales y análisis de asociación con los niveles de expresión para analizar las relaciones causales entre las variantes asociadas y la carcinogénesis gástricas, diversas líneas de evidencias indican que esta relación podría existir. Los SNPs rs10902073 y rs10794281 están en relativo elevado LD ($D'=1$, $r^2=0.773$) sugiriendo que la asociación encontrada puede ser debida al LD existente entre estos dos polimorfismos. Estos dos polimorfismos marcan un haplotipo de 7 kb localizado a 5 kb a 5' del gen *MUC2* por lo que el polimorfismo responsable de la asociación negativa con la regresión podría estar localizado en una secuencia reguladora del promotor del gen. De hecho, la expresión de los genes *MUC* del clúster en 11p15.5 está regulada por mecanismos epigenéticos en células epiteliales que, en el caso de *MUC2*, afectan a la región distal del promotor. Específicamente, el patrón de expresión es el resultado de una combinación de metilación y modificación de la cromatina en células de cáncer gástrico, de colon y pancreático, y la activación de la expresión de *MUC2* durante la diferenciación celular coincide con una pérdida de metilación de su promotor (315,316,438). Por otro lado, el rs2071174 se asoció significativamente con una mayor probabilidad de regresión. Este SNP, juntamente con el rs2856082, marca el bloque de haplotipos que contiene la región promotora y los primeros exones del gen donde se localiza el SNP no sinónimo V116M (rs11825977, $D'=1$ y $r^2=0.57$ con el rs2071174). Un estudio sugirió que este SNP puede ser responsable de la significativa reducción de la expresión de *MUC2* en la enfermedad de Crohn debido a la reducción de la expresión de su mRNA (439). Además, el polimorfismo V116M se localiza en uno de los dominios von Willebrand que median la dimerización de la mucina secretada permitiendo la formación de la capa extracelular altamente viscosa. Si se analiza la asociación del haplotipo CC formado por los dos tagSNPs anteriores, rs2856082 y rs2071174 y que marca el alelo menos frecuente del rs11825977 (V116M), se observa una asociación significativa como factor protector de la progresión y muestra una tendencia, aunque no significativa, con una mayor probabilidad de regresión. El alelo G del SNP sinónimo rs7944723 (P1832P), localizado en el extremo del exón 30 que contiene la región TR, se asoció inversamente con la regresión. El dominio central de TRs es altamente polimórfico en longitud y secuencia y contiene secuencias ricas en residuos de serina y treonina, donde se unen las cadenas de oligosacáridos (440). Este núcleo proteico glicosilado tiene diversas propiedades esenciales para las funciones protectoras de la superficie epitelial llevadas a cabo por las mucinas (262). Por el contrario, el alelo A del rs10794293 en el intrón 37, se asoció inversamente con la progresión. Ambos SNPs, juntamente con el rs6421972, marcan un haplotipo de 5 kb en la región 3' que corresponde al dominio del extremo C-terminal que contiene el sitio de proteólisis autocatalítico responsable de la disociación de *MUC2*. Se ha

sugerido que la disociación de los polímeros de *MUC2* pueda afectar a la viscosidad del mucus y pueda proporcionar a la mucina la capacidad de unirse a otros componentes (285,441,442). Es de destacar que el haplotipo GAC, que contiene el alelo G del rs7944723 asociado a menor regresión y el alelo común C del rs10794293, cuyo alelo menos frecuente T se asoció a menor riesgo de progresión, se asociaron a un mayor riesgo de progresión, sugiriendo que esta combinación específica de alelos pueda marcar un polimorfismo funcional, en uno de estos dominios particulares (TR y sitio de proteólisis autocatalítica) de la mucina secretada, responsable de la asociación encontrada. Los polimorfismos asociados significativamente a un menor riesgo de progresión, rs4077759 y rs3924453, son marcadores de un bloque de 4 kb que contiene los últimos exones del gen y región genómica adyacente. El relativo elevado LD encontrado entre estos SNPs en la población caso-control ($D'=0.992$, $r^2=0.687$) sugiere que la asociación encontrada puede ser debida al LD existente entre estos dos polimorfismos, más que a un efecto independiente de cada uno de ellos. Ambos SNPs están localizados en la región genómica a 3' del gen y marcan un SNP (rs7928098) localizado en la región 3'UTR de *MUC2*, según los datos del HapMap. Este tipo de SNPs, como se ha comentado anteriormente, pueden tener diversos efectos en la regulación de la expresión génica. Además, en el mismo bloque de LD se encuentran otros SNPs codificantes no sinónimos que pueden ser considerados como potencialmente funcionales. Es interesante destacar que no se encontró asociación con ninguno de los SNPs de cambio de aminoácido de *MUC2* analizados, lo que podría ser indicativo de que la asociación observada esté causada por variantes que regulan la expresión de *MUC2*. Tampoco se puede descartar que la posible variante causal, marcada por los tagSNPs asociados significativamente, no sea un SNP sino un polimorfismo estructural o incluso el mismo VNTR de la región central, no analizado en este estudio, ya que algunos estudios han demostrado que tanto pequeñas inserciones/deleciones como otros polimorfismos estructurales están en LD con tagSNPs (408).

Otros dos haplotipos distintos en la región 5' (CCGCG) y 3' (GGAAGAG) de *MUC2* se asociaron, respectivamente, a un mayor riesgo del cáncer gástrico del no cardias y a un menor riesgo de cáncer gástrico globalmente en el estudio caso-control, confirmando que el efecto de las variantes de *MUC2* en la evolución de las lesiones gástricas se mantiene en cáncer gástrico, aunque en menor grado y en sentidos muy variables en función de los subtipos histológicos y anatómicos y los modelos de herencia analizados. Algunos de los SNPs en la región 5' del gen asociados previamente a una menor (rs10902073) o mayor probabilidad (rs2071174) de regresión, se asociaron en concordancia a un mayor riesgo de cáncer gástrico del tipo intestinal (rs10902073) o como factor protector de cáncer gástrico, en los individuos no infectados, y del cáncer gástrico del cardias en hombres (rs2071174). Otros SNPs de la región 5' de *MUC2* que previamente no se habían asociado a la evolución de lesiones se asociaron con alguno de los subtipos de cáncer gástrico: el rs2856082, que juntamente con el SNP rs2071174 marca un bloque que engloba la región promotora y los primeros exones de *MUC2*, se asoció a un mayor riesgo de cáncer gástrico no cardias y el rs11245936 se asoció como factor protector de cáncer gástrico intestinal. Ambos SNPs presentan una frecuencia relativamente baja (MAF<0.01) por lo que es posible que no se pudiera detectar su efecto en el estudio de evolución de lesiones gástricas

debido a su limitación de potencia estadística. Referente a la región 3' de *MUC2*, el rs6421972 (G/A) se asoció significativamente como factor protector del cáncer gástrico no cardias. Este SNP es marcador del bloque de la región C-terminal donde se localiza el sitio autocatalítico. Otro SNP de la región 3' del gen, rs7480563, se asoció a un mayor riesgo de cáncer gástrico del tipo intestinal. Este SNP, no asociado previamente con la evolución de lesiones gástricas, marca un bloque de haplotipos adyacente al marcado por los SNPs rs4077759 y rs3924453, asociados con la progresión y localizados en el extremo 3' del gen. Además, este SNP se encuentra en LD ($D'=1$, $r^2=$) con el rs3924453 en la población caso-control, para la cual se excluyó el rs4077759 por problemas de genotipado. En general, de estos resultados se deduce que existe cierta variabilidad de *MUC2* que puede influir en el desarrollo del cáncer gástrico, especialmente del tipo intestinal y del no cardias, que son la localización y subtipo que mejor se relacionan con la progresión de lesiones precursoras de cáncer gástrico, iniciadas por la infección por *H.pylori*. Aunque también existieron diferencias en la asociación con cáncer gástrico en función del sexo, en la mayoría de las ocasiones la tendencia fue similar en los dos subgrupos y no se hallaron prácticamente interacciones significativas excepto en el subtipo anatómico del cardias. Esta mayor asociación de algunos de los SNPs de *MUC2* en hombres en los casos del cardias puede ser debida a las diferencias epidemiológicas existentes, ya que este subtipo de cáncer es mucho más frecuente en hombres. No obstante, para poder verificar una posible interacción de este gen con el sexo, del cual no existen precedentes en humanos aunque sí en modelos animales (443), deberían llevarse a cabo estudios con un mayor tamaño muestral para ambos sexos.

El efecto de todos los SNPs de *MUC2* asociados con la evolución de lesiones gástricas fue mayor y únicamente significativo en los pacientes infectados por *H.pylori*, siendo la interacción estadísticamente significativa para los SNPs rs10902073, rs10794281y rs4077759. A pesar del relativo bajo número de pacientes negativos para la infección (23%), que implica las evidentes limitaciones de potencia estadística, los ORs calculados no siguieron la misma tendencia en los pacientes no infectados, sugiriendo una interacción real entre estas variantes y la bacteria, en la evolución de las lesiones. Esta interacción no se observó tan claramente en todos los SNPs asociados al cáncer gástrico, aunque rs2856082, asociado a cáncer gástrico del no cardias, lo es con mucha mayor significación para los pacientes infectados, siendo la interacción significativa. La capa de mucus viscoso formado por las mucinas secretadas, incluyendo *MUC2*, proporciona una barrera física contra los microorganismos. Aunque no se ha establecido una interacción directa entre *H.pylori* y *MUC2* en el estómago, la importancia de esta mucina en la colonización e infección bacteriana ha sido ampliamente demostrada en el intestino (323). En tejido gástrico, sólo se han observado relaciones indirectas entre *H.pylori* y *MUC2*. La infección por *H.pylori* causa una expresión anormal de *MUC2* en líneas células de cáncer gástrico (237) y se ha asociado la expresión *de novo* de *MUC2* con una inducción de CDX1 por parte de la cepa *cagA* de *H.pylori* en células epiteliales gástricas que han sido retenidas en G1 (83). Aunque no conocemos el mecanismo preciso por el cual la variabilidad genética de *MUC2* puede interaccionar con la infección de *H.pylori* para influenciar la evolución de las lesiones gástricas, el hecho de que la infección se asocie con la expresión anormal de *MUC2* y que variantes en *MUC2* se asocien a la

evolución de las lesiones, sugiere que estas variantes podrían modificar la expresión o la función de la proteína de tal modo que alteren el efecto de la infección sobre la expresión de *MUC2*. Que la mayor evidencia de interacción entre la infección con *H.pylori* y variantes en *MUC2* se de en la evolución de las lesiones podría explicarse por el hecho de que el aumento de expresión de esta mucina se inicia en la metaplasia intestinal y, por tanto, la posible combinación entre el efecto de *MUC2* y la bacteria sería particularmente evidente durante la progresión de las lesiones gástricas. Por otro lado, nuestros resultados mostraron que el efecto de las variantes asociadas a la evolución de lesiones sólo fue significativo en los no consumidores de AINES, aunque no se encontró interacción entre ambos factores. Dado el efecto protector del consumo de AINES en la progresión de las lesiones comentado anteriormente, este resultado podría explicarse como que el efecto protector de la variación genética en *MUC2* se observa más claramente en aquellos pacientes no protegidos ya por ese consumo.

TFF3 es otro de los genes que se expresan normalmente en el intestino, en las células caliciformes, pero que aumenta su expresión en la carcinogénesis gástrica siguiendo un patrón muy similar al de *MUC2*, con el cual colocaliza en tejido gástrico (237,444,445). *TFF3* es el gen *TFF* cuya expresión tiene más amplia distribución tisular y, aunque este gen tiene un papel en la restitución de la mucosa gástrica, recientemente se ha relacionado con múltiples patologías (215,446-453), existiendo generalmente una correlación entre mayor expresión del gen y malignidad y proponiéndose en prácticamente todos los estudios como un potencial biomarcador de cáncer. *TFF3* promueve la migración *in vitro* y la angiogénesis y estas funciones, que son esenciales para la restitución celular, pueden asociarse también a invasión celular y progresión tumoral (226,227,454). En relación al cáncer gástrico también se ha descrito esta correlación entre mayor expresión de *TFF3* y peor pronóstico y se ha demostrado que los niveles de *TFF3* en suero, en combinación con los de pepsinógeno son eficientes marcadores del cáncer gástrico (241,248,249,455).

Aunque la asociación entre variabilidad genética de *TFF3* con la evolución de lesiones gástricas no fue muy evidente, en relación al cáncer gástrico se observaron algunas asociaciones significativas. El SNP c*41G>A en la región 3'UTR de *TFF3*, y con una $MAF \leq 0.01$ fue el único que se asoció con una mayor probabilidad de regresión de lesiones gástricas y, por el contrario, mostró una asociación significativa como factor de riesgo para cáncer gástrico difuso. Este SNP posiblemente altere, disminuyéndola, la expresión del gen, tal como observamos en los ensayos luciferasa realizados. Podría ser que una disminución de la expresión del gen pueda favorecer la regresión de lesiones, asociadas mayoritariamente al cáncer gástrico de tipo intestinal, pero ser factor de riesgo para el tipo difuso. El cáncer gástrico difuso difiere etiológica y epidemiológicamente del tipo intestinal. Este último es el que se desarrolla a partir de etapas muy diferenciadas, entre las cuales metaplasia intestinal, cuya progresión o regresión se evalúa principalmente en el estudio de evolución de lesiones gástricas. Por otro lado, estos resultados deben tratarse con mucha cautela y no se puede descartar que se trate de falsos positivos ya que se basan en un SNP muy poco frecuente ($MAF=0.01$ en la población de Soria y $MAF < 0.005$ en la población EPIC Europea), sin presencia de homocigotos para el alelo menos frecuente en ninguna

de las dos poblaciones analizadas y con unos intervalos de confianza de sus ORs excesivamente amplios, cuando se han podido calcular. Este es el caso también de otro de los SNPs potencialmente funcionales de *TFF3*, V80M, que se asoció con cáncer gástrico en mujeres. A pesar de su baja frecuencia, que dificulta la interpretación de los resultados, este SNP no sinónimo es el único identificado y validado en el gen *TFF3* hasta ahora y es seguramente perjudicial según el análisis *in silico* de su efecto utilizando distintas herramientas. Además, tres tagSNPs mucho más frecuentes del mismo bloque de LD, que abarca todo el gen *TFF3*, se asociaron a un menor (rs8133510) o mayor riesgo (rs2236705) de cáncer gástrico o como factor protector (rs225363) del intestinal, sugiriendo la existencia de alguna variante causal en el gen *TFF3* o su región más adyacente, que bien podría ser alguno de los SNPs potencialmente funcionales ya identificados. No obstante, la baja frecuencia de estos SNPs aparentemente funcionales y el poco grado de LD en la región, impiden que este punto pueda confirmarse mediante el análisis de haplotipos. Además, parece que el mayor efecto en la asociación se da con los genotipos heterocigotos de estos SNPs, sugiriendo que esta variante causal podría ser un SNP de cambio de aminoácido que afecte la correcta formación de homo/heterodímeros. Otro tagSNP, rs692356, se asoció como factor protector del tipo intestinal. Este SNP marca un bloque de LD en la región genómica a 5' de *TFF3* donde se localiza el promotor distal del gen, que contiene diversos elementos reguladores, como *enhancers* y regiones silenciadoras (193). Además el rs225363, que forma parte del bloque de LD que engloba el gen y que se asoció también como factor protector, se localiza en a 5' del gen, a 2 kb del inicio de transcripción. Recientemente se ha reportado la asociación de un SNP del promotor de *TFF3*, el rs9981660, con un menor riesgo de cáncer gástrico difuso en hombres de población china y una mayor expresión del mRNA (261). Aunque este polimorfismo se seleccionó como tagSNP en nuestro estudio, se excluyó del análisis debido a problemas en el genotipado, como también el tagSNP potencialmente funcional de la región 5'UTR, el rs11701143. No obstante, es posible que la variante causal se localice en la región promotora o 5'UTR del gen afectando la regulación de su expresión. *TFF3* contiene islas CpG en su promotor y se ha descrito la regulación epigenética del gen en algunas localizaciones, observándose una asociación entre la hipometilación del promotor y su aumento de expresión en cáncer de próstata o de hígado (447,456). Además, se ha descrito la regulación del gen por parte de CDX2, un factor de transcripción regulador específico de genes intestinales, identificándose sitios potenciales de unión a esta proteína en el promotor (457). Este factor de transcripción, como *MUC2* y *TFF3*, aumenta su expresión durante la metaplasia intestinal y el cáncer gástrico (237,458,459).

Los resultados más relevantes en relación al papel de los genes de protección de la mucosa gástrica en cáncer gástrico se observaron para *MUC1* y *TFF1*, los cuales se asociaron significativamente o casi en el test de permutaciones por genes (*MUC1* valor $p=0.045$, *TFF1*, valor $p=0.052$). De todas las mucinas analizadas, *MUC1* es la única asociada a membrana y, por tanto, desempeña un papel diferente en el mantenimiento de la mucosa gástrica. Además de su posible implicación en limitar la patología que resulta de las infecciones bacterianas en el glicocáliz (270,328-330), también actúa como transductor de señales participando en múltiples

procesos de regulación (264). Aunque en cáncer gástrico frecuentemente disminuye su expresión, se ha correlacionado una expresión positiva de MUC1 en tumores gástricos con progresión, invasividad celular y mal pronóstico (353,354,460,461). Además, su subunidad terminal (MUC1-C) se considera una oncoproteína en algunos tumores, donde además su expresión aumenta considerablemente (355,462-464).

Tres de los cuatro SNPs analizados en *MUC1* se asociaron significativamente con un mayor (alelos T de rs3814316 y rs9426886) o menor riesgo de cáncer gástrico (alelo G de rs4072037). Los SNPs rs3814316 y rs9426886 se localizan en la región genómica a 3' de *MUC1*, muy próximos al gen y se encuentran en cierto LD con rs4072037, un SNP sinónimo (T31T) localizado en el exón 2 de *MUC1*. El único SNP de *MUC1* que no se asoció a cáncer gástrico ni a ninguno de sus subtipos fue el rs1045253, localizado a una distancia de aproximadamente 40 kb a 5' del gen. Estos resultados indican que la variante causal se encuentra en el gen o en su región próxima adyacente. Los haplotipos que llevaban los alelos de riesgo para los tres SNPs asociados mostraron una tendencia de riesgo en cáncer gástrico (TTAT) o se asociaron significativamente con el del cardias (TTAC). Estos resultados con haplotipos confirmaron las asociaciones encontradas mediante el análisis individual de SNPs aunque no aportaron información adicional, a excepción quizá, del hecho que el haplotipo más común y, en consecuencia, de referencia, no es el que contiene el alelo más común (A) de rs4072037, el cual se encuentra en distintos haplotipos, sino el que contiene el alelo protector, G. Así, el haplotipo que contiene el alelo común, A, junto con los alelos de riesgo de los SNP asociados a cáncer gástrico (TTA), pasa a ser un haplotipo menos común y de riesgo. El SNP rs4072037 es uno de los polimorfismos cuya asociación con cáncer gástrico se identificó mediante GWA en población japonesa (153,166) y que se ha replicado en otras poblaciones (363,465,466), incluida la de origen europeo (168,364). Este SNP es muy común, en nuestra población caso-control MAF=0.47, siendo el alelo G el menos frecuente, y en algunas poblaciones pueden invertirse las frecuencias. Aun así, todos los estudios identificaron el alelo G como factor protector o el alelo A, generalmente el más frecuente, como factor de riesgo bajo muy diversos modelos de herencia. Este SNP es el responsable del procesamiento alternativo en el inicio del exón 2, que genera las dos isoformas más frecuentes del gen, que se diferencian en una retención intrónica alternativa de 27 pb (167,288); mientras que el alelo A genera la isoforma más corta, el alelo G genera la más larga (166) y afecta también a otras isoformas minoritarias, tanto en condiciones normales como en cáncer. Esta variante también podría afectar la función de la proteína. El evento de procesamiento alternativo modifica el péptido señal y el extremo N-terminal de la proteína madura, afectando posiblemente un sitio de proteólisis. Esta modificación podría alterar el tráfico intracelular y la glicosilación y plegamiento de la proteína, afectando a su funcionalidad (167). Además, ensayos funcionales con este SNP y otros de la región promotora que se encuentran en el mismo bloque de haplotipos y en elevado LD, demostraron que el haplotipo que contiene el alelo protector G tiene una mayor actividad transcripcional que el haplotipo que contiene el alelo A (166). Los tres SNPs de *MUC1* se asociaron únicamente al tipo intestinal, aunque el test de heterogeneidad sólo fue significativo para el rs9426886. El rs4072037 fue significativo únicamente en el cardias pero

mostró una tendencia similar en el no cardias y el test de heterogeneidad indicó que no había diferencias significativas en el efecto de este SNP sobre las dos localizaciones del cáncer gástrico. Existen pocos estudios previos en los que se analice diferencialmente el efecto del rs4072037 en los dos subtipos histológicos y anatómicos de cáncer gástrico y los resultados en algunos casos son discrepantes. En población asiática se ha asociado únicamente al cáncer gástrico difuso en un estudio (166) y en otro no se ha encontrado diferencias significativas entre subtipos (465). Recientemente, en población europea se ha asociado este SNP tanto al tipo intestinal como al cardias, con resultados similares a los de nuestro estudio (168). El test de permutaciones realizado en nuestro estudio confirmó la asociación de este gen con el tipo intestinal y se acercó a la significación estadística en cáncer gástrico globalmente y en el cardias. Por tanto, aunque estos resultados deberían reproducirse en otros estudios en población europea, que además analicen el cáncer gástrico por subtipos, todo indica que en esta población el alelo G del rs4072037 es un factor protector para cáncer gástrico y, específicamente y en mayor medida, para el tipo intestinal y, probablemente, en el cardias.

En relación a su implicación en la evolución de las lesiones gástricas, el rs4072037 se tuvo que excluir del análisis debido a problemas en el genotipado. Aun así ninguno de los otros tres tagSNPs seleccionados en la región donde se localiza *MUC1* mostró una asociación significativa. Es posible que su posible efecto funcional sea más determinante en el desarrollo final del cáncer gástrico, de los cuales hay muy pocos casos en el estudio de evolución de lesiones, y no tanto en la progresión o regresión de las lesiones previas. En este sentido, hay que considerar también las limitaciones de potencia del estudio y, por tanto, se requiere de un análisis de asociación en una muestra más amplia para dilucidar y confirmar el papel de este polimorfismo en la evolución de las lesiones gástricas precursoras de cáncer gástrico.

Otros de los resultados más relevantes en relación al cáncer gástrico fueron los obtenidos para el gen *TFF1*. Se ha propuesto a este péptido como un supresor de tumores, específico de cáncer gástrico. Entre algunas de las evidencias se ha observado que el ratón knockout para *TFF1* desarrollaba una mucosa gástrica alterada y predisposición a cáncer gástrico (207). También se han descrito mutaciones somáticas, pérdida de heterocigosidad e hipermetilación como mecanismo de inactivación de la expresión de *TFF1*, observada en un elevado porcentaje de tumores gástricos (239,250,251) y que, además, se ha asociado a un peor pronóstico (241). Sin embargo, este gen aumenta frecuentemente su expresión en el cáncer de mama (467,468), y su papel fisiopatológico en esta patología es muy controvertido ya que existen tanto estudios que lo relacionan con una capacidad oncogénica (214) como, por el contrario, otros lo relacionan una capacidad de reducir el desarrollo tumoral (469).

La variabilidad genética de *TFF1* también mostró una fuerte asociación con el cáncer gástrico del tipo intestinal ya que seis de los ocho SNPs genotipados de este gen se asociaron significativamente a un mayor (rs13276, rs2839488, rs9976977, rs13047838) o menor (rs225356 y rs424694) riesgo de este tipo de cáncer. El análisis de haplotipos, confirmó esta compleja asociación de la diversa variabilidad en el gen *TFF1* con el subtipo intestinal ya que tres haplotipos distintos se asociaron con el cáncer gástrico intestinal: CCGGGGG asociado como

factor protector y CCAGCGGA y CCGGCGGG asociados a un mayor riesgo. En general, estos haplotipos portan los distintos alelos de riesgo o protectores identificados en el análisis de SNPs individual en diferentes combinaciones. Muchas de las asociaciones con SNPs individuales de *TFF1* y cáncer gástrico o sus subtipos se observaron en los modelos codominante, dominante o sobredominante, modelos que no quedan representados cuando se considera el modelo aditivo en el análisis de asociación de haplotipos. Este hecho conjuntamente con la enorme variabilidad, el poco LD existente en el gen y región adyacente (los seis SNPs asociados abarcan una región 10 kb) y la diversidad de los modelos de herencia, dificultan la interpretación de la relación entre los resultados de asociación de haplotipos y todas las asociaciones observadas previamente en el análisis de SNPs individuales.

Tres de los SNPs asociados con el tipo intestinal (rs9976977, rs424694 y rs13047838) son marcadores del bloque de LD que engloba la región genómica a 5' del gen donde se localiza la región promotora. Estos tres SNPs también se asociaron a la evolución de lesiones con una mayor probabilidad de regresión. El mayor efecto de estos SNPs, con la excepción de rs9976977, se observó para el genotipo heterocigoto. Los genotipos CT del rs424694 y rs13047838 se asociaron a una mayor probabilidad de regresión de lesiones gástricas y también como factores protectores del cáncer gástrico del tipo intestinal (sólo en mujeres para el rs13047838). Además, el homocigoto TT del rs13047838 se asoció como factor de riesgo para el tipo intestinal, en mujeres. El rs9976977 se asoció con una mayor probabilidad de regresión en la evolución de lesiones bajo un modelo codominante pero como factor de riesgo del tipo intestinal bajo un modelo recesivo. Aunque el modelo de herencia no queda claro en estos resultados, si que se deduce que la posible variante causal, marcada por estos 3 SNPs, podría estar localizada en el promotor y, por tanto, afectar a la regulación y expresión del gen. El promotor del gen de *TFF1* contiene una región compleja de estimuladores de la transcripción (186-188) y es un gen clásicamente regulado por estrógenos en diversos tejidos. *TFF1* contiene en su promotor un ERE, un elemento de respuesta a AP1 y sitios de unión para Sp1/Sp3, todos ellos esenciales para su activación en respuesta a estrógeno en células de cáncer de mama (470,471), donde este gen aumenta frecuentemente su expresión. Aunque no hay evidencias claras de que *TFF1* sea un gen regulado por estrógeno en el estómago, y existen muy pocos estudios al respecto, los dos tipos de receptores del estrógeno (ER α and ER β) están presentes en el estómago y en la mucosa gástrica (472,473). El nivel de expresión y posible valor pronóstico de los ER en cáncer gástrico es discrepante en los diversos estudios realizados, aunque recientemente se ha descrito la disminución de la expresión de estos receptores en tumores gástricos, comparado con la mucosa normal adyacente, y la expresión positiva de estos receptores se ha asociado con el tipo intestinal (474). Aunque el mecanismo subyacente al potencial papel del estrógeno en la etiología del cáncer gástrico no está claro, la mayor incidencia de cáncer gástrico en hombres podría ser explicada, en parte, por un efecto protector de los estrógenos en mujeres, particularmente para el cáncer gástrico del tipo intestinal (6-8). Se ha propuesto que una de las vías por la cual el estrógeno pueda ejercer este efecto es la inducción de la expresión del gen *TFF1* (6). Significativamente, uno de los tagSNPs del bloque, rs13047838, mostró una interacción

significativa con el sexo en el tipo intestinal, asociándose únicamente a las mujeres. Además, un estudio reciente ha descrito la asociación de un polimorfismo localizado en el ERE del promotor de *TFF1*, a unas 400 pb del inicio de transcripción (rs3761376), con un mayor riesgo de cáncer gástrico. La sustitución nucleotídica en la secuencia del ERE, de C a T, podría resultar en una disminución de la expresión de *TFF1* y un incremento en el riesgo de cáncer gástrico (260). Este SNP, rs3761376, no se encontraba en la base de datos de HapMap en su segunda fase ni está genotipado actualmente para población europea en su tercera fase, por lo que no podemos conocer la relación exacta de LD que presenta con los tres SNPs asociados en nuestro estudio. No obstante, ya puede ser que la variante causal que estén marcando no sea el rs3761376, ya que por localización se encontraría fuera de la influencia del bloque de LD que marcan los tagSNPs a 5' del gen, sino otra variante se encuentre en una zona más distal del promotor, mucho menos explorada. La variante rs3761376 parece estar incluida en el bloque de LD que marcan los SNPs rs225356 y rs2839488 del intrón 1, ambos asociados significativamente con el tipo intestinal, y que abarca la mitad 5' del gen incluyendo la región 5'UTR y la zona más proximal del promotor. En ambos casos, el genotipo homocigoto para el alelo menos frecuente se asoció a un mayor riesgo del tipo intestinal y, por el contrario, el genotipo heterocigoto se asoció como factor protector, aunque en el caso del rs2839488 únicamente en mujeres mostrando interacción significativa con el sexo. Esta tendencia es idéntica a la observada para los tagSNPs asociados del bloque a 5' del gen, con los cuales existe cierto elevado LD en valores de D' aunque no de r^2 , por lo que no se puede descartar que estén marcando a la misma variante causal. Otra indicación del papel que las hormonas podrían ejercer en la regulación de *TFF1* es que ambos tagSNPs mostraron una interacción significativa con el sexo en los subtipos anatómicos. Mientras que el rs23839488 se asoció a un mayor riesgo de cáncer gástrico no cardias en hombres, el rs225356 se asoció como factor protector del no cardias en mujeres. Además de, probablemente, el rs3761376, en el bloque de LD que marcan los tagSNPs rs225356 y rs2839488 se localizan otros SNPs potencialmente funcionales, entre los cuales dos en la región 5'UTR: c.-29G>A (actualmente rs145185427) y c.-2C>T (rs2156310) y un cambio sinónimo T22I (rs34795821). Desafortunadamente ninguno de los tres se incluyó en el análisis por problemas de diseño o de genotipado y no conocemos su asociación con cáncer gástrico, aunque la tipología y localización de estos SNPs los hace candidatos como posibles variantes causales. El promotor mínimo de un gen, es decir, la secuencia mínima de DNA que es suficiente para iniciar la transcripción, generalmente abarca unas 80 pb alrededor del inicio de transcripción y pueden ser promotores conservados, con caja TATA, o promotores variables ricos en islas CpG. En el caso de *TFF1* se ha descrito una caja TATA entre -24 y -30 pb del inicio de transcripción así como una caja CAAT y motivos ricos en G y C (475). Aunque el rs145185427 se localiza a unas 40 pb de la caja TATA, y tampoco coincide con los otros elementos, se encuentra en la región de influencia del promotor mínimo. Además, las regiones 5'UTR contienen diversos elementos reguladores, así como motivos estructurales, que juegan un importante papel en el inicio de la traducción y, además, pueden interactuar con diferentes proteínas de unión al RNA (476). Es interesante remarcar que el rs2156310 se encuentra a tan sólo 2 pb del codón de inicio de la traducción por lo que es posible

que estas variantes puedan afectar a la regulación post-transcripcional del gen y/o afectar la traducción del mismo. Los tres SNPs potencialmente funcionales que podrían ser candidatos como variante causal de la asociación con cáncer gástrico (rs145185427, rs2156310 y rs34795821), si se analizaron en relación a la evolución de lesiones gástricas. Estos SNPs son muy poco frecuentes, en dos casos $MAF=0.01$, y aunque no se encontró asociación significativa con la progresión o regresión de lesiones no se encontraron prácticamente homocigotos para el alelo menos frecuente por lo que la potencia del estudio limitó mucho su análisis. Sin embargo, un tagSNP del intrón 1, en el mismo bloque de LD, rs225358 que se excluyó en el estudio caso-control por problemas de genotipado, si se asoció como factor de riesgo para la progresión de lesiones. Este SNP, además de encontrarse muy próximo a los SNPs potencialmente funcionales, se encuentra a tan sólo 8 pb de la unión exón-intrón y podría alterar el procesamiento alternativo creando un sitio donador según el análisis *in silico*. Finalmente, el rs13276 (c.*41G>A), localizado en la región 3'UTR, se asoció a un mayor riesgo de cáncer gástrico del tipo intestinal. Este SNP es un polimorfismo potencialmente funcional que, como se ha comentado anteriormente, se asoció a una disminución de la expresión en los ensayos luciferasa. Además, parece que el cambio de G a A aumente la afinidad por uno de los miRNAs testados, el miR-760, provocando así también una represión de esta región. El SNP el c.*41G>A no fue seleccionado como marcador y, significativamente, el tagSNP que marca el bloque de LD donde se localiza, rs225353, no se asoció a cáncer gástrico. Este hecho, juntamente con la posible funcionalidad confirmada mediante ensayos luciferasa, hace plausible que la variante de 3'UTR sea la variante causal responsable de la asociación encontrada. Este SNP poco frecuente ($MAF=0.05$) no se asoció a la evolución de lesiones gástricas. Por tanto, parece ser que distintas variantes a lo largo del gen *TFF1* y regiones reguladoras adyacentes modifican el riesgo del cáncer gástrico del tipo intestinal, y probablemente la evolución de las lesiones gástricas. El análisis de haplotipos confirmaría esta hipótesis ya que fueron tres los haplotipos asociados a este tipo de cáncer, evidenciando la enorme variabilidad del gen y distintos efectos en la carcinogénesis en función de las variantes que se combinen en el haplotipo. En la región también hay distintos SNPs candidatos a ser las variantes causales, cuya asociación con la carcinogénesis gástrica y posible funcionalidad debería abordarse con estudios con un mayor tamaño muestral y ensayos funcionales específicos

El efecto de la mayoría de los SNPs de *TFF1* se observó en los no infectados por *H.pylori*. Debido al elevado efecto de riesgo de de *H.pylori* en cáncer gástrico, muy superior al efecto de las variantes genéticas asociadas, es posible que este efecto de los SNPs sólo sea visible aquellos pacientes que no estén infectados.

MUC6 y *TFF2* colocalizan en la mucosa gástrica y su mayor nivel de expresión se encuentra en las células mucosas del cuello de las glándulas oxínticas y pilóricas (181).

MUC6 es una de las principales mucinas secretadas del estómago y su función principal es la formación del mucus que protege el epitelio subyacente contra múltiples agresiones y actúa como barrera física contra microorganismos (323). En cáncer gástrico, la expresión de esta mucina está generalmente disminuida, aunque puede ser positiva en el tipo difuso (277). En

nuestro estudio, las variantes de estos genes no se asociaron con la evolución de las lesiones gástricas y sólo se observaron algunas asociaciones puntuales con cáncer gástrico o alguno de sus subtipos. No obstante, un haplotipo poco frecuente de *MUC6*, GCTGG, se asoció al cáncer gástrico del no cardias y contenía los alelos del tagSNP de cambio de aminoácido (V619M, rs7481521A>G) asociado al no cardias y del tagSNP intrónico asociado a cáncer gástrico globalmente (rs11246384, C>T). *MUC6* se localiza en una región de elevado LD y, además, contiene múltiples SNPs descritos en sus regiones funcionales por lo que es realmente complejo, a partir de estos resultados, acotar la región donde pueda localizarse una posible variante causal de esta asociación, que debería ser confirmada en posteriores estudios de asociación en otras poblaciones. Los SNPs situados en la región 5' del gen no se asociaron con cáncer y aunque estos dos SNP están separados por unas 2.6 kb, que abarcan del intrón 7 hasta el exón 16, todos forman parte del mismo bloque de LD. Además, otro de los SNPs de cambio de aminoácido presente en esta región, G395R (rs11604757) no se asoció en nuestro estudio y el tagSNP situado más a 3' del gen, rs4076950 se excluyó del análisis. Teniendo en cuenta que el bloque de LD abarcaba unas 40 kb y el gen es relativamente grande, 22.9 kb, sería necesario hacer un mapeo más fino de la región e incluir SNPs potencialmente funcionales para poder confirmar las asociaciones encontradas y proponer posibles variantes causales. Podría ser, además, que la variante causal no sea un polimorfismo tipo SNP, ya que recientemente se han asociado determinados alelos de un minisatélite en el intrón 26 del gen con un mayor riesgo de cáncer gástrico (477).

El papel de *TFF2* en la carcinogénesis gástrica es el menos conocido de todos, aunque existen algunas evidencias, a nivel de expresión y regulación, que lo relacionan (246,331) y se ha asociado la hipermetilación de su promotor, inducida por *H.pylori*, con progresión tumoral (253). En relación a la asociación de variabilidad genética en *TFF2* y la carcinogénesis gástrica, el polimorfismo potencialmente funcional c.*67_68insCTT, que mostró cierta influencia en la expresión en ensayos luciferasa, se asoció ligeramente al cáncer gástrico localizado en el cardias en mujeres ($p=0.0460$), aunque se trata un polimorfismo muy poco frecuente (MAF=0.04) y la interpretación de estos resultados deben hacerse con mucha precaución. Sin embargo, un tagSNP rs3814896 (c.-645T>C) de la región genómica a 5' de *TFF2*, se asoció significativamente con un mayor riesgo de cáncer gástrico globalmente en los hombres, mostrando una interacción significativa con el sexo. Esta misma asociación se repitió de manera similar en los dos subtipos histológicos. Muy recientemente, este polimorfismo se ha asociado a un menor riesgo de cáncer gástrico en población china, especialmente del tipo difuso, y a gastritis atrófica, aunque únicamente en individuos menores de 50 años y en sentido inverso al observado en nuestra población (261). En nuestro estudio, además, el efecto diferencial de este SNP se encontró con el sexo aunque no existen antecedentes de la regulación de este gen por parte de hormonas sexuales o diferencias de expresión en función del sexo en condiciones normales o patológicas. Sin embargo, esta relación si se ha establecido para *TFF1*, como se ha comentado anteriormente, y para *TFF3*. Se ha descrito que la expresión de *TFF3* podría ser específica en función del sexo, siendo positiva en el caso de cáncer gástrico avanzado en hombres, y negativa en mujeres (236).

El rs3814896 se encuentra a tan sólo 500 pb del inicio de transcripción de *TFF2* y marca, junto otros tagSNPs, un bloque de LD que abarca la región 5' del gen, la región promotora y la región intergénica entre *TFF2* y *TFF1*, incluyendo también el extremo 3' de este último gen. Sin embargo, ningún otro tagSNP de este bloque se asoció con cáncer gástrico, por lo que la variante causal, en caso de confirmarse la asociación, puede encontrarse muy próxima a este SNP y en elevado LD. Uno de los SNPs potencialmente funcionales validados previamente, el rs13052596 (c.-450G>T), se encuentra a 200 pb de la variante asociada y a 308 pb del inicio de transcripción. Este SNP, que no se pudo genotipar pero que se seleccionó inicialmente porque podría afectar un posible sitio de unión a NF-κB según el análisis con PupaSuite, mostró también consenso para el factor de transcripción Sp1 en un análisis *in silico* más detallado. La regulación por parte de esta familia de factores de transcripción ya se ha descrito para *TFF1*, que se regularía por Sp1 y Sp3 (189), y para *TFF2*, por Sp3 (478). Además de la posible secuencia consenso para Sp1, *TFF2* contiene una secuencia consenso para la unión de Sp3 (-150 a -145) y dos secuencias ricas en G y C esenciales para la unión de este factor de transcripción, una de ellas localizada en -306 a -311 (478), justamente donde se localiza el rs13052596. Aunque se ha descrito que NFκB inhibe la transcripción de los genes *TFF* en líneas celulares gastrointestinales (197), no se ha establecido ninguna relación directa entre este factor de transcripción y *TFF2*. Desafortunadamente, el rs13052596 no se pudo analizar en el estudio caso-control y este polimorfismo no se ha genotipado en las poblaciones de HapMap. Sin embargo, si tomamos en cuenta los resultados de LD del estudio de evoluciones de lesiones gástricas, que en general correlacionan bien con las otras dos poblaciones (caso-control de cáncer gástrico, analizada en este estudio y Hapmap), estos dos SNP presentan una $D'=0.92$ y $r^2=0.52$, que denotan un considerable LD. Tampoco se puede descartar que existan otras variantes en esta región del promotor que puedan afectar la expresión del gen ya que se han descrito números elementos reguladores en su secuencia (192,204,479; Tu, 2009 #2495) y mecanismos epigenéticos de regulación (253).

Finalmente, referente a los genes de respuesta inflamatoria, reparación del DNA y respuesta a la infección por *H.pylori*, que en este estudio se analizaron únicamente en relación a la evolución de lesiones gástricas precursoras, los resultados más relevantes se encontraron para los genes relacionados con la vía de señalización de la bacteria.

Los alelos G y T de los dos SNPs analizados del gen *TLR4* (rs4986790, D299G y rs4986791, T399I) se asociaron a una menor probabilidad de regresión, así como su haplotipo GT. Estos dos polimorfismos se han asociado previamente a un mayor riesgo de cáncer gástrico en múltiples estudios (143,147,480). Se ha propuesto que estos SNPs pueden modificar la estructura normal de la región extracelular de TLR4 y, por tanto, alterar su unión con los liposacáridos (LPS), presentes en la pared de la bacteria, disminuyendo la capacidad de respuesta (143). Existen pocos estudios de los polimorfismos de *TLR4* D299G y T399I en relación a las lesiones gástricas precursoras y ninguno que aborde su asociación con la evolución de las mismas. Aun así, se ha asociado el SNP D299G con un mayor riesgo de gastritis atrófica e hipoclorhidria en población caucásica y el SNP T399I con un mayor riesgo de atrofia y metaplasia intestinal (140,141). A pesar de la baja frecuencia de estos SNPs en la población de pacientes con lesiones gástricas MAF

de 0.06 (D299G) y de 0.07 (T399I), su asociación con una menor probabilidad de regresión concuerda con estos estudios previos que los asocian con un mayor riesgo de determinadas lesiones gástricas precursoras de cáncer gástrico.

Un SNP frecuente (rs1125557, MAF=0.4) en el intrón 2 de *CHD1*, el gen que codifica para la E-cadherina, se asoció también con una menor probabilidad de regresión. No existen antecedentes en relación a lesiones precursoras, aunque este SNP se ha asociado a mayor riesgo de cáncer gástrico del tipo difuso en población italiana (481). No obstante, el polimorfismo que más recurrentemente se ha asociado a cáncer gástrico en este gen es un SNP localizado en el promotor, c.-160C>A (rs16260), aunque posiblemente sea factor de riesgo para población asiática pero no caucásica (482-484). Este SNP se incluyó inicialmente en nuestro estudio pero se descartó por problemas de genotipación con lo que no pudimos analizar su posible relación con la evolución de lesiones gástricas.

El alelo G de un SNP de la región 5'UTR de *CD14*, c.-260A>G, rs2569190, se asoció como factor protector de la progresión. Se trata de un SNP muy frecuente, MAF de 0.48 en nuestra población que puede invertir su frecuencia en determinadas poblaciones. El alelo A rs2569190 se ha asociado con el desarrollo de una mayor inflamación de la mucosa gástrica, en respuesta a la infección por *H.pylori* (485), y con un mayor riesgo de cáncer gástrico en población asiática, aunque existen estudios discrepantes y la asociación en población europea no esta tan clara (486). *CD14* desempeña un papel importante tanto en la respuesta inmunitaria a la infección de la bacteria como en la iniciación de la respuesta inflamatoria por parte de citocinas y está involucrado en el reconocimiento de LPS por parte del complejo TLR4/MD-2 (487,488). Ensayos de gen reportero han demostrado que los alelos de este SNP presentan diferente actividad transcripcional. En la región promotora de *CD14* se localizan distintos elementos reguladores, como los sitios de unión a Sp1 (-110) y C/EBP y, además, existe otra región, entre -360 pb y -1090 pb que parece ser crítica también para la expresión del gen (488).

Un SNP intrónico, rs2301756, del gen *PTPN11* se asoció a la progresión de lesiones en el modelo recesivo y, aunque no se pudieron calcular los ORs debido a su baja frecuencia, se deduce por las diferencias en las frecuencias genotípicas que su asociación sería como factor protector. Diversos estudios han asociado este polimorfismo con un menor riesgo de gastritis atrófica, pero no de cáncer gástrico, en población asiática (489-491). Sin embargo, no existen antecedentes de este tipo de asociación en población europea. El gen *PTPN11* codifica la fosfasa SHP-2, que se activa por fosforilación en respuesta a la interacción con CagA, iniciando la vía de señalización. Finalmente el SNP c.-488G>A (rs1800629), en la región promotora de TNF se asoció como factor de riesgo de la progresión. De nuevo es un SNP relativamente poco frecuente (MAF=0.13) y la amplitud del intervalo confianza de los OR calculados son muy amplios. No obstante, este SNP, y otros próximos en la región promotora que podrían alterar la transcripción del gen, se han asociado previamente a un mayor riesgo de cáncer gástrico tanto en población asiática como occidental (130,492-494).

Por tanto, a pesar de las limitaciones de potencia de este estudio y la baja frecuencia alélica de muchos de los SNPs analizados en genes relacionados con la infección por *H.pylori*, se

han encontrado asociaciones significativas para SNP en genes claves de la vía de señalización de *H.pylori* que concuerdan con estudios previos, en su mayoría en población asiática, y con su posible implicación en la iniciación de la progresión de las lesiones gástricas precursoras.

Como resumen final se puede decir que este trabajo no solamente nos ha permitido confirmar la hipótesis principal de este estudio que era que la variabilidad de los genes *MUC* y *TFF* de la mucosa gástrica podría asociarse al cáncer gástrico o a la evolución de las lesiones precursoras de cáncer gástrico, sino que en general hemos observado cierta concordancia entre el efecto de variantes en estos genes en la evolución de las lesiones gástricas y en el cáncer gástrico o alguno de sus subtipos anatómicos e histológicos. Además, algunas de las asociaciones identificadas confirman o están de acuerdo con las recientemente identificadas por otros grupos y, aunque todas ellas, pero muy especialmente las observadas en los análisis estratificados o por subgrupos, deberían confirmarse mediante nuevos estudios en poblaciones distintas y de mayor tamaño a las utilizadas en este trabajo, para la mayoría de estas asociaciones se encuentra una explicación biológica plausible.

Conclusiones

- Nuestra estrategia de validación y búsqueda de SNPs con frecuencias superiores a 0.01 en regiones funcionales de los genes *TFF* nos permitió detectar todas las variantes que actualmente están descritas con esa frecuencia, a excepción de un cambio sinónimo de *TFF1* (I54I, rs61735845) y un SNP de la región 3'UTR de *TFF3* (c.*15C>T, rs192423444), ambas variantes con una frecuencia inferior a 0.1.
- El análisis funcional de las regiones 3'UTR de los genes *TFF* indicó que cambios en estas secuencias podrían afectar la expresión génica. El mayor efecto se observó para el SNP c.*41G>A de *TFF1* que disminuyó significativamente la expresión de la luciferasa, y un efecto similar se observó para la variante c.*41G>A de *TFF3*. En el caso de *TFF2*, con dos polimorfismos en su región 3'UTR, el cambio de expresión fue opuesto y el mayor efecto se observa para el cambio c.*82C>T.
- Se han identificado 2 nuevas variantes poco comunes ($MAF \leq 0.01$) y aparentemente funcionales del gen *TFF3*, los SNPs c.*41G>A (rs192991660) y V80M (rs118095917), que parecen asociarse a la carcinogénesis gástrica, aunque se requieren estudios con un tamaño de muestra mayor para confirmar la asociación.
- Existe asociación genética entre variantes del gen *MUC2* y la evolución de lesiones precursoras de cáncer gástrico, especialmente en los pacientes infectados por *H.pylori*, sugiriendo una interacción con la bacteria. Además, ciertas variantes de *MUC2* también parecen asociarse con el cáncer gástrico, especialmente del tipo intestinal y del no cardia. Estos resultados sugieren una implicación de la mucina secretada en la carcinogénesis gástrica
- Se confirma que el alelo G del SNP rs4072037 del gen *MUC1* es un factor protector para cáncer gástrico y, específicamente y en mayor medida, para el tipo intestinal y probablemente, para el cardias, en población europea. Además, uno de los haplotipos que contiene el alelo A del SNP rs4072037 es un haplotipo de riesgo para cáncer gástrico, en comparación con el haplotipo de referencia, el más común, que contiene el alelo G
- Distintas variantes y haplotipos en el gen *TFF1* y regiones reguladoras adyacentes modifican el riesgo del cáncer gástrico del tipo intestinal y, probablemente, la evolución de las lesiones gástricas. Existe una gran variabilidad del gen, con distintos efectos en la carcinogénesis en función de las variantes que se combinen en el haplotipo.
- Existen indicios de una posible interacción de variantes de *TFF1* con el sexo, especialmente con los SNPs localizados en el promotor, que contiene elementos de respuesta a estrógeno.
- Aunque la relación de la variabilidad polimórfica en los genes *MUC6*, *TFF2* y *TFF3* con la carcinogénesis gástrica no es tan evidente, se han encontrado asociaciones puntuales con cáncer gástrico que indican que estos genes también podrían desempeñar un papel importante en el desarrollo de cáncer gástrico.
- Polimorfismos en genes relacionados con la vía de señalización y respuesta inmunitaria a la infección por *H.pylori*, especialmente (*TLR4*, *CDH1*, *CD14*, *PTPN11* y *TNF*), se han asociado a la evolución de lesiones gástricas, en concordancia con estudios previos, sugiriendo una posible implicación en las etapas iniciales de la carcinogénesis gástrica.
- El consumo de AINES se asocia como factor protector de la progresión de lesiones gástricas.

Bibliografia

1. Ferlay, J., *et al.* (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*, **127**, 2893-917.
2. Bertuccio, P., *et al.* (2009) Recent patterns in gastric cancer: a global overview. *Int J Cancer*, **125**, 666-73.
3. Ferlay, J., *et al.* (2010) Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer*, **46**, 765-81.
4. Aragonés, N., *et al.* (2009) The striking geographical pattern of gastric cancer mortality in Spain: environmental hypotheses revisited. *BMC Cancer*, **9**, 316.
5. García-Esquinas, E., *et al.* (2009) Gastric cancer mortality trends in Spain, 1976-2005, differences by autonomous region and sex. *BMC Cancer*, **9**, 346.
6. Chandanos, E., *et al.* (2008) Oestrogen and the enigmatic male predominance of gastric cancer. *Eur J Cancer*, **44**, 2397-403.
7. Duell, E.J., *et al.* (2010) Menstrual and reproductive factors, exogenous hormone use, and gastric cancer risk in a cohort of women from the European Prospective Investigation Into Cancer and Nutrition. *Am J Epidemiol*, **172**, 1384-93.
8. Camargo, M.C., *et al.* (2012) Sex hormones, hormonal interventions, and gastric cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*, **21**, 20-38.
9. Verdecchia, A., *et al.* (2007) Recent cancer survival in Europe: a 2000-02 period analysis of EURO-CARE-4 data. *Lancet Oncol*, **8**, 784-96.
10. Schubert, M.L., *et al.* (2008) Control of gastric acid secretion in health and disease. *Gastroenterology*, **134**, 1842-60.
11. Young, B., *et al.* (eds.) (2006) *Wheater's Functional Histology: A text and colour atlas*. Churchill Livingstone.
12. Phillipson, M., *et al.* (2008) The gastric mucus layers: constituents and regulation of accumulation. *Am J Physiol Gastrointest Liver Physiol*, **295**, G806-12.
13. Hamilton, S.R., *et al.* (2000) Fenoglio-Preiser C, Carneiro F, Correa P, Guilford P, Lambert R, Megraud F, *et al.* (2000). Gastric carcinoma. In: Hamilton SR, Aaltonen LA, editors. *Pathology & Genetics, Tumors of the Digestive System*. Lyon: IARC Press. pp. 39-52.
14. Lauren, P. (1965) The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand*, **64**, 31-49.
15. Correa, P. (1992) Human gastric carcinogenesis: a multistep and multifactorial process-- First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res*, **52**, 6735-40.
16. Nobili, S., *et al.* (2011) Genomic and genetic alterations influence the progression of gastric cancer. *World J Gastroenterol*, **17**, 290-9.
17. Lochhead, P., *et al.* (2008) Gastric cancer. *Br Med Bull*, **85**, 87-100.
18. Milne, A.N., *et al.* (2009) Nature meets nurture: molecular genetics of gastric cancer. *Hum Genet*, **126**, 615-28.
19. Machado, J.C., *et al.* (1999) E-cadherin gene mutations provide a genetic basis for the phenotypic divergence of mixed gastric carcinomas. *Lab Invest*, **79**, 459-65.
20. Humar, B., *et al.* (2009) E-cadherin deficiency initiates gastric signet-ring cell carcinoma in mice and man. *Cancer Res*, **69**, 2050-6.
21. Peek, R.M., Jr., *et al.* (2002) Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer*, **2**, 28-37.
22. Wang, H.D., *et al.* (2004) CDH1 germline mutation in hereditary gastric carcinoma. *World J Gastroenterol*, **10**, 3088-93.
23. Barber, M., *et al.* (2006) Familial gastric cancer - aetiology and pathogenesis. *Best Pract Res Clin Gastroenterol*, **20**, 721-34.

24. Guilford, P., *et al.* (1998) E-cadherin germline mutations in familial gastric cancer. *Nature*, **392**, 402-5.
25. Oliveira, C., *et al.* (2006) Genetics, pathology, and clinics of familial gastric cancer. *Int J Surg Pathol*, **14**, 21-33.
26. Humar, B., *et al.* (2009) Hereditary diffuse gastric cancer: a manifestation of lost cell polarity. *Cancer Sci*, **100**, 1151-7.
27. McColl, K.E. (2006) Cancer of the gastric cardia. *Best Pract Res Clin Gastroenterol*, **20**, 687-96.
28. Wu, X., *et al.* (2006) Incidence of esophageal and gastric carcinomas among American Asians/Pacific Islanders, whites, and blacks: subsite and histology differences. *Cancer*, **106**, 683-92.
29. Parkin, D.M. (2006) The global health burden of infection-associated cancers in the year 2002. *Int J Cancer*, **118**, 3030-44.
30. Gonzalez, C.A., *et al.* (2012) Helicobacter pylori infection assessed by ELISA and by immunoblot and noncardia gastric cancer risk in a prospective study: the Eurgast-EPIC project. *Ann Oncol*, **23**, 1320-4.
31. Hansen, S., *et al.* (2007) Two distinct aetiologies of cardia cancer; evidence from premorbid serological markers of gastric atrophy and Helicobacter pylori status. *Gut*, **56**, 918-25.
32. Abrams, J.A., *et al.* (2013) Diverging trends in the incidence of reflux-related and Helicobacter pylori-related gastric cardia cancer. *J Clin Gastroenterol*, **47**, 322-7.
33. Peek, R.M., Jr., *et al.* (2010) Role of innate immunity in Helicobacter pylori-induced gastric malignancy. *Physiol Rev*, **90**, 831-58.
34. Dixon, M.F., *et al.* (1996) Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol*, **20**, 1161-81.
35. Correa, P., *et al.* (2012) The gastric precancerous cascade. *J Dig Dis*, **13**, 2-9.
36. Konturek, P.C., *et al.* (2009) Helicobacter pylori infection in gastric cancerogenesis. *J Physiol Pharmacol*, **60**, 3-21.
37. Agreus, L., *et al.* (2012) Rationale in diagnosis and screening of atrophic gastritis with stomach-specific plasma biomarkers. *Scand J Gastroenterol*, **47**, 136-47.
38. Calmels, S., *et al.* (1991) Bacterial formation of N-nitroso compounds in the rat stomach after omeprazole-induced achlorhydria. *IARC Sci Publ*, 187-91.
39. Sanduleanu, S., *et al.* (2001) Double gastric infection with Helicobacter pylori and non-Helicobacter pylori bacteria during acid-suppressive therapy: increase of pro-inflammatory cytokines and development of atrophic gastritis. *Aliment Pharmacol Ther*, **15**, 1163-75.
40. Secretan, B., *et al.* (2009) A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol*, **10**, 1033-4.
41. Jass, J.R., *et al.* (1981) The mucin profiles of normal gastric mucosa, intestinal metaplasia and its variants and gastric carcinoma. *Histochem J*, **13**, 931-9.
42. Reis, C.A., *et al.* (1999) Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, and MUC6) expression. *Cancer Res*, **59**, 1003-7.
43. Correa, P., *et al.* (2010) Pathology of gastric intestinal metaplasia: clinical implications. *Am J Gastroenterol*, **105**, 493-8.
44. Filipe, M.I., *et al.* (1994) Intestinal metaplasia types and the risk of gastric cancer: a cohort study in Slovenia. *Int J Cancer*, **57**, 324-9.
45. Gonzalez, C.A., *et al.* (2010) Gastric cancer occurrence in preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. *Int J Cancer*, **127**, 2654-60.
46. Kang, K.P., *et al.* (2009) Role of intestinal metaplasia subtyping in the risk of gastric cancer in Korea. *J Gastroenterol Hepatol*, **24**, 140-8.
47. Gonzalez, C.A., *et al.* (2013) Utility of subtyping intestinal metaplasia as marker of gastric cancer risk. A review of the evidence. *Int J Cancer*, **133**, 1023-32.

48. Busuttill, R.A., *et al.* (2009) Intestinal metaplasia: a premalignant lesion involved in gastric carcinogenesis. *J Gastroenterol Hepatol*, **24**, 193-201.
49. Dinis-Ribeiro, M., *et al.* (2012) Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSg), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED). *Endoscopy*, **44**, 74-94.
50. Shiotani, A., *et al.* (2008) Sonic hedgehog and CDX2 expression in the stomach. *J Gastroenterol Hepatol*, **23 Suppl 2**, S161-6.
51. Noto, J.M., *et al.* (2012) Gastric-to-intestinal transdifferentiation and cancer. *Proc Natl Acad Sci U S A*, **109**, 20173-4.
52. Gutierrez-Gonzalez, L., *et al.* (2008) Biology of intestinal metaplasia in 2008: more than a simple phenotypic alteration. *Dig Liver Dis*, **40**, 510-22.
53. Schmidt, P.H., *et al.* (1999) Identification of a metaplastic cell lineage associated with human gastric adenocarcinoma. *Lab Invest*, **79**, 639-46.
54. Goldenring, J.R., *et al.* (2006) Differentiation of the gastric mucosa III. Animal models of oxyntic atrophy and metaplasia. *Am J Physiol Gastrointest Liver Physiol*, **291**, G999-1004.
55. Correa, P., *et al.* (2007) Carcinogenesis of Helicobacter pylori. *Gastroenterology*, **133**, 659-72.
56. Stolte, M. (2003) The new Vienna classification of epithelial neoplasia of the gastrointestinal tract: advantages and disadvantages. *Virchows Arch*, **442**, 99-106.
57. Lauwers, G.Y., *et al.* (2007) Gastric preneoplastic lesions and epithelial dysplasia. *Gastroenterol Clin North Am*, **36**, 813-29, vi.
58. Farinati, F., *et al.* (2008) Helicobacter pylori, inflammation, oxidative damage and gastric cancer: a morphological, biological and molecular pathway. *Eur J Cancer Prev*, **17**, 195-200.
59. Marshall, B.J., *et al.* (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*, **1**, 1311-5.
60. Humans, I.W.G.o.t.E.o.C.R.t. (1994) Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. *IARC Monogr Eval Carcinog Risks Hum*, **61**, 1-241.
61. Eslick, G.D. (2006) Helicobacter pylori infection causes gastric cancer? A review of the epidemiological, meta-analytic, and experimental evidence. *World J Gastroenterol*, **12**, 2991-9.
62. Goh, K.L., *et al.* (2011) Epidemiology of Helicobacter pylori infection and public health implications. *Helicobacter*, **16 Suppl 1**, 1-9.
63. Laine, L., *et al.* (2008) Gastric mucosal defense and cytoprotection: bench to bedside. *Gastroenterology*, **135**, 41-60.
64. Weeks, D.L., *et al.* (2000) A H⁺-gated urea channel: the link between Helicobacter pylori urease and gastric colonization. *Science*, **287**, 482-5.
65. Zanotti, G., *et al.* (2010) Functional and structural aspects of Helicobacter pylori acidic stress response factors. *IUBMB Life*, **62**, 715-23.
66. Backert, S., *et al.* (2011) Molecular mechanisms of gastric epithelial cell adhesion and injection of CagA by Helicobacter pylori. *Cell Commun Signal*, **9**, 28.
67. Ilver, D., *et al.* (1998) Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science*, **279**, 373-7.
68. Rad, R., *et al.* (2002) The Helicobacter pylori blood group antigen-binding adhesin facilitates bacterial colonization and augments a nonspecific immune response. *J Immunol*, **168**, 3033-41.
69. Mahdavi, J., *et al.* (2002) Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science*, **297**, 573-8.
70. Tomb, J.F., *et al.* (1997) The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature*, **388**, 539-47.

71. Israel, D.A., *et al.* (2001) Helicobacter pylori genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A*, **98**, 14625-30.
72. Atherton, J.C., *et al.* (2009) Coadaptation of Helicobacter pylori and humans: ancient history, modern implications. *J Clin Invest*, **119**, 2475-87.
73. Censini, S., *et al.* (1996) cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A*, **93**, 14648-53.
74. Palli, D., *et al.* (2007) CagA+ Helicobacter pylori infection and gastric cancer risk in the EPIC-EURGAST study. *Int J Cancer*, **120**, 859-67.
75. Wroblewski, L.E., *et al.* (2010) Helicobacter pylori and gastric cancer: factors that modulate disease risk. *Clin Microbiol Rev*, **23**, 713-39.
76. Odenbreit, S., *et al.* (2000) Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion. *Science*, **287**, 1497-500.
77. Stein, M., *et al.* (2002) c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol*, **43**, 971-80.
78. Higashi, H., *et al.* (2002) SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein. *Science*, **295**, 683-6.
79. Higashi, H., *et al.* (2004) Helicobacter pylori CagA induces Ras-independent morphogenetic response through SHP-2 recruitment and activation. *J Biol Chem*, **279**, 17205-16.
80. Tammer, I., *et al.* (2007) Activation of Abl by Helicobacter pylori: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology*, **132**, 1309-19.
81. Amieva, M.R., *et al.* (2003) Disruption of the epithelial apical-junctional complex by Helicobacter pylori CagA. *Science*, **300**, 1430-4.
82. Franco, A.T., *et al.* (2005) Activation of beta-catenin by carcinogenic Helicobacter pylori. *Proc Natl Acad Sci U S A*, **102**, 10646-51.
83. Murata-Kamiya, N., *et al.* (2007) Helicobacter pylori CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene*, **26**, 4617-26.
84. Brandt, S., *et al.* (2005) NF-kappaB activation and potentiation of proinflammatory responses by the Helicobacter pylori CagA protein. *Proc Natl Acad Sci U S A*, **102**, 9300-5.
85. Mimuro, H., *et al.* (2007) Helicobacter pylori dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance colonization of the stomach. *Cell Host Microbe*, **2**, 250-63.
86. Buti, L., *et al.* (2011) Helicobacter pylori cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. *Proc Natl Acad Sci U S A*, **108**, 9238-43.
87. Ohnishi, N., *et al.* (2008) Transgenic expression of Helicobacter pylori CagA induces gastrointestinal and hematopoietic neoplasms in mouse. *Proc Natl Acad Sci U S A*, **105**, 1003-8.
88. Cover, T.L., *et al.* (1992) Purification and characterization of the vacuolating toxin from Helicobacter pylori. *J Biol Chem*, **267**, 10570-5.
89. Telford, J.L., *et al.* (1994) Gene structure of the Helicobacter pylori cytotoxin and evidence of its key role in gastric disease. *J Exp Med*, **179**, 1653-58.
90. Cover, T.L., *et al.* (2005) Helicobacter pylori VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol*, **3**, 320-32.
91. Ricci, V., *et al.* (2011) Molecular cross-talk between Helicobacter pylori and human gastric mucosa. *World J Gastroenterol*, **17**, 1383-99.
92. Cover, T.L., *et al.* (2003) Induction of gastric epithelial cell apoptosis by Helicobacter pylori vacuolating cytotoxin. *Cancer Res*, **63**, 951-7.
93. Jain, P., *et al.* (2011) Helicobacter pylori vacuolating cytotoxin A (VacA) engages the mitochondrial fission machinery to induce host cell death. *Proc Natl Acad Sci U S A*, **108**, 16032-7.

94. Torres, V.J., *et al.* (2007) Helicobacter pylori vacuolating cytotoxin inhibits activation-induced proliferation of human T and B lymphocyte subsets. *J Immunol*, **179**, 5433-40.
95. Wang, F., *et al.* (2008) Helicobacter pylori VacA disrupts apical membrane-cytoskeletal interactions in gastric parietal cells. *J Biol Chem*, **283**, 26714-25.
96. Atherton, J.C., *et al.* (1995) Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration. *J Biol Chem*, **270**, 17771-7.
97. Rhead, J.L., *et al.* (2007) A new Helicobacter pylori vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology*, **133**, 926-36.
98. Wen, S., *et al.* (2009) Helicobacter pylori virulence factors in gastric carcinogenesis. *Cancer Lett*, **282**, 1-8.
99. Strober, W., *et al.* (2006) Signalling pathways and molecular interactions of NOD1 and NOD2. *Nat Rev Immunol*, **6**, 9-20.
100. Fukata, M., *et al.* (2008) Role of Toll-like receptors in gastrointestinal malignancies. *Oncogene*, **27**, 234-43.
101. Shi, Y., *et al.* (2010) Helicobacter pylori-induced Th17 responses modulate Th1 cell responses, benefit bacterial growth, and contribute to pathology in mice. *J Immunol*, **184**, 5121-9.
102. Watson, S.A., *et al.* (2006) Gastrin - active participant or bystander in gastric carcinogenesis? *Nat Rev Cancer*, **6**, 936-46.
103. El-Omar, E.M. (2001) The importance of interleukin 1beta in Helicobacter pylori associated disease. *Gut*, **48**, 743-7.
104. Caruso, R., *et al.* (2007) IL-21 is highly produced in Helicobacter pylori-infected gastric mucosa and promotes gelatinases synthesis. *J Immunol*, **178**, 5957-65.
105. Caputo, R., *et al.* (2003) Helicobacter pylori VacA toxin up-regulates vascular endothelial growth factor expression in MKN 28 gastric cells through an epidermal growth factor receptor-, cyclooxygenase-2-dependent mechanism. *Clin Cancer Res*, **9**, 2015-21.
106. Strowski, M.Z., *et al.* (2004) Helicobacter pylori stimulates host vascular endothelial growth factor-A (vegf-A) gene expression via MEK/ERK-dependent activation of Sp1 and Sp3. *FASEB J*, **18**, 218-20.
107. Gonzalez, C.A., *et al.* (2003) Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *Int J Cancer*, **107**, 629-34.
108. Tramacere, I., *et al.* (2011) Tobacco smoking and esophageal and gastric cardia adenocarcinoma: a meta-analysis. *Epidemiology*, **22**, 344-9.
109. Duell, E.J., *et al.* (2011) Alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Am J Clin Nutr*, **94**, 1266-75.
110. Tramacere, I., *et al.* (2012) A meta-analysis on alcohol drinking and gastric cancer risk. *Ann Oncol*, **23**, 28-36.
111. Renehan, A.G., *et al.* (2008) Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*, **371**, 569-78.
112. Yang, P., *et al.* (2009) Overweight, obesity and gastric cancer risk: results from a meta-analysis of cohort studies. *Eur J Cancer*, **45**, 2867-73.
113. Riboli, E., *et al.* (2003) Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *Am J Clin Nutr*, **78**, 559S-569S.
114. Gonzalez, C.A., *et al.* (2006) Fruit and vegetable intake and the risk of stomach and oesophagus adenocarcinoma in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Int J Cancer*, **118**, 2559-66.
115. Lunet, N., *et al.* (2007) Fruit and vegetable consumption and gastric cancer by location and histological type: case-control and meta-analysis. *Eur J Cancer Prev*, **16**, 312-27.
116. Soerjomataram, I., *et al.* (2010) Increased consumption of fruit and vegetables and future cancer incidence in selected European countries. *Eur J Cancer*, **46**, 2563-80.

117. Wang, X.Q., *et al.* (2009) Review of salt consumption and stomach cancer risk: epidemiological and biological evidence. *World J Gastroenterol*, **15**, 2204-13.
118. Gonzalez, C.A., *et al.* (2006) Meat intake and risk of stomach and esophageal adenocarcinoma within the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst*, **98**, 345-54.
119. Brown, L.M. (2000) Helicobacter pylori: epidemiology and routes of transmission. *Epidemiol Rev*, **22**, 283-97.
120. Lichtenstein, P., *et al.* (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*, **343**, 78-85.
121. International HapMap, C. (2003) The International HapMap Project. *Nature*, **426**, 789-96.
122. International HapMap, C. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299-320.
123. International HapMap, C., *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851-61.
124. International HapMap, C., *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52-8.
125. Gonzalez, C.A., *et al.* (2002) Genetic susceptibility and gastric cancer risk. *Int J Cancer*, **100**, 249-60.
126. El-Omar, E.M. (2006) Role of host genes in sporadic gastric cancer. *Best Pract Res Clin Gastroenterol*, **20**, 675-86.
127. El-Omar, E.M., *et al.* (2000) Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature*, **404**, 398-402.
128. Crusius, J.B., *et al.* (2008) Cytokine gene polymorphisms and the risk of adenocarcinoma of the stomach in the European prospective investigation into cancer and nutrition (EPIC-EURGAST). *Ann Oncol*, **19**, 1894-902.
129. Schneider, B.G., *et al.* (2008) Cytokine polymorphisms and gastric cancer risk: an evolving view. *Cancer Biol Ther*, **7**, 157-62.
130. Gorouhi, F., *et al.* (2008) Tumour-necrosis factor-A polymorphisms and gastric cancer risk: a meta-analysis. *Br J Cancer*, **98**, 1443-51.
131. Vincenzi, B., *et al.* (2008) Interleukin 1beta-511T gene (IL1beta) polymorphism is correlated with gastric cancer in the Caucasian population: results from a meta-analysis. *Oncol Rep*, **20**, 1213-20.
132. Zhang, J., *et al.* (2008) Polymorphisms of tumor necrosis factor-alpha are associated with increased susceptibility to gastric cancer: a meta-analysis. *J Hum Genet*, **53**, 479-89.
133. Xue, H., *et al.* (2010) Interleukin-1B and interleukin-1 RN polymorphisms and gastric carcinoma risk: a meta-analysis. *J Gastroenterol Hepatol*, **25**, 1604-17.
134. Persson, C., *et al.* (2011) Polymorphisms in inflammatory response genes and their association with gastric cancer: A HuGE systematic review and meta-analyses. *Am J Epidemiol*, **173**, 259-70.
135. Peleteiro, B., *et al.* (2010) Association between cytokine gene polymorphisms and gastric precancerous lesions: systematic review and meta-analysis. *Cancer Epidemiol Biomarkers Prev*, **19**, 762-76.
136. Capella, G., *et al.* (2008) DNA repair polymorphisms and the risk of stomach adenocarcinoma and severe chronic gastritis in the EPIC-EURGAST study. *Int J Epidemiol*, **37**, 1316-25.
137. Geng, J., *et al.* (2008) XRCC1 genetic polymorphism Arg399Gln and gastric cancer risk: A meta-analysis. *World J Gastroenterol*, **14**, 6733-7.
138. Xue, H., *et al.* (2011) X-ray repair cross-complementing group 1 (XRCC1) genetic polymorphisms and gastric cancer risk: A HuGE review and meta-analysis. *Am J Epidemiol*, **173**, 363-75.

139. Chen, B., *et al.* (2011) ERCC2 Lys751Gln and Asp312Asn polymorphisms and gastric cancer risk: a meta-analysis. *J Cancer Res Clin Oncol*, **137**, 939-46.
140. Hold, G.L., *et al.* (2007) A functional polymorphism of toll-like receptor 4 gene increases risk of gastric carcinoma and its precursors. *Gastroenterology*, **132**, 905-12.
141. Achyut, B.R., *et al.* (2007) Association of Toll-like receptor-4 (Asp299Gly and Thr399Ileu) gene polymorphisms with gastritis and precancerous lesions. *Hum Immunol*, **68**, 901-7.
142. Santini, D., *et al.* (2008) Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms in gastric cancer of intestinal and diffuse histotypes. *Clin Exp Immunol*, **154**, 360-4.
143. El-Omar, E.M., *et al.* (2008) Polymorphisms in Toll-like receptor genes and risk of cancer. *Oncogene*, **27**, 244-52.
144. Kutikhin, A.G. (2011) Impact of Toll-like receptor 4 polymorphisms on risk of cancer. *Hum Immunol*, **72**, 193-206.
145. Rosenstiel, P., *et al.* (2006) Influence of polymorphisms in the NOD1/CARD4 and NOD2/CARD15 genes on the clinical outcome of Helicobacter pylori infection. *Cell Microbiol*, **8**, 1188-98.
146. Kutikhin, A.G. (2011) Role of NOD1/CARD4 and NOD2/CARD15 gene polymorphisms in cancer etiology. *Hum Immunol*, **72**, 955-68.
147. Companioni, O., *et al.* (2013) Polymorphisms of Helicobacter pylori signaling pathway genes and gastric cancer risk in the European prospective investigation into cancer-eurgast cohort. *Int J Cancer*.
148. Agudo, A., *et al.* (2006) Polymorphisms in metabolic genes related to tobacco smoke and the risk of gastric cancer in the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev*, **15**, 2427-34.
149. Piao, J.M., *et al.* (2009) Glutathione-S-transferase (GSTM1, GSTT1) and the risk of gastrointestinal cancer in a Korean population. *World J Gastroenterol*, **15**, 5716-21.
150. Tripathi, S., *et al.* (2008) Gastric carcinogenesis: Possible role of polymorphisms of GSTM1, GSTT1, and GSTP1 genes. *Scand J Gastroenterol*, **43**, 431-9.
151. Garcia-Gonzalez, M.A., *et al.* (2012) Relevance of GSTM1, GSTT1, and GSTP1 gene polymorphisms to gastric cancer susceptibility and phenotype. *Mutagenesis*, **27**, 771-7.
152. Ma, W., *et al.* (2013) Association between glutathione S-transferase T1 null genotype and gastric cancer risk: a meta-analysis of 48 studies. *PLoS One*, **8**, e60833.
153. Sakamoto, H., *et al.* (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet*, **40**, 730-40.
154. Wang, L.D., *et al.* (2010) Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat Genet*, **42**, 759-63.
155. Abnet, C.C., *et al.* (2010) A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet*, **42**, 764-7.
156. Shi, Y., *et al.* (2011) A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet*, **43**, 1215-8.
157. Hirschhorn, J.N., *et al.* (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, **6**, 95-108.
158. McCarthy, M.I., *et al.* (2008) Genome-wide association studies: past, present and future. *Hum Mol Genet*, **17**, R100-1.
159. Matsuo, K., *et al.* (2009) Association of prostate stem cell antigen gene polymorphisms with the risk of stomach cancer in Japanese. *Int J Cancer*, **125**, 1961-4.
160. Wu, C., *et al.* (2009) Two genetic variants in prostate stem cell antigen and gastric cancer susceptibility in a Chinese population. *Mol Carcinog*, **48**, 1131-8.
161. Ou, J., *et al.* (2010) Association and haplotype analysis of prostate stem cell antigen with gastric cancer in Tibetans. *DNA Cell Biol*, **29**, 319-23.

162. Lu, Y., *et al.* (2010) Genetic variation of PSCA gene is associated with the risk of both diffuse- and intestinal-type gastric cancer in a Chinese population. *Int J Cancer*, **127**, 2183-9.
163. Song, H.R., *et al.* (2011) Association of a common genetic variant in prostate stem-cell antigen with gastric cancer susceptibility in a Korean population. *Mol Carcinog*, **50**, 871-5.
164. Lochhead, P., *et al.* (2011) Genetic variation in the prostate stem cell antigen gene and upper gastrointestinal cancer in white individuals. *Gastroenterology*, **140**, 435-41.
165. Sala, N., *et al.* (2012) Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: Results from the EPIC-EURGAST study. *Int J Cancer*, **130**, 2417-27.
166. Saeki, N., *et al.* (2011) A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology*, **140**, 892-902.
167. Ng, W., *et al.* (2008) Genetic regulation of MUC1 alternative splicing in human tissues. *Br J Cancer*, **99**, 978-85.
168. Palmer, A.J., *et al.* (2012) Genetic variation in C20orf54, PLCE1 and MUC1 and the risk of upper gastrointestinal cancers in Caucasian populations. *Eur J Cancer Prev*, **21**, 541-4.
169. Wang, M., *et al.* (2012) Potentially functional variants of PLCE1 identified by GWASs contribute to gastric adenocarcinoma susceptibility in an eastern Chinese population. *PLoS One*, **7**, e31932.
170. Song, H.R., *et al.* (2013) Genetic variations in the PRKAA1 and ZBTB20 genes and gastric cancer susceptibility in a Korean population. *Mol Carcinog*.
171. Du, J., *et al.* (2013) Genetic variants at 5p15 are associated with risk and early-onset of gastric cancer in Chinese populations. *Carcinogenesis*.
172. Thim, L., *et al.* (2005) Structure of mammalian trefoil factors and functional insights. *Cell Mol Life Sci*, **62**, 2956-73.
173. Taupin, D., *et al.* (2003) Trefoil factors: initiators of mucosal healing. *Nat Rev Mol Cell Biol*, **4**, 721-32.
174. May, F.E., *et al.* (2003) The closely related estrogen-regulated trefoil proteins TFF1 and TFF3 have markedly different hydrodynamic properties, overall charge, and distribution of surface charge. *Biochemistry*, **42**, 8250-9.
175. Thim, L. (1997) Trefoil peptides: from structure to function. *Cell Mol Life Sci*, **53**, 888-903.
176. May, F.E., *et al.* (2009) The trefoil factor interacting protein TFIZ1 binds the trefoil protein TFF1 preferentially in normal gastric mucosal cells but the co-expression of these proteins is deregulated in gastric cancer. *Int J Biochem Cell Biol*, **41**, 632-40.
177. Kjellek, S. (2009) The trefoil factor family - small peptides with multiple functionalities. *Cell Mol Life Sci*, **66**, 1350-69.
178. Seib, T., *et al.* (1997) The three human trefoil genes TFF1, TFF2, and TFF3 are located within a region of 55 kb on chromosome 21q22.3. *Genomics*, **40**, 200-2.
179. Gott, P., *et al.* (1996) Human trefoil peptides: genomic structure in 21q22.3 and coordinated expression. *Eur J Hum Genet*, **4**, 308-15.
180. Madsen, J., *et al.* (2007) Tissue localization of human trefoil factors 1, 2, and 3. *J Histochem Cytochem*, **55**, 505-13.
181. Longman, R.J., *et al.* (2000) Coordinated localisation of mucins and trefoil peptides in the ulcer associated cell lineage and the gastrointestinal mucosa. *Gut*, **47**, 792-800.
182. Hanby, A.M., *et al.* (1993) Spasmolytic polypeptide is a major antral peptide: distribution of the trefoil peptides human spasmolytic polypeptide and pS2 in the stomach. *Gastroenterology*, **105**, 1110-6.
183. Tomasetto, C., *et al.* (1990) hSP, the domain-duplicated homolog of pS2 protein, is co-expressed with pS2 in stomach but not in breast carcinoma. *EMBO J*, **9**, 407-14.

184. Podolsky, D.K., *et al.* (1993) Identification of human intestinal trefoil factor. Goblet cell-specific expression of a peptide targeted for apical secretion. *J Biol Chem*, **268**, 6694-702.
185. Kouznetsova, I., *et al.* (2004) A gradient of TFF3 (trefoil factor family 3) peptide synthesis within the normal human gastric mucosa. *Cell Tissue Res*, **316**, 155-65.
186. Nunez, A.M., *et al.* (1989) The 5' flanking region of the pS2 gene contains a complex enhancer region responsive to oestrogens, epidermal growth factor, a tumour promoter (TPA), the c-Ha-ras oncoprotein and the c-jun protein. *EMBO J*, **8**, 823-9.
187. Beck, S., *et al.* (1999) Hepatocyte nuclear factor 3 (winged helix domain) activates trefoil factor gene TFF1 through a binding motif adjacent to the TATAA box. *DNA Cell Biol*, **18**, 157-64.
188. Al-zazeh, E.D., *et al.* (2000) Transcription factor GATA-6 activates expression of gastroprotective trefoil genes TFF1 and TFF2. *Biochim Biophys Acta*, **1490**, 324-32.
189. Sun, J.M., *et al.* (2005) Estrogen regulation of trefoil factor 1 expression by estrogen receptor alpha and Sp proteins. *Exp Cell Res*, **302**, 96-107.
190. Baus-Loncar, M., *et al.* (2005) Multiple regulatory pathways for trefoil factor (TFF) genes. *Cell Mol Life Sci*, **62**, 2921-31.
191. Bulitta, C.J., *et al.* (2002) Autoinduction of the trefoil factor 2 (TFF2) promoter requires an upstream cis-acting element. *Biochem Biophys Res Commun*, **293**, 366-74.
192. Chi, A.L., *et al.* (2004) Characterization of a CCAAT-enhancer element of trefoil factor family 2 (TFF2) promoter in MCF-7 cells. *Peptides*, **25**, 839-47.
193. Sands, B.E., *et al.* (1995) Molecular cloning of the rat intestinal trefoil factor gene. Characterization of an intestinal goblet cell-associated promoter. *J Biol Chem*, **270**, 9353-61.
194. Ogata, H., *et al.* (1998) Identification of a goblet cell-specific enhancer element in the rat intestinal trefoil factor gene promoter bound by a goblet cell nuclear protein. *J Biol Chem*, **273**, 3060-7.
195. Itoh, H., *et al.* (1999) Goblet-cell-specific transcription of mouse intestinal trefoil factor gene results from collaboration of complex series of positive and negative regulatory elements. *Biochem J*, **341 (Pt 2)**, 461-72.
196. Iwakiri, D., *et al.* (2001) Keratinocyte growth factor promotes goblet cell differentiation through regulation of goblet cell silencer inhibitor. *Gastroenterology*, **120**, 1372-80.
197. Dossinger, V., *et al.* (2002) Down-regulation of TFF expression in gastrointestinal cell lines by cytokines and nuclear factors. *Cell Physiol Biochem*, **12**, 197-206.
198. Tebbutt, N.C., *et al.* (2002) Reciprocal regulation of gastrointestinal homeostasis by SHP2 and STAT-mediated trefoil gene activation in gp130 mutant mice. *Nat Med*, **8**, 1089-97.
199. Loncar, M.B., *et al.* (2003) Tumour necrosis factor alpha and nuclear factor kappaB inhibit transcription of human TFF3 encoding a gastrointestinal healing peptide. *Gut*, **52**, 1297-303.
200. Baus-Loncar, M., *et al.* (2004) Transcriptional control of TFF3 (intestinal trefoil factor) via promoter binding sites for the nuclear factor kappaB and C/EBPbeta. *Peptides*, **25**, 849-54.
201. Koike, T., *et al.* (2007) Up-regulation of TFF1 (pS2) expression by TNF-alpha in gastric epithelial cells. *J Gastroenterol Hepatol*, **22**, 936-42.
202. Blanchard, C., *et al.* (2004) IL-4 and IL-13 up-regulate intestinal trefoil factor expression: requirement for STAT6 and de novo protein synthesis. *J Immunol*, **172**, 3775-83.
203. Khan, Z.E., *et al.* (2003) Transcriptional regulation of the human trefoil factor, TFF1, by gastrin. *Gastroenterology*, **125**, 510-21.
204. Tu, S., *et al.* (2007) Gastrin regulates the TFF2 promoter through gastrin-responsive cis-acting elements and multiple signaling pathways. *Am J Physiol Gastrointest Liver Physiol*, **292**, G1726-37.
205. Ribieras, S., *et al.* (2001) Mouse Trefoil factor genes: genomic organization, sequences and methylation analyses. *Gene*, **266**, 67-75.

206. Giraud, A.S., *et al.* (2007) Differentiation of the Gastric Mucosa IV. Role of trefoil peptides and IL-6 cytokine family signaling in gastric homeostasis. *Am J Physiol Gastrointest Liver Physiol*, **292**, G1-5.
207. Lefebvre, O., *et al.* (1996) Gastric mucosa abnormalities and tumorigenesis in mice lacking the pS2 trefoil protein. *Science*, **274**, 259-62.
208. Farrell, J.J., *et al.* (2002) TFF2/SP-deficient mice show decreased gastric proliferation, increased acid secretion, and increased susceptibility to NSAID injury. *J Clin Invest*, **109**, 193-204.
209. Baus-Loncar, M., *et al.* (2005) Trefoil factor family 2 deficiency and immune response. *Cell Mol Life Sci*, **62**, 2947-55.
210. Baus-Loncar, M., *et al.* (2005) Trefoil factor 2 (TFF2) deficiency in murine digestive tract influences the immune system. *Cell Physiol Biochem*, **16**, 31-42.
211. Mashimo, H., *et al.* (1996) Impaired defense of intestinal mucosa in mice lacking intestinal trefoil factor. *Science*, **274**, 262-5.
212. Bossenmeyer-Pourie, C., *et al.* (2002) The trefoil factor 1 participates in gastrointestinal cell differentiation by delaying G1-S phase transition and reducing apoptosis. *J Cell Biol*, **157**, 761-70.
213. Uchino, H., *et al.* (2000) Overexpression of intestinal trefoil factor in human colon carcinoma cells reduces cellular growth in vitro and in vivo. *Gastroenterology*, **118**, 60-9.
214. Amiry, N., *et al.* (2009) Trefoil factor-1 (TFF1) enhances oncogenicity of mammary carcinoma cells. *Endocrinology*, **150**, 4473-83.
215. Kannan, N., *et al.* (2010) Trefoil factor 3 is oncogenic and mediates anti-estrogen resistance in human mammary carcinoma. *Neoplasia*, **12**, 1041-53.
216. Tomasetto, C., *et al.* (2005) Pleiotropic effects of Trefoil Factor 1 deficiency. *Cell Mol Life Sci*, **62**, 2916-20.
217. Siu, L.S., *et al.* (2004) TFF2 (trefoil family factor2) inhibits apoptosis in breast and colorectal cancer cell lines. *Peptides*, **25**, 855-63.
218. Frisch, S.M., *et al.* (2001) Anoikis mechanisms. *Curr Opin Cell Biol*, **13**, 555-62.
219. Chen, Y.H., *et al.* (2000) Transcription factor NF-kappaB signals antianoikic function of trefoil factor 3 on intestinal epithelial cells. *Biochem Biophys Res Commun*, **274**, 576-82.
220. Kinoshita, K., *et al.* (2000) Distinct pathways of cell migration and antiapoptotic response to epithelial injury: structure-function analysis of human intestinal trefoil factor. *Mol Cell Biol*, **20**, 4680-90.
221. Taupin, D.R., *et al.* (2000) Intestinal trefoil factor confers colonic epithelial resistance to apoptosis. *Proc Natl Acad Sci U S A*, **97**, 799-804.
222. Dignass, A., *et al.* (1994) Trefoil peptides promote epithelial migration through a transforming growth factor beta-independent pathway. *J Clin Invest*, **94**, 376-83.
223. Efstathiou, J.A., *et al.* (1999) Mutated epithelial cadherin is associated with increased tumorigenicity and loss of adhesion and of responsiveness to the motogenic trefoil factor 2 in colon carcinoma cells. *Proc Natl Acad Sci U S A*, **96**, 2316-21.
224. Meyer zum Buschenfelde, D., *et al.* (2004) Molecular mechanisms involved in TFF3 peptide-mediated modulation of the E-cadherin/catenin cell adhesion complex. *Peptides*, **25**, 873-83.
225. FitzGerald, A.J., *et al.* (2004) Synergistic effects of systemic trefoil factor family 1 (TFF1) peptide and epidermal growth factor in a rat model of colitis. *Peptides*, **25**, 793-801.
226. Rodrigues, S., *et al.* (2003) Trefoil peptides as proangiogenic factors in vivo and in vitro: implication of cyclooxygenase-2 and EGF receptor signaling. *FASEB J*, **17**, 7-16.
227. Hernandez, C., *et al.* (2009) Induction of trefoil factor (TFF)1, TFF2 and TFF3 by hypoxia is mediated by hypoxia inducible factor-1: implications for gastric mucosal healing. *Br J Pharmacol*, **156**, 262-72.
228. Giraud, A.S., *et al.* (2004) TFF-2 inhibits iNOS/NO in monocytes, and nitrated protein in healing colon after colitis. *Peptides*, **25**, 803-9.

229. Tan, X.D., *et al.* (1999) Intestinal trefoil factor binds to intestinal epithelial cells and induces nitric oxide production: priming and enhancing effects of mucin. *Biochem J*, **338** (Pt 3), 745-51.
230. Iizuka, M., *et al.* (2011) Wound healing of intestinal epithelial cells. *World J Gastroenterol*, **17**, 2161-71.
231. Tan, X.D., *et al.* (2000) Prostanoids mediate the protective effect of trefoil factor 3 in oxidant-induced intestinal epithelial cell injury: role of cyclooxygenase-2. *J Cell Sci*, **113** (Pt 12), 2149-55.
232. Tanaka, S., *et al.* (1997) Human spasmolytic polypeptide decreases proton permeation through gastric mucus in vivo and in vitro. *Am J Physiol*, **272**, G1473-80.
233. Thim, L., *et al.* (2002) Effect of trefoil factors on the viscoelastic properties of mucus gels. *Eur J Clin Invest*, **32**, 519-27.
234. Machado, J.C., *et al.* (2000) Gastric carcinoma exhibits distinct types of cell differentiation: an immunohistochemical study of trefoil peptides (TFF1 and TFF2) and mucins (MUC1, MUC2, MUC5AC, and MUC6). *J Pathol*, **190**, 437-43.
235. Emami, S., *et al.* (2004) Trefoil factor family (TFF) peptides and cancer progression. *Peptides*, **25**, 885-98.
236. Dhar, D.K., *et al.* (2005) Expression of trefoil factor family members correlates with patient prognosis and neoangiogenesis. *Clin Cancer Res*, **11**, 6472-8.
237. Matsuda, K., *et al.* (2008) Quantitative analysis of the effect of *Helicobacter pylori* on the expressions of SOX2, CDX2, MUC2, MUC5AC, MUC6, TFF1, TFF2, and TFF3 mRNAs in human gastric carcinoma cells. *Scand J Gastroenterol*, **43**, 25-33.
238. Taupin, D., *et al.* (2001) Augmented intestinal trefoil factor (TFF3) and loss of pS2 (TFF1) expression precedes metaplastic differentiation of gastric epithelium. *Lab Invest*, **81**, 397-408.
239. Beckler, A.D., *et al.* (2003) Decreased abundance of trefoil factor 1 transcript in the majority of gastric carcinomas. *Cancer*, **98**, 2184-91.
240. Tomita, H., *et al.* (2011) Inhibition of gastric carcinogenesis by the hormone gastrin is mediated by suppression of TFF1 epigenetic silencing. *Gastroenterology*, **140**, 879-91.
241. Im, S., *et al.* (2013) Reduced expression of TFF1 and increased expression of TFF3 in gastric cancer: correlation with clinicopathological parameters and prognosis. *Int J Med Sci*, **10**, 133-40.
242. Soutto, M., *et al.* (2011) Loss of TFF1 is associated with activation of NF-kappaB-mediated inflammation and gastric neoplasia in mice and humans. *J Clin Invest*, **121**, 1753-67.
243. Cobler, L., *et al.* (2013) Activation of the NF-kB pathway downregulates TFF-1 in gastric carcinogenesis. *Virchows Arch*.
244. Kirikoshi, H., *et al.* (2002) Expression of TFF1, TFF2 and TFF3 in gastric cancer. *Int J Oncol*, **21**, 655-9.
245. Leung, W.K., *et al.* (2002) Expression of trefoil peptides (TFF1, TFF2, and TFF3) in gastric carcinomas, intestinal metaplasia, and non-neoplastic gastric tissues. *J Pathol*, **197**, 582-8.
246. Shi, S.Q., *et al.* (2006) Expression of trefoil factors 1 and 2 in precancerous condition and gastric cancer. *World J Gastroenterol*, **12**, 3119-22.
247. Yamachika, T., *et al.* (2002) Intestinal trefoil factor: a marker of poor prognosis in gastric carcinoma. *Clin Cancer Res*, **8**, 1092-9.
248. Aikou, S., *et al.* (2011) Tests for serum levels of trefoil factor family proteins can improve gastric cancer screening. *Gastroenterology*, **141**, 837-845 e1-7.
249. Kaise, M., *et al.* (2011) The combination of serum trefoil factor 3 and pepsinogen testing is a valid non-endoscopic biomarker for predicting the presence of gastric cancer: a new marker for gastric cancer risk. *J Gastroenterol*, **46**, 736-45.
250. Park, W.S., *et al.* (2000) Somatic mutations of the trefoil factor family 1 gene in gastric cancer. *Gastroenterology*, **119**, 691-8.

251. Carvalho, R., *et al.* (2002) Loss of heterozygosity and promoter methylation, but not mutation, may underlie loss of TFF1 in gastric carcinoma. *Lab Invest*, **82**, 1319-26.
252. Tanaka, T., *et al.* (2013) Loss of trefoil factor 1 is regulated by DNA methylation and is an independent predictive factor for poor survival in advanced gastric cancer. *Int J Oncol*, **42**, 894-902.
253. Peterson, A.J., *et al.* (2010) Helicobacter pylori infection promotes methylation and silencing of trefoil factor 2, leading to gastric tumor development in mice and humans. *Gastroenterology*, **139**, 2005-17.
254. Tu, S.P., *et al.* (2009) p53 inhibition of AP1-dependent TFF2 expression induces apoptosis and inhibits cell migration in gastric cancer cells. *Am J Physiol Gastrointest Liver Physiol*, **297**, G385-96.
255. Guleng, B., *et al.* (2012) TFF3 mediated induction of VEGF via hypoxia in human gastric cancer SGC-7901 cells. *Mol Biol Rep*, **39**, 4127-34.
256. Zheng, Q., *et al.* (2013) Trefoil factor 3 peptide regulates migration via a Twist-dependent pathway in gastric cell. *Biochem Biophys Res Commun*, **438**, 6-12.
257. Clyne, M., *et al.* (2004) Helicobacter pylori interacts with the human single-domain trefoil protein TFF1. *Proc Natl Acad Sci U S A*, **101**, 7409-14.
258. Dolan, B., *et al.* (2012) The interaction of Helicobacter pylori with the adherent mucus gel layer secreted by polarized HT29-MTX-E12 cells. *PLoS One*, **7**, e47300.
259. dos Santos Silva, E., *et al.* (1999) Variable distribution of TFF2 (Spasmodysin) alleles in Europeans does not indicate predisposition to gastric cancer. *Hum Hered*, **49**, 45-7.
260. Moghanibashi, M., *et al.* (2012) Polymorphism of estrogen response element in TFF1 gene promoter is associated with an increased susceptibility to gastric cancer. *Gene*, **492**, 100-3.
261. Xu, Q., *et al.* (2013) Promoter polymorphisms in trefoil factor 2 and trefoil factor 3 genes and susceptibility to gastric cancer and atrophic gastritis among Chinese population. *Gene*, **529**, 104-12.
262. Andrianifahanana, M., *et al.* (2006) Regulation of mucin expression: mechanistic aspects and implications for cancer and inflammatory diseases. *Biochim Biophys Acta*, **1765**, 189-222.
263. Filipe, M.I. (1979) Mucins in the human gastrointestinal epithelium: a review. *Invest Cell Pathol*, **2**, 195-216.
264. Singh, P.K., *et al.* (2006) Cell surface-associated mucins in signal transduction. *Trends Cell Biol*, **16**, 467-76.
265. Linden, S.K., *et al.* (2008) Mucins in the mucosal barrier to infection. *Mucosal Immunol*, **1**, 183-97.
266. Gendler, S.J., *et al.* (1995) Epithelial mucin genes. *Annu Rev Physiol*, **57**, 607-34.
267. Parry, S., *et al.* (2001) Identification of MUC1 proteolytic cleavage sites in vivo. *Biochem Biophys Res Commun*, **283**, 715-20.
268. Macao, B., *et al.* (2006) Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nat Struct Mol Biol*, **13**, 71-6.
269. Thathiah, A., *et al.* (2003) Tumor necrosis factor-alpha converting enzyme/ADAM 17 mediates MUC1 shedding. *J Biol Chem*, **278**, 3386-94.
270. Linden, S.K., *et al.* (2009) MUC1 limits Helicobacter pylori infection both by steric hindrance and by acting as a releasable decoy. *PLoS Pathog*, **5**, e1000617.
271. Wreschner, D.H., *et al.* (1990) Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein forms. *Eur J Biochem*, **189**, 463-73.
272. Baruch, A., *et al.* (1997) Preferential expression of novel MUC1 tumor antigen isoforms in human epithelial tumors and their tumor-potentiating function. *Int J Cancer*, **71**, 741-9.
273. Levitin, F., *et al.* (2005) A novel protein derived from the MUC1 gene by alternative splicing and frameshifting. *J Biol Chem*, **280**, 10655-63.

274. Hollingsworth, M.A., *et al.* (2004) Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer*, **4**, 45-60.
275. Kinlough, C.L., *et al.* (2006) Recycling of MUC1 is dependent on its palmitoylation. *J Biol Chem*, **281**, 12112-22.
276. Pinto-de-Sousa, J., *et al.* (2004) MUC5B expression in gastric carcinoma: relationship with clinico-pathological parameters and with expression of mucins MUC1, MUC2, MUC5AC and MUC6. *Virchows Arch*, **444**, 224-30.
277. Senapati, S., *et al.* (2008) The MUC gene family: their role in the diagnosis and prognosis of gastric cancer. *Histol Histopathol*, **23**, 1541-52.
278. Buisine, M.P., *et al.* (1998) Genomic organization of the 3'-region of the human MUC5AC mucin gene: additional evidence for a common ancestral gene for the 11p15.5 mucin gene family. *Biochem J*, **332 (Pt 3)**, 729-38.
279. Escande, F., *et al.* (2001) Human mucin gene MUC5AC: organization of its 5'-region and central repetitive region. *Biochem J*, **358**, 763-72.
280. Rousseau, K., *et al.* (2004) The complete genomic organization of the human MUC6 and MUC2 mucin genes. *Genomics*, **83**, 936-9.
281. Lopez-Ferrer, A., *et al.* (2000) Role of fucosyltransferases in the association between apomucin and Lewis antigen expression in normal and malignant gastric epithelium. *Gut*, **47**, 349-56.
282. Perez-Vilar, J., *et al.* (1999) The structure and assembly of secreted mucins. *J Biol Chem*, **274**, 31751-4.
283. Patsos, G., *et al.* (2009) Management of the human mucosal defensive barrier: evidence for glycan legislation. *Biol Chem*, **390**, 581-90.
284. Lidell, M.E., *et al.* (2006) Cleavage in the GDPH sequence of the C-terminal cysteine-rich part of the human MUC5AC mucin. *Biochem J*, **399**, 121-9.
285. Lidell, M.E., *et al.* (2003) An autocatalytic cleavage in the C terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. *J Biol Chem*, **278**, 13944-51.
286. Thornton, D.J., *et al.* (2008) Structure and function of the polymeric mucins in airways mucus. *Annu Rev Physiol*, **70**, 459-86.
287. Gendler, S.J., *et al.* (1990) Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J Biol Chem*, **265**, 15286-93.
288. Ligtenberg, M.J., *et al.* (1990) Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini. *J Biol Chem*, **265**, 5573-8.
289. Pigny, P., *et al.* (1996) Human mucin genes assigned to 11p15.5: identification and organization of a cluster of genes. *Genomics*, **38**, 340-52.
290. Rousseau, K., *et al.* (2007) Allelic association and recombination hotspots in the mucin gene (MUC) complex on chromosome 11p15.5. *Ann Hum Genet*, **71**, 561-9.
291. Desseyn, J.L., *et al.* (2000) Evolution of the large secreted gel-forming mucins. *Mol Biol Evol*, **17**, 1175-84.
292. Vinall, L.E., *et al.* (1998) Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. *Hum Genet*, **102**, 357-66.
293. Kovarik, A., *et al.* (1993) Analysis of the tissue-specific promoter of the MUC1 gene. *J Biol Chem*, **268**, 9917-26.
294. Kovarik, A., *et al.* (1996) Two GC boxes (Sp1 sites) are involved in regulation of the activity of the epithelium-specific MUC1 promoter. *J Biol Chem*, **271**, 18140-7.
295. Morris, J.R., *et al.* (2001) The Sp1 transcription factor regulates cell type-specific transcription of MUC1. *DNA Cell Biol*, **20**, 133-9.
296. Gaemers, I.C., *et al.* (2001) A stat-responsive element in the promoter of the episialin/MUC1 gene is involved in its overexpression in carcinoma cells. *J Biol Chem*, **276**, 6191-9.

297. Abe, M., *et al.* (1993) Characterization of cis-acting elements regulating transcription of the human DF3 breast carcinoma-associated antigen (MUC1) gene. *Proc Natl Acad Sci U S A*, **90**, 282-6.
298. Gum, J.R., *et al.* (1997) Identification and characterization of the MUC2 (human intestinal mucin) gene 5'-flanking region: promoter activity in cultured cells. *Biochem J*, **325 (Pt 1)**, 259-67.
299. Li, D., *et al.* (1998) Cloning of the amino-terminal and 5'-flanking region of the human MUC5AC mucin gene and transcriptional up-regulation by bacterial exoproducts. *J Biol Chem*, **273**, 6812-20.
300. Sakai, H., *et al.* (2005) Upregulation of MUC6 mucin gene expression by NFkappaB and Sp factors. *Biochem Biophys Res Commun*, **333**, 1254-60.
301. Aslam, F., *et al.* (2001) The Sp family of transcription factors in the regulation of the human and mouse MUC2 gene promoters. *Cancer Res*, **61**, 570-6.
302. Perrais, M., *et al.* (2002) Induction of MUC2 and MUC5AC mucins by factors of the epidermal growth factor (EGF) family is mediated by EGF receptor/Ras/Raf/extracellular signal-regulated kinase cascade and Sp1. *J Biol Chem*, **277**, 32258-67.
303. Yamamoto, H., *et al.* (2003) Homeodomain protein CDX2 regulates goblet-specific MUC2 gene expression. *Biochem Biophys Res Commun*, **300**, 813-8.
304. Cornberg, M., *et al.* (1999) Variation of human mucin gene expression in gastric cancer cell lines and gastric mucous cell primary cultures. *Eur J Cell Biol*, **78**, 832-41.
305. Grohmann, G.P., *et al.* (2003) Modulation of MUC1 and blood group antigen expression in gastric adenocarcinoma cells by cytokines. *Cytokine*, **23**, 86-93.
306. Kang, W., *et al.* (2005) Interferon gamma induction of gastric mucous neck cell hypertrophy. *Lab Invest*, **85**, 702-15.
307. McGuckin, M.A., *et al.* (1998) Progesterone stimulates production and secretion of MUC1 epithelial mucin in steroid-responsive breast cancer cell lines. *Int J Oncol*, **12**, 939-45.
308. Meseguer, M., *et al.* (2001) Human endometrial mucin MUC1 is up-regulated by progesterone and down-regulated in vitro by the human blastocyst. *Biol Reprod*, **64**, 590-601.
309. Zaretsky, J.Z., *et al.* (1999) Analysis of the promoter of the MUC1 gene overexpressed in breast cancer. *FEBS Lett*, **461**, 189-95.
310. Zaretsky, J.Z., *et al.* (2006) MUC1 gene overexpressed in breast cancer: structure and transcriptional activity of the MUC1 promoter and role of estrogen receptor alpha (ERalpha) in regulation of the MUC1 gene expression. *Mol Cancer*, **5**, 57.
311. Herman, J.G., *et al.* (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med*, **349**, 2042-54.
312. de Bustros, A., *et al.* (1988) The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia. *Proc Natl Acad Sci U S A*, **85**, 5693-7.
313. Hanski, C., *et al.* (1997) MUC2 gene suppression in human colorectal carcinomas and their metastases: in vitro evidence of the modulatory role of DNA methylation. *Lab Invest*, **77**, 685-95.
314. Gratchev, A., *et al.* (2001) Regulation of the intestinal mucin MUC2 gene expression in vivo: evidence for the role of promoter methylation. *Cancer Lett*, **168**, 71-80.
315. Mesquita, P., *et al.* (2003) Role of site-specific promoter hypomethylation in aberrant MUC2 mucin expression in mucinous gastric carcinomas. *Cancer Lett*, **189**, 129-36.
316. Vincent, A., *et al.* (2007) Epigenetic regulation (DNA methylation, histone modifications) of the 11p15 mucin genes (MUC2, MUC5AC, MUC5B, MUC6) in epithelial cancer cells. *Oncogene*, **26**, 6566-76.
317. Yamada, N., *et al.* (2008) MUC1 expression is regulated by DNA methylation and histone H3 lysine 9 modification in cancer cells. *Cancer Res*, **68**, 2708-16.
318. Okudaira, K., *et al.* (2010) MUC2 gene promoter methylation in mucinous and non-mucinous colorectal cancer tissues. *Int J Oncol*, **36**, 765-75.

319. Kent, O.A., *et al.* (2006) A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. *Oncogene*, **25**, 6188-96.
320. Jin, C., *et al.* (2010) miR-1226 targets expression of the mucin 1 oncoprotein and induces cell death. *Int J Oncol*, **37**, 61-9.
321. Rajabi, H., *et al.* (2010) MUCIN 1 ONCOPROTEIN EXPRESSION IS SUPPRESSED BY THE miR-125b ONCOMIR. *Genes Cancer*, **1**, 62-68.
322. Sachdeva, M., *et al.* (2010) MicroRNA-145 suppresses cell invasion and metastasis by directly targeting mucin 1. *Cancer Res*, **70**, 378-87.
323. Moran, A.P., *et al.* (2011) Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut*, **60**, 1412-25.
324. Atuma, C., *et al.* (2001) The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *Am J Physiol Gastrointest Liver Physiol*, **280**, G922-9.
325. Ho, S.B., *et al.* (2004) The adherent gastric mucous layer is composed of alternating layers of MUC5AC and MUC6 mucin proteins. *Dig Dis Sci*, **49**, 1598-606.
326. Phillipson, M. (2004) Acid transport through gastric mucus. *Ups J Med Sci*, **109**, 1-24.
327. Van de Bovenkamp, J.H., *et al.* (2003) The MUC5AC glycoprotein is the primary receptor for *Helicobacter pylori* in the human stomach. *Helicobacter*, **8**, 521-32.
328. McGuckin, M.A., *et al.* (2011) Mucin dynamics and enteric pathogens. *Nat Rev Microbiol*, **9**, 265-78.
329. McGuckin, M.A., *et al.* (2007) Muc1 mucin limits both *Helicobacter pylori* colonization of the murine gastric mucosa and associated gastritis. *Gastroenterology*, **133**, 1210-8.
330. Guang, W., *et al.* (2010) Muc1 cell surface mucin attenuates epithelial inflammation in response to a common mucosal pathogen. *J Biol Chem*, **285**, 20547-57.
331. Hoffmann, W. (2005) Trefoil factors TFF (trefoil factor family) peptide-triggered signals promoting mucosal restitution. *Cell Mol Life Sci*, **62**, 2932-8.
332. Taupin, D., *et al.* (1999) The trefoil gene family are coordinately expressed immediate-early genes: EGF receptor- and MAP kinase-dependent interregulation. *J Clin Invest*, **103**, R31-8.
333. Strugnell, R.A., *et al.* (2010) The role of secretory antibodies in infection immunity. *Nat Rev Microbiol*, **8**, 656-67.
334. Kobayashi, K., *et al.* (2002) Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids. *Gut*, **51**, 169-76.
335. Yu, H., *et al.* (2011) The rat IgG Fc gamma BP and Muc2 C-terminal domains and TFF3 in two intestinal mucus layers bind together by covalent interaction. *PLoS One*, **6**, e20334.
336. Kawakubo, M., *et al.* (2004) Natural antibiotic function of a human gastric mucin against *Helicobacter pylori* infection. *Science*, **305**, 1003-6.
337. Kam, J.L., *et al.* (1998) MUC1 synthetic peptide inhibition of intercellular adhesion molecule-1 and MUC1 binding requires six tandem repeats. *Cancer Res*, **58**, 5577-81.
338. Rahn, J.J., *et al.* (2004) MUC1 initiates a calcium signal after ligation by intercellular adhesion molecule-1. *J Biol Chem*, **279**, 29386-90.
339. Wen, Y., *et al.* (2003) Nuclear association of the cytoplasmic tail of MUC1 and beta-catenin. *J Biol Chem*, **278**, 38029-39.
340. Wei, X., *et al.* (2005) Human MUC1 oncoprotein regulates p53-responsive gene transcription in the genotoxic stress response. *Cancer Cell*, **7**, 167-78.
341. Wei, X., *et al.* (2006) MUC1 oncoprotein stabilizes and activates estrogen receptor alpha. *Mol Cell*, **21**, 295-305.
342. Thompson, E.J., *et al.* (2006) Tyrosines in the MUC1 cytoplasmic tail modulate transcription via the extracellular signal-regulated kinase 1/2 and nuclear factor-kappaB pathways. *Mol Cancer Res*, **4**, 489-97.
343. Ahmad, R., *et al.* (2007) MUC1 oncoprotein activates the I kappa B kinase beta complex and constitutive NF-kappaB signalling. *Nat Cell Biol*, **9**, 1419-27.

344. Byrd, J.C., *et al.* (2000) Inhibition of gastric mucin synthesis by *Helicobacter pylori*. *Gastroenterology*, **118**, 1072-9.
345. Marques, T., *et al.* (2005) Topographic expression of MUC5AC and MUC6 in the gastric mucosa infected by *Helicobacter pylori* and in associated diseases. *Pathol Res Pract*, **201**, 665-72.
346. Babu, S.D., *et al.* (2006) Expression profile of mucins (MUC2, MUC5AC and MUC6) in *Helicobacter pylori* infected pre-neoplastic and neoplastic human gastric epithelium. *Mol Cancer*, **5**, 10.
347. Van den Brink, G.R., *et al.* (2000) H pylori colocalises with MUC5AC in the human stomach. *Gut*, **46**, 601-7.
348. Jass, J.R. (2000) Mucin core proteins as differentiation markers in the gastrointestinal tract. *Histopathology*, **37**, 561-4.
349. Silva, E., *et al.* (2002) Mucins as key molecules for the classification of intestinal metaplasia of the stomach. *Virchows Arch*, **440**, 311-7.
350. Kufe, D.W. (2009) Mucins in cancer: function, prognosis and therapy. *Nat Rev Cancer*, **9**, 874-85.
351. Pinto-de-Sousa, J., *et al.* (2002) Mucins MUC1, MUC2, MUC5AC and MUC6 expression in the evaluation of differentiation and clinico-biological behaviour of gastric carcinoma. *Virchows Arch*, **440**, 304-10.
352. Lee, H.S., *et al.* (2001) MUC1, MUC2, MUC5AC, and MUC6 expressions in gastric carcinomas: their roles as prognostic indicators. *Cancer*, **92**, 1427-34.
353. Ohno, T., *et al.* (2006) Prognostic significance of combined expression of MUC1 and adhesion molecules in advanced gastric cancer. *Eur J Cancer*, **42**, 256-63.
354. Ando, H., *et al.* (2009) Prognostic significance of the expression of MUC1 and collagen type IV in advanced gastric carcinoma. *Br J Surg*, **96**, 901-9.
355. Bafna, S., *et al.* (2010) Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene*, **29**, 2893-904.
356. Velcich, A., *et al.* (2002) Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science*, **295**, 1726-9.
357. Carvalho, F., *et al.* (1997) MUC1 gene polymorphism and gastric cancer--an epidemiological study. *Glycoconj J*, **14**, 107-11.
358. Silva, F., *et al.* (2001) MUC1 gene polymorphism in the gastric carcinogenesis pathway. *Eur J Hum Genet*, **9**, 548-52.
359. Silva, F., *et al.* (2003) MUC1 polymorphism confers increased risk for intestinal metaplasia in a Colombian population with chronic gastritis. *Eur J Hum Genet*, **11**, 380-4.
360. Vinall, L.E., *et al.* (2002) Altered expression and allelic association of the hypervariable membrane mucin MUC1 in *Helicobacter pylori* gastritis. *Gastroenterology*, **123**, 41-9.
361. Garcia, E., *et al.* (1997) MUC6 gene polymorphism in healthy individuals and in gastric cancer patients from northern Portugal. *Cancer Epidemiol Biomarkers Prev*, **6**, 1071-4.
362. Nguyen, T.V., *et al.* (2006) Short mucin 6 alleles are associated with H pylori infection. *World J Gastroenterol*, **12**, 6021-5.
363. Xu, Q., *et al.* (2009) Risk of gastric cancer is associated with the MUC1 568 A/G polymorphism. *Int J Oncol*, **35**, 1313-20.
364. Jia, Y., *et al.* (2010) A comprehensive analysis of common genetic variation in MUC1, MUC5AC, MUC6 genes and risk of stomach cancer. *Cancer Causes Control*, **21**, 313-21.
365. Frank, B., *et al.* (2011) Polymorphisms in MUC1, MUC2, MUC5B and MUC6 genes are not associated with the risk of chronic atrophic gastritis. *Eur J Cancer*.
366. Gonzalez, C.A., *et al.* (2011) *Helicobacter pylori* cagA and vacA Genotypes as Predictors of Progression of Gastric Preneoplastic Lesions: A Long-Term Follow-Up in a High-Risk Area in Spain. *Am J Gastroenterol*, **106**, 867-74.
367. Riboli, E., *et al.* (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr*, **5**, 1113-24.

368. Carneiro, F., *et al.* (2007) Pathology findings and validation of gastric and esophageal cancer cases in a European cohort (EPIC/EUR-GAST). *Scand J Gastroenterol*, **42**, 618-27.
369. Teng, M., *et al.* (2012) regSNPs: a strategy for prioritizing regulatory single nucleotide substitutions. *Bioinformatics*, **28**, 1879-86.
370. Macintyre, G., *et al.* (2010) is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics*, **26**, i524-30.
371. Desmet, F.O., *et al.* (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*, **37**, e67.
372. Adzhubei, I.A., *et al.* (2010) A method and server for predicting damaging missense mutations. *Nat Methods*, **7**, 248-9.
373. Kumar, P., *et al.* (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, **4**, 1073-81.
374. Barrett, J.C., *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263-5.
375. Gabriel, S.B., *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225-9.
376. Montgomery, J., *et al.* (2007) Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nat Protoc*, **2**, 59-66.
377. Gonzalez, J.R., *et al.* (2007) SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, **23**, 644-5.
378. Landis, J.R., *et al.* (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-74.
379. Genomes Project, C., *et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56-65.
380. Lee, W., *et al.* (2009) Analytical methods for inferring functional effects of single base pair substitutions in human cancers. *Hum Genet*, **126**, 481-98.
381. May, F.E., *et al.* (2000) The human two domain trefoil protein, TFF2, is glycosylated in vivo in the stomach. *Gut*, **46**, 454-9.
382. Hanisch, F.G., *et al.* (2013) Human gastric TFF2 peptide contains an N-linked fucosylated N,N'-diacetyllactosamine (LacdiNAc) oligosaccharide. *Glycobiology*, **23**, 2-11.
383. Chen, J.M., *et al.* (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Hum Genet*, **120**, 1-21.
384. Chen, J.M., *et al.* (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. *Hum Genet*, **120**, 301-33.
385. Xie, X., *et al.* (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, **434**, 338-45.
386. Siepel, A., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, **15**, 1034-50.
387. Sethupathy, P., *et al.* (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat Methods*, **3**, 881-6.
388. Chen, K., *et al.* (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet*, **38**, 1452-6.
389. Shaw, G., *et al.* (2002) Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J*, **16**, 869-71.
390. Quante, M., *et al.* (2010) TFF2 mRNA transcript expression marks a gland progenitor cell of the gastric oxyntic mucosa. *Gastroenterology*, **139**, 2018-2027 e2.
391. Hansen, T.B., *et al.* (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J*, **30**, 4414-22.
392. Faghihi, M.A., *et al.* (2009) Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol*, **10**, 637-43.

393. Rutnam, Z.J., *et al.* (2012) The non-coding 3' UTR of CD44 induces metastasis by regulating extracellular matrix functions. *J Cell Sci*, **125**, 2075-85.
394. Kim, H.S., *et al.* (2011) Genome-wide molecular characterization of mucinous colorectal adenocarcinoma using cDNA microarray analysis. *Oncol Rep*, **25**, 717-27.
395. Wang, Q., *et al.* (2012) Plasma miR-601 and miR-760 are novel biomarkers for the early detection of colorectal cancer. *PLoS One*, **7**, e44398.
396. Cicatiello, L., *et al.* (2010) Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. *Am J Pathol*, **176**, 2113-30.
397. Green, K.A., *et al.* (2007) Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state. *Nat Rev Cancer*, **7**, 713-22.
398. Wang, X.Y., *et al.* (2010) Differential miRNA expression and their target genes between NGX6-positive and negative colon cancer cells. *Mol Cell Biochem*, **345**, 283-90.
399. Suh, Y., *et al.* (2005) SNP discovery in associating genetic variation with human disease phenotypes. *Mutat Res*, **573**, 41-53.
400. International HapMap, C. (2004) Integrating ethics and science in the International HapMap Project. *Nat Rev Genet*, **5**, 467-75.
401. Ribas, G., *et al.* (2006) Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum Genet*, **118**, 669-79.
402. Montpetit, A., *et al.* (2006) An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. *PLoS Genet*, **2**, e27.
403. Gonzalez-Neira, A., *et al.* (2006) The portability of tagSNPs across populations: a worldwide survey. *Genome Res*, **16**, 323-30.
404. Mueller, J.C., *et al.* (2005) Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am J Hum Genet*, **76**, 387-98.
405. Pettersson, F.H., *et al.* (2009) Marker selection for genetic case-control association studies. *Nat Protoc*, **4**, 743-52.
406. Pe'er, I., *et al.* (2006) Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet*, **38**, 663-7.
407. de Bakker, P.I., *et al.* (2005) Efficiency and power in genetic association studies. *Nat Genet*, **37**, 1217-23.
408. Frazer, K.A., *et al.* (2009) Human genetic variation and its contribution to complex traits. *Nat Rev Genet*, **10**, 241-51.
409. Xu, Z., *et al.* (2007) Tag SNP selection for candidate gene association studies using HapMap and gene resequencing data. *Eur J Hum Genet*, **15**, 1063-70.
410. Hong, E.P., *et al.* (2012) Sample size and statistical power calculation in genetic association studies. *Genomics Inform*, **10**, 117-22.
411. Gauderman, W.J. (2002) Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol*, **155**, 478-84.
412. Gauderman, W.J. (2002) Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med*, **21**, 35-50.
413. Li, W.Q., *et al.* (2009) Association between genetic polymorphisms of DNA base excision repair genes and evolution of precancerous gastric lesions in a Chinese population. *Carcinogenesis*, **30**, 500-5.
414. Brenner, H., *et al.* (2004) Is *Helicobacter pylori* infection a necessary condition for noncardia gastric cancer? *Am J Epidemiol*, **159**, 252-8.
415. Sjodahl, K., *et al.* (2007) Smoking and alcohol drinking in relation to risk of gastric cancer: a population-based, prospective cohort study. *Int J Cancer*, **120**, 128-32.
416. Freedman, N.D., *et al.* (2007) A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am J Epidemiol*, **165**, 1424-33.
417. Nomura, A.M., *et al.* (2012) The association of cigarette smoking with gastric cancer: the multiethnic cohort study. *Cancer Causes Control*, **23**, 51-8.

418. Ladeiras-Lopes, R., *et al.* (2008) Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. *Cancer Causes Control*, **19**, 689-701.
419. Koivisto, T.T., *et al.* (2008) Effect of smoking on gastric histology in Helicobacter pylori-positive gastritis. *Scand J Gastroenterol*, **43**, 1177-83.
420. Redeen, S., *et al.* (2010) Natural history of chronic gastritis in a population-based cohort. *Scand J Gastroenterol*, **45**, 540-9.
421. Adamu, M.A., *et al.* (2011) Incidence and risk factors for the development of chronic atrophic gastritis: five year follow-up of a population-based cohort study. *Int J Cancer*, **128**, 1652-8.
422. Felley, C., *et al.* (2012) Age, smoking and overweight contribute to the development of intestinal metaplasia of the cardia. *World J Gastroenterol*, **18**, 2076-83.
423. Russo, A., *et al.* (2001) Effect of lifestyle, smoking, and diet on development of intestinal metaplasia in H. pylori-positive subjects. *Am J Gastroenterol*, **96**, 1402-8.
424. Peleteiro, B., *et al.* (2007) Smoking, Helicobacter pylori virulence, and type of intestinal metaplasia in Portuguese males. *Cancer Epidemiol Biomarkers Prev*, **16**, 322-6.
425. Din, F.V., *et al.* (2010) Effect of aspirin and NSAIDs on risk and survival from colorectal cancer. *Gut*, **59**, 1670-9.
426. Yang, P., *et al.* (2010) Aspirin use and the risk of gastric cancer: a meta-analysis. *Dig Dis Sci*, **55**, 1533-9.
427. Bosetti, C., *et al.* (2012) Aspirin and cancer risk: a quantitative review to 2011. *Ann Oncol*, **23**, 1403-15.
428. Wang, W.H., *et al.* (2003) Non-steroidal anti-inflammatory drug use and the risk of gastric cancer: a systematic review and meta-analysis. *J Natl Cancer Inst*, **95**, 1784-91.
429. Shirin, H., *et al.* (2006) Non-steroidal anti-inflammatory drugs have bacteriostatic and bactericidal activity against Helicobacter pylori. *J Gastroenterol Hepatol*, **21**, 1388-93.
430. Ubukata, H., *et al.* (2011) Why is the coexistence of gastric cancer and duodenal ulcer rare? Examination of factors related to both gastric cancer and duodenal ulcer. *Gastric Cancer*, **14**, 4-12.
431. Thun, M.J., *et al.* (2002) Nonsteroidal anti-inflammatory drugs as anticancer agents: mechanistic, pharmacologic, and clinical issues. *J Natl Cancer Inst*, **94**, 252-66.
432. Dai, Y., *et al.* (2006) Non-steroidal anti-inflammatory drugs in prevention of gastric cancer. *World J Gastroenterol*, **12**, 2884-9.
433. Elwood, P.C., *et al.* (2009) Aspirin, salicylates, and cancer. *Lancet*, **373**, 1301-9.
434. Chang, S.K., *et al.* (1994) Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology*, **107**, 28-36.
435. Kim, Y.S., *et al.* (2010) Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Curr Gastroenterol Rep*, **12**, 319-30.
436. Ho, S.B., *et al.* (1995) Mucin gene expression in normal, preneoplastic, and neoplastic human gastric epithelium. *Cancer Res*, **55**, 2681-90.
437. Mejias-Luque, R., *et al.* (2010) Inflammation modulates the expression of the intestinal mucins MUC2 and MUC4 in gastric tumors. *Oncogene*, **29**, 1753-62.
438. Ho, J.J., *et al.* (2003) Methylation status of promoters and expression of MUC2 and MUC5AC mucins in pancreatic cancer cells. *Int J Oncol*, **22**, 273-9.
439. Moehle, C., *et al.* (2006) Aberrant intestinal expression and allelic variants of mucin genes associated with inflammatory bowel disease. *J Mol Med*, **84**, 1055-66.
440. Swallow, D.M., *et al.* (1987) The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. *Nature*, **328**, 82-4.
441. Lidell, M.E., *et al.* (2003) The recombinant C-terminus of the human MUC2 mucin forms dimers in Chinese-hamster ovary cells and heterodimers with full-length MUC2 in LS 174T cells. *Biochem J*, **372**, 335-45.

442. Lidell, M.E., *et al.* (2006) Entamoeba histolytica cysteine proteases cleave the MUC2 mucin in its C-terminal domain and dissolve the protective colonic mucus gel. *Proc Natl Acad Sci U S A*, **103**, 9298-303.
443. Wallace, J.L., *et al.* (2011) Muc-2-deficient mice display a sex-specific, COX-2-related impairment of gastric mucosal repair. *Am J Pathol*, **178**, 1126-33.
444. Wong, W.M., *et al.* (1999) Trefoil peptides. *Gut*, **44**, 890-5.
445. Kim, B.W., *et al.* (2004) Expression of trefoil peptides in the subtypes of intestinal metaplasia. *Peptides*, **25**, 779-83.
446. Bignotti, E., *et al.* (2008) Trefoil factor 3: a novel serum marker identified by gene expression profiling in high-grade endometrial carcinomas. *Br J Cancer*, **99**, 768-73.
447. Vestergaard, E.M., *et al.* (2010) Promoter hypomethylation and upregulation of trefoil factors in prostate cancer. *Int J Cancer*, **127**, 1857-65.
448. Kosriwong, K., *et al.* (2011) Trefoil factors: tumor progression markers and mitogens via EGFR/MAPK activation in cholangiocarcinoma. *World J Gastroenterol*, **17**, 1631-41.
449. Davidson, B., *et al.* (2012) Gene expression signatures differentiate adenocarcinoma of lung and breast origin in effusions. *Hum Pathol*, **43**, 684-94.
450. Tonacchera, M., *et al.* (2012) Genetic markers to discriminate benign and malignant thyroid nodules with undetermined cytology in an area of borderline iodine deficiency. *J Endocrinol Invest*, **35**, 754-9.
451. Karger, S., *et al.* (2012) ADM3, TFF3 and LGALS3 are discriminative molecular markers in fine-needle aspiration biopsies of benign and malignant thyroid tumours. *Br J Cancer*, **106**, 562-8.
452. Qu, Y., *et al.* (2012) Increased trefoil factor 3 levels in the serum of patients with three major histological subtypes of lung cancer. *Oncol Rep*, **27**, 1277-83.
453. Koebel, M., *et al.* (2013) Biomarker-based ovarian carcinoma typing: a histological investigation in the Ovarian Tumor Tissue Analysis consortium. *Cancer Epidemiol Biomarkers Prev*.
454. Emami, S., *et al.* (2001) Induction of scattering and cellular invasion by trefoil peptides in src- and RhoA-transformed kidney and colonic epithelial cells. *FASEB J*, **15**, 351-61.
455. Meng, J.R., *et al.* (2012) TFF3 and survivin expressions associate with a lower survival rate in gastric cancer. *Clin Exp Med*.
456. Okada, H., *et al.* (2005) Frequent trefoil factor 3 (TFF3) overexpression and promoter hypomethylation in mouse and human hepatocellular carcinomas. *Int J Oncol*, **26**, 369-77.
457. Shimada, T., *et al.* (2007) Regulation of TFF3 expression by homeodomain protein CDX2. *Regul Pept*, **140**, 81-7.
458. Almeida, R., *et al.* (2003) Expression of intestine-specific transcription factors, CDX1 and CDX2, in intestinal metaplasia and gastric carcinomas. *J Pathol*, **199**, 36-40.
459. Bornschein, J., *et al.* (2013) Dysregulation of CDX1, CDX2 and SOX2 in patients with gastric cancer also affects the non-malignant mucosa. *J Clin Pathol*, **66**, 819-22.
460. Kocer, B., *et al.* (2004) Prognostic significance of mucin expression in gastric carcinoma. *Dig Dis Sci*, **49**, 954-64.
461. Tamura, Y., *et al.* (2012) MUC4 and MUC1 expression in adenocarcinoma of the stomach correlates with vessel invasion and lymph node metastasis: an immunohistochemical study of early gastric cancer. *PLoS One*, **7**, e49251.
462. Rajabi, H., *et al.* (2012) MUC1-C oncoprotein confers androgen-independent growth of human prostate cancer cells. *Prostate*, **72**, 1659-68.
463. Raina, D., *et al.* (2012) Targeting cysteine-mediated dimerization of the MUC1-C oncoprotein in human cancer cells. *Int J Oncol*, **40**, 1643-9.
464. Kufe, D.W. (2013) MUC1-C oncoprotein as a target in breast cancer: activation of signaling pathways and therapeutic approaches. *Oncogene*, **32**, 1073-81.
465. Zhang, H., *et al.* (2011) Genetic variants at 1q22 and 10q23 reproducibly associated with gastric cancer susceptibility in a Chinese population. *Carcinogenesis*, **32**, 848-52.

466. Yang, X.X., *et al.* (2012) Association of genetic polymorphisms at 1q22 but not 10q23 with gastric cancer in a southern Chinese population. *Asian Pac J Cancer Prev*, **13**, 2519-22.
467. Balleine, R.L., *et al.* (1999) Expression of the oestrogen responsive protein pS2 in human breast cancer. *Histol Histopathol*, **14**, 571-8.
468. Crosier, M., *et al.* (2001) High expression of the trefoil protein TFF1 in interval breast cancers. *Am J Pathol*, **159**, 215-21.
469. Buache, E., *et al.* (2011) Deficiency in trefoil factor 1 (TFF1) increases tumorigenicity of human breast cancer cells and mammary tumor development in TFF1-knockout mice. *Oncogene*, **30**, 3261-73.
470. Baron, S., *et al.* (2007) Estrogen receptor alpha and the activating protein-1 complex cooperate during insulin-like growth factor-I-induced transcriptional activation of the pS2/TFF1 gene. *J Biol Chem*, **282**, 11732-41.
471. Li, L., *et al.* (2008) Association of Sp3 and estrogen receptor alpha with the transcriptionally active trefoil factor 1 promoter in MCF-7 breast cancer cells. *J Cell Biochem*, **105**, 365-9.
472. Taylor, A.H., *et al.* (2000) Immunolocalisation of oestrogen receptor beta in human tissues. *J Mol Endocrinol*, **24**, 145-55.
473. Campbell-Thompson, M., *et al.* (2001) Immunolocalization of estrogen receptor alpha and beta in gastric epithelium and enteric neurons. *J Endocrinol*, **171**, 65-73.
474. Gan, L., *et al.* (2012) Expression profile and prognostic role of sex hormone receptors in gastric cancer. *BMC Cancer*, **12**, 566.
475. Jeltsch, J.M., *et al.* (1987) Structure of the human oestrogen-responsive gene pS2. *Nucleic Acids Res*, **15**, 1401-14.
476. Barrett, L.W., *et al.* (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci*, **69**, 3613-34.
477. Kwon, J.A., *et al.* (2010) Short rare MUC6 minisatellites-5 alleles influence susceptibility to gastric carcinoma by regulating gene. *Hum Mutat*, **31**, 942-9.
478. Liu, J., *et al.* (2012) The regulation of trefoil factor 2 expression by the transcription factor Sp3. *Biochem Biophys Res Commun*, **427**, 410-4.
479. Shimada, T., *et al.* (2007) Peroxisome proliferator-activated receptor gamma (PPARgamma) regulates trefoil factor family 2 (TFF2) expression in gastric epithelial cells. *Int J Biochem Cell Biol*, **39**, 626-37.
480. Castano-Rodriguez, N., *et al.* (2013) The role of TLR2, TLR4 and CD14 genetic polymorphisms in gastric carcinogenesis: a case-control study and meta-analysis. *PLoS One*, **8**, e60327.
481. Nasri, S., *et al.* (2008) A novel diffuse gastric cancer susceptibility variant in E-cadherin (CDH1) intron 2: a case control study in an Italian population. *BMC Cancer*, **8**, 138.
482. Loh, M., *et al.* (2009) Meta-analysis of genetic polymorphisms and gastric cancer risk: variability in associations according to race. *Eur J Cancer*, **45**, 2562-8.
483. Wang, Q., *et al.* (2011) The E-cadherin (CDH1) -160C>A polymorphism associated with gastric cancer among Asians but not Europeans. *DNA Cell Biol*, **30**, 395-400.
484. Li, Y.L., *et al.* (2012) CDH1 promoter polymorphism and stomach cancer susceptibility. *Mol Biol Rep*, **39**, 1283-6.
485. Kim, E.J., *et al.* (2013) Helicobacter pylori Infection Enhances Gastric Mucosal Inflammation in Individuals Carrying the 260-T Allele of the CD14 Gene. *Gut Liver*, **7**, 317-22.
486. Hold, G.L., *et al.* (2009) CD14-159C/T and TLR9-1237T/C polymorphisms are not associated with gastric cancer risk in Caucasian populations. *Eur J Cancer Prev*, **18**, 117-9.
487. Jiang, Z., *et al.* (2005) CD14 is required for MyD88-independent LPS signaling. *Nat Immunol*, **6**, 565-70.
488. Zhao, D., *et al.* (2007) Role of CD14 promoter polymorphisms in Helicobacter pylori infection--related gastric carcinoma. *Clin Cancer Res*, **13**, 2362-8.

489. Kawai, S., *et al.* (2006) Significant association between PTPN11 polymorphism and gastric atrophy among Japanese Brazilians. *Gastric Cancer*, **9**, 277-83.
490. Goto, Y., *et al.* (2006) Association between serum pepsinogens and polymorphism of PTPN11 encoding SHP-2 among Helicobacter pylori seropositive Japanese. *Int J Cancer*, **118**, 203-8.
491. Hishida, A., *et al.* (2010) Genetic predisposition to Helicobacter pylori-induced gastric precancerous conditions. *World J Gastrointest Oncol*, **2**, 369-79.
492. Yang, J.J., *et al.* (2009) The role of TNF genetic variants and the interaction with cigarette smoking for gastric cancer risk: a nested case-control study. *BMC Cancer*, **9**, 238.
493. Hong, Y., *et al.* (2013) Functional promoter -308G>A variant in tumor necrosis factor alpha gene is associated with risk and progression of gastric cancer in a Chinese population. *PLoS One*, **8**, e50856.
494. Costa, N.L., *et al.* (2013) Tumor-associated macrophages and the profile of inflammatory cytokines in oral squamous cell carcinoma. *Oral Oncol*, **49**, 216-23.

Apéndice I: Protocolos

Índice de protocolos

PROTOCOLO I- Extracción de DNA a partir de sangre periférica	252
PROTOCOLO II- Extracción de DNA a partir de tejido incluido en parafina	253
PROTOCOLO III- Reacción en cadena de la polimerasa (PCR)	254
PROTOCOLO IV- PCR de secuencia	255
PROTOCOLO V- Purificación del producto de PCR de secuencia	255
PROTOCOLO VI- Electorforesis en geles de agarosa	256
PROTOCOLO VII- Electroforesis en gel de poliacrilamida	258
PROTOCOLO VIII- Tinción de geles con nitrato de plata.....	259
PROTOCOLO IX- Digestión del DNA con enzimas de restricción	259
PROTOCOLO X- Single Strand Conformation Polymorphism (SSCP)	261
PROTOCOLO XI- Genotipado mediante sondas FRET	261
PROTOCOLO XII- Genotipado mediante sondas TaqMan [®]	262
PROTOCOLO XIII- Análisis de las variantes polimórficas con HRM.....	263
PROTOCOLO XIV- Defosforilación de un vector	264
PROTOCOLO XV- Ligación	265
PROTOCOLO XVI- Preparación de células competentes de E.coli por el método del cloruro cálcico.....	265
PROTOCOLO XVII- Transformación de células competentes mediante choque térmico.....	266
PROTOCOLO XVIII- PCR de cribaje	268
PROTOCOLO XIX- Mutagénesis dirigida	269
PROTOCOLO XX- Minipreparación de DNA plasmídico	270
PROTOCOLO XXI- Midi y Maxipreparacion de DNA plasmídico	271
PROTOCOLO XXII- Tripsinización y subcultivo de células adherentes	273
PROTOCOLO XXIII- Congelación y descongelación de células adherentes	274
PROTOCOLO XXIV- Transfección de células con lípidos catiónicos	275
PROTOCOLO XXV- Ensayos luciferasa dual.....	276

PROTOCOLO I- Extracción de DNA a partir de sangre periférica.

Materiales

- Reactivos del Flexigene DNA kit (Qiagen):
 - ◆ Tampón de lisis FG1
 - ◆ Tampón desnaturalizante FG2 y proteasa QIAGEN: Mezclar 500 µl de tampón con 5 µl de proteasa por cada ml de sangre a procesar. Esta mezcla debe prepararse en el momento de su uso.
 - ◆ Tampón de hidratación FG3
- Etanol al 70% (v/v)
- Isopropanol al 100% (v/v)

1- Añadir 25-30 ml del tampón FG1 en función del volumen de sangre de partida:

1.1- Para extracción de DNA a partir de 12 ml de sangre:

- a- Transferir 6 ml de sangre total a un tubo de 50 ml, añadir 15 ml de tampón FG1 e invertir 5 veces el tubo.
- b- Centrifugar 5 minutos a 2000 x g.
- c- Aspirar el sobrenadante con una pipeta conectada a una bomba de vacío o decantando el tubo cuidadosamente, sin tocar el precipitado.
- d- Transferir los 6 ml de sangre restantes al mismo tubo, añadir 15 ml de tampón FG1 e invertir 5 veces el tubo.

1.2- Para extracción de DNA a partir de 10 ml de sangre:

- a- Transferir los 10 ml de sangre total a un tubo de 50 ml, añadir 25 ml de tampón FG1 e invierte 5 veces el tubo.

2- Centrifugar 5 minutos a 2000 x g.

3- Aspirar el sobrenadante, por aspiración o decantación, y dejar el tubo invertido sobre un papel absorbente durante 2 minutos. En este punto, el precipitado puede congelarse a -20°C hasta su procesamiento.

4- Añadir la mezcla del tampón FG2 y proteasa, y vortear rápidamente hasta que el precipitado esté completamente homogeneizado. Si se procesan varias muestras, vortear cada tubo inmediatamente después de añadir la mezcla.

5- Invertir el tubo 3 veces e incubar a 65°C durante 10 minutos.

6- Añadir 5 ml de isopropanol al 100% y mezclar por inversión hasta que se forme la medusa de DNA.

7- Centrifugar 3 minutos a 2000 x g.

8- Descartar el sobrenadante e invertir el tubo sobre un papel absorbente durante 2-3 minutos.

9- Añadir 5 ml de etanol al 70% y golpear suavemente el tubo para que el DNA se desprenda y se limpie.

10- Centrifugar 3 minutos a 2000 x g.

11- Descartar el sobrenadante e invertir el tubo sobre un papel absorbente durante, aproximadamente, 10 minutos.

- 12- Poner el tubo en una gradilla a temperatura ambiente para que se evapore el etanol residual.
- 13- Añadir 0,5-1 ml, dependiendo del tamaño de la medusa, de dH₂O e incubar a 65°C durante 1 hora o toda la noche a 4°C.

PROTOCOLO II- Extracción de DNA a partir de tejido incluido en parafina

Desparafinación y digestión

Materiales

- Portaobjetos
 - Solución de poli-L-lisina al 10%
 - Batería de desparafinación
 - ◆ 3 cubetas de xilol puro
 - ◆ 3 cubetas de etanol al 96% (v/v)
 - Etanol al 70% (v/v)
 - Tampón PBS (Phosphate-buffered saline)
 - Proteinasa K 20 mg/ml
 - Reactivo ATL de Qiagen
- 1- Preparar los portaobjetos con una solución de poli-L-lisina al 10% y dejarlos secar a temperatura ambiente.
 - 2- Colocar los bloques de parafina en el micrótopo, ajustar la inclinación de la cuchilla y realizar 10 cortes a 10 µm para cada extracción de DNA.
 - 3- Dejar secar los cortes 30 minutos a 60°C, para que queden adheridos al vidrio, antes de iniciar la desparafinación.
 - 4- Colocar los cortes obtenidos anteriormente en una cesta y sumergirlos 10 minutos en cada una de las cubetas que contienen xilol.
 - 5- Seguidamente sumergir los cortes 10 minutos en cada una de las cubetas que contienen etanol al 96%.
 - 6- Secar los portaobjetos en una estufa a 65°C durante 1 hora.
 - 7- Rascar el tejido del portaobjetos con un bisturí y colocar el material de cada muestra en tubos de 1,5 ml.
 - 8- Añadir 500 µl de etanol al 70% a cada tubo, vortear 10 segundos, hasta que el tejido se desprenda de las paredes del tubo, y centrifugar a 14000 rpm durante 10 minutos.
 - 9- Extraer el etanol con una pipeta.
 - 10- Añadir 500 µl de PBS 1X, vortear y centrifugar a 14 000 rpm durante 10 segundos.
 - 11- Extraer el PBS de la misma manera que el etanol i dejar secar a temperatura ambiente.
 - 12- Para cada tubo de 1,5 ml añadir 180 µl de ATL y 20 µl de proteinasa K.
 - 13- Vortear hasta que el tejido se disgregue completamente, ayudarse con una punta de pipeta.

- 14- Envolver la tapa del tubo con film adherente y colocar los tubos en un baño seco a 56°C en agitación durante 3 días.
- 15- Añadir 20 µl de proteinasa K una vez al día. Centrifugar brevemente todos los tubos antes de añadir la enzima.

Extracción del DNA con QIAamp DNA Mini Kit

Materiales

- Reactivos del QIAamp DNA Mini kit (Qiagen)
 - ♦ Tampón ATL de lisis de tejido
 - ♦ Tampón AL de lisis para purificación del DNA
 - ♦ Tampón AW1 y AW2 de lavado
- Etanol absoluto

- 1- Realizar un pulso de centrifuga a los tubos y añadir 200 µl del reactivo AL, vortear e incubar a 70°C durante 10 minutos en un baño seco.
- 2- Realizar un pulso de centrifuga a los tubos y añadir 200 µl de etanol absoluto, vortear y repetir el pulso de centrifuga.
- 3- Recoger el sobrenadante, descartando el material no digerido ya que podría obturar la membrana de las columnas, y pasarlo a una columna proporcionada por el kit. Centrifugar a 8000 rpm durante 1 minuto. Cambiar la columna a un nuevo tubo de 2 ml, descartando el que contiene el producto filtrado.
- 4- Añadir 500 µl del tampón AW1 y centrifugar a 8000 rpm durante 1 minuto. Colocar la columna en un tubo nuevo y descartar el que contiene el producto filtrado.
- 5- Añadir 500 µl del tampón AW2 y centrifugar a 14000 rpm durante 3 minutos. Decantar el producto filtrado y, en el mismo tubo colector, volver a centrifugar durante 1 minuto.
- 6- Colocar la columna en un tubo de 1,5 ml, añadir 100 µl de dH₂O, incubar 5 minutos a temperatura ambiente y centrifugar a 8000 rpm durante 2 minutos.

PROTOCOLO III- Reacción en cadena de la polimerasa (PCR)

Materiales

- Tampón de PCR (10X) (100 mM Tris-HCl, 15 mM MgCl₂, 500 mM KCl, pH 8,3) (Roche)
- dNTPs (dTTP, dATP, dCTP, dGTP) a 2 mM
- Oligonucleótidos cebadores a 10 µM
- MgCl₂ a 25 mM
- Taq DNA polimerasa (5 U/µl) o Pfu Turbo (2,5 U/ µl) (Roche)
- DNA molde a 10-100 ng/µl

- 1- Preparar la mezcla siguiente para todas las muestras, teniendo en cuenta un volumen final de 25 µl, y mantenerla en hielo hasta que se coloquen los tubos en el termociclador:
 - a- Tampón de PCR 10X a concentración final 1X.
 - b- dNTPs a concentración final 0,240 mM.

- c- $MgCl_2$, cuando sea necesario, hasta una concentración final de 1,5-5 mM (incluyendo el que incorpora el tampón de PCR)
- d- Oligonucleótidos cebadores a concentración final de 0,5 μM (TABLAS M-1 y M-2).
- e- Taq o Pfu DNA polimerasa 0,5 unidades por reacción
- f- dH_2O hasta alcanzar un volumen de 24 μl para cada muestra a amplificar

- 2- Agitar suavemente y hacer un pulso de centrifuga.
- 3- Repartir la mezcla (24 μl /tubo) en tubos estériles de 0,2 ml:
- 4- Añadir a cada tubo 10-100 ng de DNA molde.
- 5- Programar el termociclador con las condiciones específicas que correspondan en cada caso. (TABLA M-3)
- 6- Cuando la temperatura supere los 80°C, colocar los tubos en el termociclador programado.

PROTOCOLO IV- PCR de secuencia

Materiales

- BigDye Taq DyeDeoxi™ Terminador Cycle Sequencing 5X (Applied Biosystems)
- Tampón BigDye 5X (Applied Biosystems)
- Oligonucleótidos cebadores a 3,2 μM
- DNA molde para secuenciar

- 1- Poner en un tubo de 0,2 ml los siguientes reactivos:
 - a- 2 μl de Tampón de BigDye (5X)
 - b- 200-300 ng de DNA
 - c- 1 μl de uno de los cebadores a 3,2 μM
 - d- 2 μl de BigDye Taq DyeDeoxi™ Terminador Cycle Sequencing 5X
 - e- El volumen necesario de dH_2O hasta un volumen 10 μl
- 2- La reacción se realiza en un termociclador, en treinta ciclos de amplificación, con el programa siguiente:
 - a- Desnaturalización: 95 °C, 30 s
 - b- Hibridación: 50 °C, 30 s.
 - c- Extensión: 60 °C, 90 s.

PROTOCOLO V- Purificación del producto de PCR de secuencia

Purificación por columnas de Sephadex G-50

Materiales

- Sephadex G-50 Fine (0,1 mg/ml en ddH_2O) (Amersham Biosciences)
- Columnas Auto Seq™ G-50 (Amersham Biosciences)

Preparación del Sephadex G-50:

- ◆ Pesar 30 g de Sephadex G-50
- ◆ Agregar 300 ml de agua bidestilada y desionizada en un aparato Milli-Q (Millipore).

- ◆ Dejar la mezcla 3 horas en agitación a temperatura ambiente. Cambiar la fase superior tres veces por agua bidestilada Milli-Q fresca.
- ◆ Conservar la resina de Sephadex a 4°C.

- 1- Colocar una columna Auto Seq™ G-50 en un tubo de 1,5 ml y añadir 750µl Sephadex G-50.
- 2- Centrifugar a 450 g durante un minuto. Desechar el eluido.
- 3- Añadir en el centro de la columna de resina 10 µl de dH₂O y centrifugar a 2000 rpm durante un minuto. Es importante colocar el tubo con la columna en la centrífuga siempre en la misma posición para mantener la estructura de la resina.
- 4- Colocar la columna en un nuevo tubo de 1,5 ml. Añadir en la zona más alta de la columna de resina los 10 µl del producto de la reacción de secuenciación y centrifugar a 2000 rpm durante un minuto.
- 5- Liofilizar la muestra eluida en un concentrador centrífugo al vacío Speed-Vac®.
- 6- Conservar la muestra protegida de la luz a -20°C hasta su análisis en el secuenciador automático

Purificación por precipitación con etanol

Materiales

- Etanol al 70% (v/v)
- Etanol al 96% (v/v)

- 1- Añadir 26 µl de agua a cada muestra y pasarlas a tubos de 1,5 ml.
- 2- Añadir 64 µl de etanol al 96% a cada tubo y vortear.
- 3- Dejar incubar durante 30 minutos a temperatura ambiente.
- 4- Centrifugar a 13200 rpm durante 20 minutos. Decantar todo el sobrenadante cuidadosamente.
- 5- Realizar 3 lavados con 100 µl de etanol al 70% centrifugando 3 minutos a 13200 rpm y decantado, después de cada centrifugación, el etanol.
- 6- Colocar el tubo boca abajo en un papel de filtro y dejar secar completamente la muestra a temperatura ambiente.
- 7- Conservar la muestra protegida de la luz a -20°C hasta su análisis en el secuenciador automático.

PROTOCOLO VI- Electroforesis en geles de agarosa

Materiales

- Agarosa NMP (normal melting point)
- TBE 10X: 900 mM Tris, 720 mM ácido bórico, 24 mM EDTA
- TBE 1X
- Marcador estándar de pesos moleculares de 1 Kb (Invitrogene):
 - ◆ 960µl TE (Tris 10mM, EDTA 1mM) pH 7.5

- ◆ 210 µl Colorante Naranja (0,25% orange G (p/v), 40% sacarosa (p/v))
- ◆ 60µl Marcador 1kb
- Solución de bromuro de etidio a 0,5 µg/ml
- Tampón de carga de ácidos nucleicos 6X: 0,25% azul de bromofenol (p/v), 40% sacarosa (p/v)
- Tampón de carga de ácidos nucleicos 2X

Preparación gel de agarosa

- 1- Pesar la cantidad de agarosa necesaria para obtener una concentración del 0,7-3%. Esta concentración varía en función del tamaño del fragmento a separar.
- 2- Añadir el volumen de TBE 1X correspondiente a la concentración necesaria que queremos obtener.
- 3- Disolver completamente la agarosa en TBE 1X aplicando calor en un horno microondas a máxima potencia.
- 4- Dejar enfriar la agarosa, hasta unos 60°, y verterla en un soporte horizontal, que actuará como molde del gel y donde se habrán colocado previamente los peines que crearán en el gel los pocillos necesarios para la carga de las muestras.
- 5- Dejar solidificar el gel, retirar el peine y colocar el gel en una cubeta de electroforesis horizontal.
- 6- Añadir TBE 1X hasta cubrir completamente el gel.

Preparación y carga de las muestras

- 1- Añadir el tampón de carga 2X a las muestras en proporción 1:1. Si el volumen de la muestra a cargar supera el volumen disponible en el pocillo se utiliza el tampón a una concentración 6X y se añade a la muestra en proporción 1:5.
- 2- Cargar las muestras en los pocillos del gel. En uno de los pocillos cargar el marcador de 1kb, el mismo volumen que se ha utilizado para las muestras.
- 3- Conectar la cubeta a una fuente de electroforesis, teniendo en cuenta los polos positivo y negativo, y ajustar el voltaje a 70-120 V, dependiendo del tamaño de los fragmentos y el tiempo que necesitará para migrar. La migración del tampón de carga incluido en la electroforesis servirá de referencia para controlar la migración de las muestras.

Visualización

- 1- Teñir el gel en una solución de bromuro de etidio a 0,5 µg/ml durante 10 min. Alternativamente, puede añadirse el bromuro de etidio en la fase de preparación del gel, siempre y cuando se tengan las precauciones necesarias y exista una zona destinada exclusivamente a este fin.
- 2- Lavar el gel con dH₂O para eliminar el exceso de bromuro del gel.
- 3- Visualizar los ácidos nucleicos sometiendo el gel a luz ultravioleta en un transiluminador. En la mayoría de casos la visualización se llevó a cabo en un transiluminador GelDoc (BioRad).

PROTOCOLO VII- Electroforesis en gel de poliacrilamida

Materiales

- Solución de acrilamida:bisacrilamida 29:1 o 19:1 al 40% (Pronadisa National Diagnostics)
- TEMED (SIGMA)
- APS (Persulfato amónico) al 1% (w/v)
- TBE 10X
- TBE 1X
- Marcador estándar de pesos moleculares de 1 Kb (Invitrogene)
- Tampón de carga de ácidos nucleicos (6X): 0,25 azul de bromofenol (w/v), 0,25% xileno-cianol (w/v), 0,40% sacarosa (w/v)

El procedimiento que se describe a continuación corresponde a un gel de 20x20 cm.

- 1- Lavar los cristales con agua y jabón. Dejar que se sequen y pulverizarlos con etanol al 70%.
- 2- Disponer los vidrios enfrentados y horizontales sobre un soporte en la mesa de trabajo y colocar separadores de 0,6 mm entre los dos cristales. Sujetarlos con tres pinzas en cada lateral para mantener la presión entre los vidrios.
- 3- Mezclar en un vaso de precipitado el volumen de acrilamida:bisacrilamida 29:1 o 19:1 al 40% necesario para obtener la concentración final deseada, 5 ml de TBE 5X, dH₂O hasta 25 ml, 180 µl de APS al 10 % y 20 µl de TEMED. Mezclar con la ayuda de una jeringa de 50 ml.
- 4- Utilizar la jeringa para inyectar la acrilamida entre los dos cristales lentamente y evitando la formación de burbujas.
- 5- Colocar el peine que servirá de molde para los pocillos donde se cargarán las muestras. Dejar solidificar un mínimo de 30 minutos.
- 6- Retirar las pinzas y el peine, y colocar los cristales, donde ha polimerizado la acrilamida, en una cubeta de electroforesis vertical. Colocar sobre los cristales una placa metálica, si la electroforesis se va realizar a 4°C (ver apartado SSCPs), y ajustar los cristales en la cubeta, ayudándose si fuera necesario con unas pinzas.
- 7- Añadir TBE 1X, tanto en el ánodo como en el cátodo de la cubeta, asegurándose que los pocillos queden completamente cubiertos por el tampón y que el TBE 1X no se filtre.
- 8- Conectar la cubeta a una fuente de electroforesis, ajustar el voltaje a 180V, y dejarla conectada durante 20 min, para permitir que el gel de acrilamida se impregne del TBE y se iguale la carga iónica del gel.
- 9- Añadir el tampón de carga a las muestras en proporción 1:5, y cargar las muestras en los pocillos. Cargar también el marcador de 1 kb en uno de los pocillos.
- 10- Conectar la cubeta a una fuente de electroforesis, ajustar el voltaje a 150-400 V, y dejar que migren las muestras durante el tiempo necesario para la correcta separación de los fragmentos.
- 11- Finalizada la electroforesis, separar los cristales con cuidado, procurando que el gel quede adherido a uno de ellos.
- 12- Recuperar el gel y teñirlo mediante tinción con bromuro de etidio o tinción con nitrato de plata.

PROTOCOLO VIII- Tinción de geles con nitrato de plata

Materiales

- Etanol al 10% (v/v)
- Ácido nítrico (HNO₃) al 1% (v/v)
- Nitrato de plata (AgNO₃) 2 gr/l
- Solución de revelado: 30 gr de carbonato sódico (Na₂CO₃) y 500 µl de formaldehído al 37% por litro.
- Ácido acético al 10% (v/v)
- Papel Whatman 3 MM y una transparencia

Cuando se vaya a realizar una tinción con nitrato plata es importante la utilización de guantes libres de polvo tanto en el proceso de preparación y manipulación del gel como durante la tinción con el nitrato de plata.

- 1- **Fijación:** Sumergir el gel durante 10 minutos en una solución de etanol al 10 % en agitación. En este punto el gel se desprende del cristal y queda en la cubeta de tinción. Descartar la solución de fijación.
- 2- **Oxidación:** Añadir la solución de ácido nítrico al 1% hasta cubrir el gel e incubarlo durante exactamente 3 minutos en agitación. Decantar el ácido nítrico.
- 3- **Lavado:** Lavar el gel en dH₂O durante 20 segundos. Realizar el proceso dos veces.
- 4- **Tinción:** Sumergir el gel en durante 20 minutos en la solución de nitrato de plata en agitación. Tapar la bandeja donde se esté realizando la tinción con papel de aluminio. Retirar el nitrato de plata.
- 5- **Lavado:** Repetir el paso 3, pero con una duración de 10 segundos por lavado. Realizar el lavado durante más de un minuto podría eliminar iones de plata del gel resultando en una menor sensibilidad.
- 6- **Revelado:** Añadir una pequeña cantidad solución de revelado sobre el gel hasta cubrirlo y agitar unos 20 segundos hasta que se oscurezca, momento en que la solución se ha saturado con el exceso de nitrato de plata. Descartar la solución de revelado. Añadir solución reveladora fresca, unos 300 ml, y agitar hasta conseguir la intensidad y contraste deseados. Descartar la solución reveladora
- 7- **Parada:** Sumergir inmediatamente el gel en una solución de ácido acético al 10 % para parar la reacción. Mantener el gel en la solución de parada durante 10 minutos. Retirar el ácido acético.
- 8- **Lavado:** Repetir el paso 3.
- 9- **Secado:** Recoger el gel adhiriéndolo en un papel Whatman un poco más grande que el gel. Cubrirlo con plástico transparente y secarlo al vacío a 80 °C durante 50 minutos

PROTOCOLO IX- Digestión del DNA con enzimas de restricción

Materiales

- Tampón de digestión específico de cada enzima

- Enzima de restricción
- BSA (albúmina bovina sérica) 100 mg/ml
- DNA a digerir

Digestión sencilla

- 1- Añadir, en tubos de 0,2 ml, los siguientes componentes de la reacción para cada muestra:
 - a- Tampón de digestión específico de cada enzima, a concentración final 1X.
 - b- Cuando el enzima lo requiera, 1 mg/ml de BSA.
 - c- 1-5 unidades de enzima (dependiendo de la capacidad de corte de cada una).
 - d- 5-10 μ l de producto de PCR o 0,1-1 μ g de DNA plasmídico.
 - e- dH₂O hasta un volumen de 20 μ l.
- 2- Agitar suavemente y realizar un pulso en la centrifuga.
- 3- Incubar la reacción a la temperatura de digestión óptima y durante el tiempo indicado para cada enzima.
- 4- Visualizar los fragmentos producto de la digestión mediante electroforesis en geles de agarosa o acrilamida, dependiendo de la resolución que sea necesaria.

Digestión doble

- 5- Realizar la primera digestión con la enzima que requiera el tampón con menor concentración de sales, tal y como se describe en el protocolo de digestión sencilla.
- 6- Separar una alícuota de 2 μ l, y utilizar 1 μ l para comprobar el resultado de la primera digestión en un gel de agarosa al 1%, añadiendo el fragmento sin digerir como control. Si la digestión no ha sido completa añadir más enzima de restricción y continuar la incubación.
- 7- Cuando la primera digestión sea completa, mezclar los siguientes componentes, para cada muestra, en tubos de 0,2 ml.
 - a- Tampón de digestión específico de cada enzima, a concentración final 1X.
 - b- Cuando el enzima lo requiera, 1 mg/ml de BSA.
 - c- 1-5 unidades de enzima (dependiendo de la capacidad de corte de cada una).
 - d- 19 μ l del producto digerido con la primera enzima.
 - e- dH₂O hasta un volumen de 40 μ l.
- 8- Agitar suavemente y realizar un pulso en la centrífuga.
- 9- Incubar la reacción a la temperatura de digestión óptima y durante el tiempo indicado para cada enzima.
- 10- Visualizar los fragmentos producto de la digestión mediante electroforesis en geles de agarosa al 1% y tinción con bromuro de etidio, utilizando como controles 1 μ l del fragmento sin digerir y 1 μ l del digerido por la primera enzima. Si la digestión no ha sido completa añadir más enzima de restricción y continuar la incubación.

PROTOCOLO X- Single Strand Conformation Polymorphism (SSCP)**Materiales**

- Tampón de carga desnaturizante: 95% formamida, 0,05% azul de bromofenol (w/v), 0,05% xileno- cianol (w/v), 20 mM EDTA a pH 8,0, 10 mM NaOH.
- Materiales necesarios para electroforesis en gel de poliacrilamida
- Materiales necesarios para la tinción de geles con nitrato de plata

- 1- Utilizar entre 2-6 μ l de cada producto de PCR, dependiendo de la intensidad de la banda visualizada por electroforesis en un gel de agarosa al 2%.
- 2- Diluir la muestra en dH₂O en un volumen final de 8 μ l y añadir el tampón de carga desnaturizante en proporción 1:1.
- 3- Desnaturizar las muestras 3 minutos a 95°C. Colocar los tubos inmediatamente en hielo para facilitar el plegamiento de las cadenas sencillas y evitar que renaturalicen.
- 4- Cargar toda la muestra en un gel de poliacrilamida 29:1 del 10-12%.
- 5- Realizar la electroforesis de los productos amplificados en una cámara fría a 4°C, a 300 V durante 16-24 horas, dependiendo del fragmento a analizar.
- 6- Teñir el gel de acrilamida con nitrato de plata y secar el gel tal como se describe en el PROTOCOLO VIII.
- 7- Comparar el patrón de bandas de los fragmentos analizados.
- 8- Cuando se detecte más de un patrón, correspondiente a una variación en la secuencia, caracterizar el cambio mediante secuenciación directa de un nuevo producto de PCR.

PROTOCOLO XI- Genotipado mediante sondas FRET**Materiales**

- Placas blancas de 384 pocillos (Roche Diagnostics)
- Film adherente
- Oligonucleótidos cebadores a 10 μ M
- Sondas (*Anchor* y *Sensor*) a 4 μ M
- LightCycler® 480 Genotyping Master Mix 5X (Roche Diagnostics)
- MgCl₂ a 25 mM
- gDNA a 10 ng/ μ l

- 1- Preparar la mezcla siguiente para todas las muestras, teniendo en cuenta un volumen final de 5 μ l.
 - a- LightCycler 480 Genotyping Master Mix 5X a concentración final 1X.
 - b- MgCl₂, cuando sea necesario, hasta una concentración final de 1,5-4 mM (incluyendo el que incorpora la Master Mix).
 - c- Oligonucleótidos cebadores a concentración final de 0,5 μ M.
 - d- Sonda de hibridación *Sensor* (concentración final 0,12 μ M).
 - e- Sonda de hibridación *Anchor* (concentración final 0,12 μ M).
 - f- dH₂O hasta alcanzar un volumen de 5 μ l para cada muestra a amplificar.

- 2- Repartir la mezcla anterior en los pocillos de la placa (4 µl / pocillo) y añadir a cada pocillo 10 ng del DNA correspondiente (1 µl).
- 3- Sellar cuidadosamente las placas con el film adherente.
- 4- Centrifugar la placa durante 2 minutos a 3.500 x g.
- 5- Introducir las placas en el termociclador LC 480 y programarlo específicamente para cada fragmento (tabla M-5), siguiendo este esquema general:

Programa	Ciclos	Subprograma	T°C	Tiempo	Rampa (°C/s)	(Adq/°C) ^a	Modo Adq. ^b
Desnaturalización	1	Desnaturalización	5	10 min	4,8	-	-
		Desnaturalización	95	10-15 s	4,8	-	-
Amplificación	35	Anillamiento	50-65 ^c	20-30 s	2	-	Single
		Elongación	72	X s ^f	4,8	-	-
		Desnaturalización	95	10-15 s	4,8	-	-
Curva de Fusión	1	Re-anillamiento	40-50 ^d	30 s	2,5	-	-
		Desnatulalización	80-95 ^e	0 s	-	1-5	Continuos
Enfriamiento	1	Enfriamiento	40	20 s	4,8	-	-

a, n° de adquisiciones de fluorescencia por °C. b, modo de realizar las adquisiciones. c, T^a de anillamiento específica de los cebadores usados en la amplificación. d, Las T^a de re-anillamiento y de desnaturalización dependen del fragmento amplificado. e, varían en función del tamaño del fragmento a amplificar.

PROTOCOLO XII- Genotipado mediante sondas TaqMan[®]

Materiales

- Placas de 384 pocillos, MicroAmp[®] Optical 384-Well Reaction Plate (Applied Biosystems)
- Film adherente
- 20X o 40X TaqMan[®] SNP Genotyping Assays (Applied Biosystems). Contiene sondas TaqMan específicas para cada alelo del polimorfismo (VIC y FAM) y los cebadores correspondientes.
- TaqMan[®] Universal PCR Master Mix (2X) (Applied Biosystems)
- gDNA a 10 ng/µl

- 1- Preparar la mezcla siguiente de los componentes de la reacción para todas las muestras a amplificar, teniendo en cuenta un volumen final de 5 µl para las placas de 384 pocillos.
 - a- Sondas TaqMan y cebadores a concentración final 1X
 - b- 2,5 µl de TaqMan[®] Universal PCR Master Mix (concentración final 1X)
 - c- dH₂O hasta alcanzar un volumen de 5 µl/muestra
- 2- Agitar suavemente y hacer un pulso de centrifuga.
- 3- Repartir la mezcla en los pocillos (4 µl/pocillo) de la placa.
- 4- Añadir a cada pocillo 10 ng de DNA molde (1 µl).
- 5- Sellar las placas con el film adherente y centrifugar las placas 2 min a 3.500 x g.
- 6- Introducir las placas en un termociclador de PCR a tiempo real (ABI PRISM 7900HT) donde se producirá la reacción específica de amplificación. Programar el termociclador siguiendo este esquema:

Programa	Ciclos	Subprograma	T°C	Tiempo
Preincubación	1	Preincubación	50	2 min
Activación de la enzima	1	Desnaturalización	95	10 min
		Desnaturalización	95	15 s
Amplificación	45	Anillamiento-Elongación	60	1 min 20 s

- 7- Una vez acabado el programa anterior, realizar la lectura de la fluorescencia de cada pocillo para poder analizar la discriminación alélica.

PROTOCOLO XIII- Análisis de las variantes polimórficas con HRM

Materiales

- Placas blancas de 384 pocillos (Roche Diagnostics)
- Film adherente
- Oligonucleótidos cebadores a 10 μ M
- LightCycler® 480 High Resolution Melting Master Mix 2X (Roche Diagnostics)
- MgCl₂ a 25 mM
- gDNA a 10 ng/ μ l

- 1- Preparar la mezcla siguiente para todas las muestras, teniendo en cuenta un volumen final de 10 μ l.
 - a- LightCycler 480 High Resolution Melting Master Mix 2X a concentración final 1X.
 - b- MgCl₂, cuando sea necesario, hasta una concentración final de 1-3,5 mM.
 - c- Oligonucleótidos cebadores a concentración final de 0,2-0,3 μ M.
 - d- dH₂O hasta alcanzar un volumen de 10 μ l para cada muestra a amplificar.
- 2- Repartir la mezcla anterior en los pocillos de la placa (9 μ l / pocillo).
- 3- Añadir a cada pocillo 10-25 ng del DNA correspondiente.
- 4- Sellar cuidadosamente las placas con el film adherente.
- 5- Centrifugar la placa durante 2 minutos a 3.500 x g.
- 6- Introducir las placas en el termociclador LC 480 y programarlo específicamente para cada fragmento (ver APARTADO 7.3.3 de Materiales y Métodos), siguiendo este esquema general:

Programa	Ciclos	Subprograma	T°C	Tiempo	Rampa (°C/s)	(Adq/°C) ^a	Modo Adq. ^b
Desnaturalización	1	Desnaturalización	95	10 min	4,8	-	-
		Desnaturalización	95	10 s	4,8	-	-
Amplificación	40	Anillamiento	50-65 ^c	20-30 s	2	-	-
		Elongación	72	15 s	4,8	-	Single
		Desnaturalización	95	1 min	4,8	-	-
Curva de Fusión	1	Re-anillamiento	40	1 min	1,5	-	-
			65	1 s	1	-	-
		Desnatulización	95	0 s	-	15-25	Continuos
Enfriamiento	1	Enfriamiento	40	20 s	2,5	-	-

a, n° de adquisiciones de fluorescencia por °C. b, modo de realizar las adquisiciones. c, T^a de anillamiento específica de los cebadores usados en la amplificación.

7- Una vez finalizado el programa anterior, realizar el análisis de la curva de fusión con el LightCycler® 480 HRM Gene Scanning Software.

PROTOCOLO XIV- Defosforilación de un vector

Materiales

- *Calf intestinal alkaline phosphatase*, CIP, 1 U/μl (Roche)
- Tampón de defosforilación 10X (0.5 M Tris-HCl, 1 mM EDTA) (Roche)
- EDTA 5 mM a pH 8,0

- 1- Digerir el plásmido de DNA circular con la enzima de restricción apropiada durante 2 horas y comprobar el resultado de la digestión mediante electroforesis en un gel de agarosa al 1% y tinción con bromuro de etidio.
- 2- Cuando la digestión sea completa, purificar el DNA mediante extracción de banda de un gel de agarosa donde se haya cargado la totalidad de la reacción. Eluir el DNA en 30 μl de agua bidestilada.
- 3- Mezclar en un tubo de 0,2 ml:
 - a- 20 μl del plásmido linearizado anterior
 - b- 6 μl de la fosfatasa CIP
 - c- 4 μl de tampón de defosforilación correspondiente
 - d- Hasta 40 μl de agua bidestilada dH₂O
- 4- Incubar 2 horas a 37°C.
- 5- Inactivar la actividad fosfatasa de CIP al final del periodo de incubación añadiendo 0,4 μl de EDTA (pH 8,0) 5 mM.
- 6- Incubar durante 15 minutos a 65°C.

- 7- Enfriar la mezcla de reacción a temperatura ambiente y purificar el DNA plásmidico a partir de extracción de banda de gel de agarosa previamente a la reacción de ligación. Eluir en 30 µl de dH₂O.
- 8- Realizar alícuotas y conservar a -20°C.

PROTOCOLO XV- Ligación

Materiales

- Ligasa de DNA T4, 3 U/µl (New England Biolabs)
- Tampón de ligasa T4 10X (New England Biolabs)
- Vector digerido y purificado
- Inserto digerido y purificado

Para calcular la cantidad apropiada de producto de PCR se utilizó la siguiente ecuación:

$$\text{ng inserto} = (\text{ng del vector} \times \text{kb tamaño inserto} \times \text{inserto:vector molar ratio}) / \text{kb tamaño del vector}$$

- 1- Centrifugar brevemente los tubos que contienen el inserto y el vector a ligar.
- 2- Vortear vigorosamente el tampón de la ligasa antes de cada uso. Este contiene ATP, que se degrada durante las fluctuaciones de temperatura, por lo que se realizan alícuotas de un sólo uso para evitar congelar-descongelar múltiples veces y exponerlo a frecuentes cambios de temperatura.
- 3- Llevar a cabo las reacciones de ligación estándar en tubos de 0,2 ml:
 - a- 1 µl del tampón de ligación (10X).
 - b- El volumen necesario de vector para añadir 50 ng del mismo.
 - c- El volumen necesario del fragmento a clonar para obtener los ng calculados anteriormente.
 - d- dH₂O hasta alcanzar un volumen de 10 µl.
 - e- 1 µl de ligasa (3 U/µl).
- 4- Realizar tres reacciones control adicionales en las que faltará alguno de los componentes anteriores:
 - a- en una no añadir fragmento
 - b- en otra no añadir vector
 - c- no añadir ligasa
- 5- Incubar a 16 °C durante 12-16 h.
- 6- Congelar la reacción a -20°C hasta que se proceda a la transformación

PROTOCOLO XVI- Preparación de células competentes de E.coli por el método del cloruro cálcico

Materiales

- Células E.Coli XL 1-Blue
- LB (1% (p/v) bacto triptona, 0,5% (p/v) extracto de levadura y 0,5 % (p/v) de NaCl)
- LB-Tetraciclina

- ◆ Tetraciclina (50 µg/ml en etanol)
 - CaCl₂ 50 mM estéril
 - Glicerol estéril 100%
- 1- Sembrar una estría de E.coli XL1-Blue en una placa de LB-Tetraciclina. Incubar a 37°C 14-16 horas.
 - 2- Inocular una colonia aislada de E.coli XL1-Blue a 2,5 ml de medio LB-Tetraciclina. Incubar a 37°C a 250 rpm durante 10-14 horas.
 - 3- Medir la absorbancia del cultivo a una longitud de onda de 600 nm (A₆₀₀) de una dilución 1:10 y comprobar que sea menor que 2.
 - 4- Centrifugar a 13000 rpm durante 1 minuto.
 - 5- Resuspender el sedimento celular en 2,5 ml de medio LB-Tetraciclina fresco. Mantener las células resuspendidas en hielo picado.
 - 6- Atemperar 80 ml de medio LB en un frasco tipo Erlenmeyer (300-500ml) a 37°C.
 - 7- Inocular los 80 ml de medio LB con 0.8 ml de las células resuspendidas en el paso 5 (1:100). Incubar a 37°C a 250 rpm.
 - 8- Medir la A₆₀₀ a intervalos de 60 min.
 - 9- Incubar el cultivo en hielo picado 1 hora a 4°C cuando el crecimiento bacteriano alcance una densidad óptica de 0,4-0,5 (aproximadamente, 3 horas). En este punto el cultivo de encuentra en la fase de crecimiento exponencial. A partir de este momento es muy importante no romper la cadena de frío y seguir operando en condiciones de esterilidad.
 - 10- Resuspender las células suavemente con vórtex en la mitad del volumen inicial (40 ml) de cultivo de una solución acuosa 50 mM CaCl₂ estéril y fría. Incubar en hielo picado 15 minutos.
 - 11- Centrifugar la suspensión de células a 2000 rpm y a 4°C durante 5 min. Descartar el sobrenadante.
 - 12- Resuspender las células suavemente con vórtex en 1/15 (5,5 ml) del volumen inicial de cultivo de una solución acuosa 50 mM CaCl₂ estéril y fría. Incubar en hielo picado 1-2h.
 - 13- Incubar la suspensión celular a 4°C 12-24 horas. Esta incubación aumenta en 4-6 veces la eficiencia de transformación de las células competentes.
 - 14- Repartir la suspensión de células en alícuotas y añadiendo glicerol estéril a una concentración final del 10-12% (v/v). Homogeneizar la suspensión.
 - 15- Conservar los gliceroles a -80°C.

PROTOCOLO XVII- Transformación de células competentes mediante choque térmico

Materiales

- Células E.Coli competentes por choque térmico
- Medio LB (1% (p/v) bacto triptona, 0,5% (p/v) extracto de levadura y 0,5 % (p/v) de NaCl)

- Medio SOC (2% (p/v) bacto triptona, 0,5% (p/v) extracto de levadura, 0,05 g (p/v) NaCl, 2,5 mM KCl, 20 mM MgCl₂ y 20mM glucosa)
- Placas LB-Antibiótico selectivo
 - ♦ LB + 1,5% (p/v) agar
 - ♦ Ampicilina 100 µg/ml en dH₂O o Kanamicina 50 µg/ml en dH₂O
- Placas LB-Ampicilina/IPTG/X-Gal
 - ♦ LB + 1,5% (p/v) agar
 - ♦ Ampicilina 100 µg/ml en dH₂O
 - ♦ 50 µl de IPTG 200 mM
 - ♦ 20 µl de X-Gal a 50 mg/µl en dimetilformamida
- Plásmido control
- Plásmido de interés

La manipulación de las células competentes y las placas de cultivo siempre debe realizarse en condiciones estériles. Para la transformación con plásmidos procedentes de preparaciones plasmídicas, transformar 50 µl de células competentes con 1-20 ng de plásmido y preparar 2 placas por transformación, donde se siembra en cada una de ellas 1/10 de las bacterias transformadas.

- 1- Centrifugar los tubos de la reacción de ligación.
- 2- Sacar las células del congelador de -80°C y colocarlas en hielo. Dejarlas descongelar aproximadamente 5 minutos. Mezclar las células moviendo suavemente el tubo evitando pipetear excesivamente, ya que las células competentes son extremadamente frágiles.
- 3- Añadir 5 µl de cada reacción de ligación en un tubo de microcentrífuga estéril de 1,5 ml en hielo. Preparar otro tubo en hielo con 0,2 ng de un plásmido control para determinar la eficiencia de transformación.
- 4- Cuidadosamente transferir 50 µl de las células a los tubos preparados con la reacción de ligación y 100 µl de células al tubo con el plásmido no cortado para la determinación de la eficiencia de la transformación.
- 5- Agitar suavemente los tubos y dejarlos en hielo 20 minutos.
- 6- Choque térmico: dejar las células 45 segundos exactamente en un baño húmedo a 42°C, no agitar.
- 7- Inmediatamente poner los tubos en hielo durante 2 minutos.
- 8- Añadir 950 µl de medio SOC, a temperatura ambiente, a los tubos con las células transformadas con la reacción de ligación y 900 µl al tubo con las células transformadas con el plásmido sin cortar. Se puede utilizar también medio LB, aunque el número de colonias puede ser menor.
- 9- Incubar 1,5 horas a 37°C con agitación a 250 rpm.
- 10- Preparar placas de LB-antibiótico o LB-Ampicilina/IPTG/X-Gal, si se utiliza el vector pGEM, tal y como se especifica en el siguiente listado, y equilibrarlas a temperatura ambiente antes de plaquar:
 - 2 X cada reacción estándar
 - 2 X cada reacción estándar diluída 1:10
 - 2 X cada reacción estándar concentrada X2
 - 2 X cada uno de los 3 controles de la reacción de ligación

- 2 X control de transformación
- 11- Sembrar 100 µl de cada cultivo de transformación por duplicado en las placas. Para el control de transformación sembrar una dilución 1:10. Si se requiere un mayor número de colonias, las células deben centrifugarse a velocidad máxima durante 1 minuto, resuspenderlas con la mitad de volumen aproximadamente (concentración 2x) y sembrar 100 µl por duplicado.
 - 12- Incubar las placas o/n (16-24 h) a 37°C. En el caso de la placas LB-Antibiótico/IPTG/ X-Gal utilizar incubaciones más largas o a 4°C (después de la incubación a 37°C o/n) para facilitar la aparición de colonias azules.
 - 13- Comprobar los plásmidos recombinantes mediante PCR de cribaje y realizar minipreparaciones plásmidicas de las colonias.

PROTOCOLO XVIII- PCR de cribaje

Materiales

- Tampón de PCR (10X) (100 mM Tris-HCl, 15 mM MgCl₂, 500 mM KCl, pH 8,3)
 - dNTPs (dTTP, dATP, dCTP, dGTP) a 2 mM
 - Oligonucleótidos cebadores a 10 µM
 - Taq DNA polimerasa (5 U/µl)
 - LB-Antibiótico
 - ♦ 1% (p/v) bacto triptona, 0,5% (p/v) extracto de levadura y 0,5 % (p/v) de NaCl
 - ♦ Ampicilina 100 µg/ml en dH₂O o Kanamicina 50 µg/ml en dH₂O
 - Glicerol estéril al 100%
- 1- Transformar el producto de ligación tal y como se describe en el PROTOCOLO XVIII.
 - 2- Dispensar 150 µl de LB selectivo, con el correspondiente antibiótico al que el plásmido tenga resistencia, en una placa estéril de 96 pocillos.
 - 3- Picar 30-40 colonias, con una punta estéril, e inocularlas en los 150 µl de LB selectivo. Si las colonias han crecido en placas LB-Ampicillin/IPTG/ X-Gal picar 20 colonias blancas y una azul.
 - 4- Incubar a 37°C durante 1 h.
 - 5- Preparar la mezcla siguiente para todas las muestras, calculada para un volumen final de 25 µl, y dispensarla en tubos de 0,2 ml. La mezcla se debe mantener en hielo hasta que se coloquen los tubos en el termociclador:
 - a- 2.5 µl de tampón de reacción de la polimerasa (10X).
 - b- 1 µl de dNTPs 2mM.
 - c- 1 µl de cada primer a 10 µM.
 - d- 0.1 µl de Taq DNA polimerasa 5 U/µl.
 - e- el volumen de dH₂O necesario para completar 23 µl.
 - 6- Añadir en condiciones estériles 2 µl del cultivo anterior a los tubos preparados con la mezcla de PCR. La placa de 96 pocillos con el cultivo restante se conserva a 4°C mientras se realiza la reacción de PCR.

- 7- Programar el termociclador con las condiciones siguientes y colocar los tubos cuando la temperatura supere los 80°C:
 - Desnaturalización a 95°C 4 min
 - 30 ciclos de desnaturalización a 95°C 30 s, anillamiento a la temperatura específica para cada cebador 30-60 s y extensión a 72°C 30 s
 - Extensión final: 72°C 4 min
- 8- Visualizar las colonias positivas cargando 10 µl de la reacción de PCR anterior mezclada con el tampón de carga (6x), a una relación 6:1, en un gel de agarosa al 1-2% y posterior tinción con bromuro de etidio.
- 9- Inocular los 150 µl de los cultivos anteriores en 5 ml de LB-antibiótico, de las colonias positivas y de una negativa. Incubar a 37°C a 225-250 rpm durante 12-16 h.
- 10- Preparar gliceroles al 15% de cada uno de los cultivos en criotubos de 2 ml (150 µl glicerol estéril + 850 µl de cultivo líquido), congelar inmediatamente en nieve carbónica y conservar a -80°C.
- 11- Del cultivo restante, realizar minipreparaciones plásmidicas (ver PROTOCOLO XXI, a partir del paso 3) y secuenciarlas para comprobar el plásmido recombinante.

PROTOCOLO XIX- Mutagénesis dirigida

Materiales

- Pfu Turbo (2,5 U/µl)
 - Tampón Pfu Turbo 10X
 - Cebadores de mutagénesis directo y reverso 10 µM
 - dNTPs (dTTP, dATP, dCTP, dGTP) a 2mM
 - Plásmido original a mutagenizar
 - DMSO
 - Enzima Dpnl (10 U/µl)
 - Reactivos necesarios para realizar la transformación bacteriana por choque térmico
- 1- Preparar la siguiente mezcla de componentes en un tubo de 0,2 ml para cada reacción de mutagénesis:
 - a- 1,25 µl del tampón Pfu Turbo 10X.
 - b- 1,25 µl de dNTPs 2mM.
 - c- 0,125 µl de cada uno de los cebadores a 10 µM.
 - d- 0,375 µl de DMSO.
 - e- 2-5 ng del plásmido a mutagenizar.
 - f- 0,25 de la enzima Pfu Turbo 2,5 U/µl.
 - g- El volumen necesario de dH₂O hasta un volumen de 12,5 µl.
 - 2- Realizar la reacción de PCR de mutagénesis en un termociclador programado tal y como se describe en la siguiente tabla. Tanto el número de ciclos como el tiempo de extensión se ajustan dependiendo del tamaño del plásmido (el tiempo) y del tipo de mutación a introducir (el número de ciclos). Así el número de ciclos es 12 para mutaciones puntuales y 16 para inserciones o deleciones.

Programa	Ciclos	Subprograma	T°C	Tiempo
Desnaturalización inicial	1	Desnaturalización inicial	94°C	1 min 30 s
		Desnaturalización	94°C	30 s
Amplificación	12 o 16	Anillamiento	55°C	30 s
		Extensión	68°C	2 min/kb
Extensión final	1	Extensión final	68°C	4 min 30 s
Enfriamiento	1	Enfriamiento	15°C	0 s

- 3- Una vez finalizada la reacción y enfriada, añadir 1 µl de la enzima DpnI para digerir el plásmido no mutado.
- 4- Incubar 1 hora a 37°C en agitación.
- 5- Separar una alícuota de 2,5 µl y visualizarla en un gel de agarosa al 1%, para comprobar la reacción de mutagénesis y la digestión del plásmido parental.
- 6- Transformar 5 µl o 10 µl de la mezcla restante en 100 µl de células competentes de *E. coli* por choque térmico. Sembrar la totalidad de las células transformadas en una placa de LB-agar-antibiótico.
- 7- Cultivar en 5 ml de LB-antibiótico unas 5 colonias de cada placa a 37 °C 12-16 h, a 250 rpm.
- 8- Del cultivo líquido anterior, realizar:
 - a- Glicerol al 15%: 400 µl de cultivo en 800 µl de glicerol estéril al 60% y congelar a -80 °C
 - b- Minipreparaciones de DNA
- 9- Comprobar las minipreparaciones que hayan incorporado la mutación deseada mediante secuenciación directa del fragmento que debe contener la mutación.
- 10- De las minipreparaciones positivas, secuenciar el inserto completo con el fin de comprobar que no se ha introducido otra mutación no deseada por error de la polimerasa.

PROTOCOLO XX- Minipreparación de DNA plasmídico

Materiales

- QIAprep Spin Miniprep kit (Qiagen)
- Colonias bacterianas transformadas en placas selectivas LB-agar-Antibiótico
- LB-Antibiótico
 - ◆ 1% (p/v) bacto triptona, 0,5% (p/v) extracto de levadura y 0,5 % (p/v) de NaCl
 - ◆ Ampicilina 100 µg/ml en dH₂O o Kanamicina 50 µg/ml en dH₂O
- Glicerol estéril al 100%
- Reactivos necesarios para realizar electroforesis horizontal de ácidos nucleicos en agarosa

- 1- Inocular una colonia en 5 ml de LB líquido-Antibiótico. Incubar durante 12-16 h a 37 °C y a 250 rpm.
- 2- Preparar glicerol al 15% de cada cultivo (150 µl de glicerol al 100% + 850 µl del cultivo líquido).

- 3- Del cultivo restante, repartir 3 ml en 2 tubos y centrifugar 1 minuto a 13200 rpm. Eliminar totalmente el sobrenadante.
- 4- Resuspender el precipitado de bacterias de uno de los tubos con 125 µl de tampón P1. Pasar el líquido al otro tubo que contiene el precipitado del mismo cultivo y añadir otros 125 µl de tampón P1. Las bacterias deben resuspenderse completamente vorteadando o pipeteando arriba y abajo hasta que no se observen grumos.
- 5- Añadir 250 µl de tampón P2 e invertir suavemente el tubo 10 veces. Colocar los diversos tubos en una gradeta y realizar la inversión conjuntamente. No vortear ya que podría degradarse el DNA genómico. Si es necesario continuar invirtiendo el tubo hasta que la solución quede viscosa y ligeramente clara. No dejar la reacción de lisis más de 5 minutos.
- 6- Añadir 350 µl de tampón N3 y mezclar inmediatamente invirtiendo el tubo 10 veces para evitar precipitaciones. En este punto la solución debe presentar un aspecto turbio.
- 7- Centrifugar 10 minutos a 13000 rpm. Se formará un precipitado blanco compacto.
- 8- Pasar el sobrenadante a las columnas QIAprep por decantación o pipeteo.
- 9- Centrifugar 1 minuto a 13000 rpm. Descartar el eluido.
- 10- Lavar la columna QIAprep añadiendo 0,5 ml de tampón PB y centrifugar 1 minuto a 13000 rpm. Descartar el eluido.
- 11- Lavar la columna QIAprep añadiendo 0,75 ml de tampón PE y centrifugar 1min. Descartar el eluido y centrifugar 1 minuto adicional a 13000 rpm para eliminar totalmente el tampón de lavado, ya que los residuos de etanol pueden inhibir las consiguientes reacciones enzimáticas.
- 12- Colocar las columnas en un tubo de 1,5 ml. Añadir 30 µl de dH₂O en el centro de la columna, dejar 5 minutos a temperatura ambiente y eluir el plásmido centrifugando los tubos 1,5 minutos a 13000 rpm.
- 13- Incubar los tubos durante 10 min a 65°C para resuspender totalmente el plásmido.
- 14- Cuantificar la minipreparación por espectrofotometría y comprobar la calidad de la misma mediante electroforesis en un gel de agarosa al 1%.

PROTOCOLO XXI- Midi y Maxipreparacion de DNA plasmídico

Materiales

- Placas selectivas LB-Agar-Antibiótico
- Tampones P1, P2, P3, QBT, QC y QF (incluidos en el kit)
- LB-Antibiótico
 - ◆ 1% (p/v) bacto triptona, 0,5% (p/v) extracto de levadura y 0,5 % (p/v) de NaCl
 - ◆ Ampicilina 100 µg/ml en dH₂O o Kanamicina 50 µg/ml en dH₂O
- Isopropanol
- Etanol 70 %
- Reactivos necesarios para realizar electroforesis horizontal de ácidos nucleicos en agarosa

El protocolo para midi y maxipreparaciones es exactamente el mismo, diferenciándose únicamente en el volumen de reactivos utilizados en cada caso. Los volúmenes de los reactivos

que difieren en los dos tipos de preparaciones plásmidicas se representan en el protocolo de la siguiente manera: Volumen midipreparación/Volumen maxipreparación

- 1- Obtener colonias de células competentes transformadas con el plásmido de interés, o sembrar una estría en una placa LB-Antibiótico desde el glicerol almacenado a -80°C.
- 2- Picar una colonia aislada e inocular un cultivo de 2-5 ml de LB con el antibiótico selectivo apropiado. Incubar durante 8 horas, aproximadamente, a 37°C a 250 rpm.
- 3- Diluir el cultivo inicial 1:500 o 1:1000 en 100/500 ml de LB selectivo. Incubar a 37°C durante 12-16 horas a 250 rpm.
- 4- Repartir el cultivo en tubos de 50 ml y centrifugar durante 15 minutos a 5000 x g durante 15 minutos. Decantar el sobrenadante.
- 5- Resuspender las células bacterianas, vorteadando o pipeteando, en 4/10 ml del tampón P1 de lisis hasta que la solución sea homogénea. Juntar el resuspendido del mismo cultivo en un sólo tubo de 50 ml.
- 6- Añadir 4/10 ml de tampón P2, e invertir los tubos 10 veces. Incubar a temperatura ambiente durante 5 minutos máximo. En este punto la mezcla adquiere una apariencia viscosa. Es importante no vortear en este paso para evitar la degradación del DNA.
- 7- Añadir 4/10 ml de tampón P3, previamente enfriado a 4°C. Invertir 10 veces los tubos e incubar en hielo picado durante 15-20 minutos. En este punto se forman el precipitado de DNA genómico, proteínas y otros componentes celulares.
- 8- Centrifugar a 5000 x g durante 30 minutos a 4°C. Recuperar inmediatamente el sobrenadante en un nuevo tubo por decantación y descartar el anterior.
- 9- Centrifugar el sobrenadante otra vez a 5000 x g durante 15 minutos a 4°C. Volver a recuperar el sobrenadante.
- 10- Equilibrar las columnas proporcionadas en el kit con 4/10 ml de tampón QBT. Permitir que la columna se vacíe y seque por gravedad.
- 11- Aplicar el sobrenadante del paso 8 a la columna y permitir su paso por la resina por gravedad.
- 12- Lavar dos veces la columna con 10/30 ml de tampón QC cada lavado.
- 13- Eluir el DNA con 5/15 ml de tampón QF. En este paso se puede parar el protocolo y conservar el eluido a 4°C hasta el día siguiente.
- 14- Precipitar el DNA añadiendo 0,7 volúmenes de isopropanol (3,5/19,5 ml) a temperatura ambiente al DNA eluido. Mezclar por inversión y centrifugar durante 1 hora a 5000 x g a 4°C. Decantar el sobrenadante, invirtiendo el tubo por el lado opuesto a donde se encuentre el precipitado.
- 15- Lavar la el precipitado de DNA con 2/5 ml de etanol al 70%, y centrifugar a 5000 x g durante 30 minutos. Decantar el sobrenadante cuidadosamente sin mover el sedimento.
- 16- Secar el precipitado a temperatura ambiente durante 5-10 minutos y disolverlo en 100/500 µl de dH₂O.
- 17- Acabar de resuspender el plásmido incubándolo 15 minutos a 65°C.

- 18- Cuantificar las preparaciones por espectrofotometría y comprobar la calidad mediante electroforesis en un gel de agarosa al 1%.

PROTOCOLO XXII- Tripsinización y subcultivo de células adherentes

Materiales

- Tampón PBS (*Phosphate-buffered saline*) 1X sin Ca^{2+} ni Mg^{2+} , a 37° C
- Tripsina 0,25% (1X)-EDTA, a 37° C
- Medio DMEM o RPMI 1640 completo, a 37° C
 - ◆ 10% FBS (suero fetal bovino)
 - ◆ 100 U/ml penicilina y 100 µg/ml estreptomycin

- 1- Eliminar el medio del cultivo subconfluyente con una pipeta Pasteur estéril conectada al vacío.
- 2- Lavar la monocapa de células adherentes con un pequeño volumen de PBS (4 ml) para eliminar el suero residual que podría inhibir la acción de la tripsina. Utilizar PBS libre de Ca^{2+} y Mg^{2+} , ya que estos componentes podrían causar la agregación de las células.
- 3- Añadir 1 ml de tripsina/EDTA. Mover solución para que cubra toda la monocapa.
- 4- Colocar la placa en un incubador a 37°C durante 2-3 minutos. Golpear suavemente la parte inferior de la placa para desprender las células. Comprobar en el microscopio invertido que las células estén desenganchadas de la placa. Si no lo están, incubar la placa a 37°C 1-2 minutos adicionales.
- 5- Añadir 4 ml del medio apropiado y pipetear la suspensión de células.
- 6- Añadir el volumen necesario de la suspensión celular y medio apropiado, en placas de 100 mm, para realizar la dilución necesaria (1/3-1/10). Alternativamente las células pueden ser contadas, utilizando un hemacitómetro Neubauer, y diluirlas a la concentración deseada.
- 7- Si es necesario, reemplazar el medio de los cultivos por medio fresco completo hasta que el cultivo alcance la subconfluencia, momento en que se repetirá el proceso.

Recuento del número de células en una cámara Neubauer

Materiales

- Etanol al 70% (v/v)
- Hemacitómetro Neubauer y cubreobjetos (Baxter Scientific)
- Reactivos necesarios para la tripsinización

- 1- Lavar la superficie del hemacitómetro y el cubreobjetos con etanol al 70%. Dejar secar.
- 2- Colocar el cubreobjetos sobre los surcos del hemacitómetro, de manera que cubra el área central.
- 3- Tripsinizar las células.
- 4- Resuspender las células en medio completo, 5-10 ml de medio para cultivos subconfluentes, hasta alcanzar una suspensión uniforme.

- 5- Cargar 10 µl de la suspensión celular en cada extremo del cubreobjetos. Sostener la punta de la pipeta por debajo del cubreobjetos y dispensar, gota a gota, la suspensión, que entrará en la cámara por capilaridad.
- 6- Colocar el hemacitómetro en el microscopio y realizar el recuento con una magnificación de 400X (objetivo 40X).
- 7- Contar, como mínimo, 2 cuadrículas, de 1x1 mm, de 16 cuadrados, de cada una de las 2 cámaras y calcular el promedio (nº total de células/ nº de cuadrículas contadas)
- 8- Calcular el número de células de la suspensión:
 - c- Concentración en la suspensión (células / ml) = promedio de células x 10000
 - d- Número total de células = células/ml x volumen total de la suspensión de células de la cual se tomó la muestra.

PROTOCOLO XXIII- Congelación y descongelación de células adherentes

Materials

- Tripsina 0,25% (1X)-EDTA, a 37°C
- Tampón PBS (Phosphate-buffered saline) 1X sin Ca²⁺ ni Mg²⁺, a 37°C
- Medio DMEM o RPMI 1640 completo, a 37°C
 - ♦ 10% FBS (suero fetal bovino)
 - ♦ 100 U/ml penicilina y 100 µg/ml estreptomina
- FBS (Suero Fetal Bovino), a 37°C
- DMSO
- 5100 Cryo 1°C Freezing Container (Nalgene)
- Isopropanol al 100%

Protocolo para la congelación

- 1- Tripsinizar las células de la placa de cultivo (ver PROTOCOLO XXIII, pasos 1-4)
- 2- Añadir 4 ml de medio completo, pipetear la suspensión de células sobre la placa, y transferir todo el volumen a un tubo estéril de 15 ml.
- 3- Centrifugar 5 minutos a 1000 rpm, a temperatura ambiente.
- 4- Eliminar el sobrenadante con una pipeta Pasteur conectada al vacío.
- 5- Añadir 4,5 ml de FBS y resuspender las células.
- 6- Añadir 500 µl de DMSO y mezclar rápidamente la suspensión, ya que el DMSO es altamente tóxico para las células.
- 7- Repartir, inmediatamente, 1 ml de la suspensión anterior en criotubos de 2 ml y colocar rápidamente en el contenedor de congelación, que se conservará a -80°C. El contenedor de congelación debe haber sido llenado previamente con isopropanol al 100% que se reemplaza cada 3 tandas de congelación.

Protocolo para la descongelación

- 1- Recuperar el vial que contiene las células criopreservadas y descongelarlo manualmente por fricción, durante, aproximadamente, 1 minuto.
- 2- Cuando la suspensión de células empiece a descongelarse, transferir el contenido del vial a un tubo estéril de 15 ml con 5 ml de medio completo atemperado a 37°C.
- 3- Centrifugar el tubo durante 5 minutos a 1000 rpm y eliminar el sobrenadante.
- 4- Resuspender las células con 3 ml de medio fresco completo y transferir la suspensión a una placa de 100 mm de diámetro con 7 ml de medio completo adicionales.
- 5- Comprobar el cultivo cada 24 horas para asegurarse que las células quedan adheridas en la placa.
- 6- Cambiar el medio cada 2-3 días hasta que las células alcancen la subconfluencia.

PROTOCOLO XXIV- Transfección de células con lípidos catiónicos**Materiales**

- Lipofectamine™ 2000 Reagent (Invitrogen)
- DNA plasmídico (200-1200 ng/μl)
- Línea celular y medio de crecimiento apropiado.
- PBS

- 1- Un día antes de la transfección cultivar las células en placas de 24 pocillos en 500 μl del medio apropiado, de manera que alcancen el 80-90% de confluencia en el momento de la transfección. Incubar las placas a 37°C y 5% CO₂ durante 24 horas, hasta alcanzar la confluencia deseada.
- 2- Lavar las células con PBS, y añadir 500 μl de medio fresco sin antibiótico ni suero.
- 3- Preparar los complejos para cada transfección de la manera siguiente:
 - a- Diluir la cantidad de DNA apropiada para cada condición experimental en 50 μl de medio sin suero ni antibiótico.
 - b- Mezclar vigorosamente el vial que contiene la lipofectamina 2000 y diluir 1,5 μl del reactivo en 50 μl de medio sin suero ni antibiótico. Incubar durante 5 minutos a temperatura ambiente
 - c- Combinar el DNA diluido con la lipofectamina, preparados en el paso 3a y 3b. Mezclar suavemente e incubar durante 20 minutos.
- 4- Añadir los 100 μl de los complejos a los pocillos de la placa correspondiente a cada condición experimental, y mezclar moviendo la placa adelante y atrás.
- 5- Incubar la placa a 37°C y 5 % CO₂ durante 24 horas. El medio debe ser substituido por 500 μl de medio fresco 4-6 horas después de añadir los complejos.
- 6- A las 24 horas llevar a cabo los ensayos luciferasa.

PROTOCOLO XXV- Ensayos luciferasa dual

Materiales

- Placa de 24 pocillos con células transfectadas con los vectores reporteros (luciferasa *Firefly* y *Renilla*)
- Tampón PBS (Phosphate-buffered saline) 1X sin Ca^{2+} ni Mg^{2+} , a 37 °C
- Dual-Luciferase® Reporter Assay System (Promega):
 - ◆ Tampón de lisis (*Passive Lysis Buffer*, PLB)
 - ◆ Preparar el reactivo de la luciferasa (Luciferase Assay Reagent II, LARII), resuspendiendo el sustrato liofilizado en 10 ml del tampón Luciferase Assay Buffer II. Realizar alícuotas de 1 ml.
 - ◆ Preparar el volumen adecuado del reactivo Stop&Glo® para el número de muestras a cuantificar. Añadir el sustrato al tampón Stop&Glo® en una relación 1/55
- Tubos para luminómetro de poliestireno 11,5x55 mm.

- 1- Aspirar completamente el medio de las células y lavarlas con PBS.
- 2- Añadir 150 µl de PLB 1X a cada pocillo, y lisar las células incubándolos en agitación a temperatura ambiente durante 15 minutos. Mantener las placas en hielo picado.
- 3- Ajustar los parámetros de cuantificación de la luminiscencia en un luminómetro.
- 4- Sensibilidad de la lectura, depende de la actividad luciferasa del ensayo (40-80%).
- 5- Tiempo de lectura de la luminiscencia (10 s).
- 6- Atemperar los reactivos LARII y Stop&Glo®
- 7- Transferir 10 µl del extracto celular lisado a un tubo de poliestireno y colocar el tubo en la cámara de lectura del luminómetro.
- 8- Añadir 40 µl del reactivo LARII, pipetear 2 veces y realizar la lectura de la luciferasa *Firefly*.
- 9- Añadir 40 µl del reactivo Stop&Glo, pipetear 2 veces y realizar la lectura de la luciferasa *Renilla*.
- 10- Repetir los pasos 7 a 9 para todas las muestras del ensayo.

Apéndice II: Tablas

Tabla A-1. Polimorfismos potencialmente funcionales descritos en las bases de datos en el momento de su selección (dbSNP v126).

Polimorfismo	Gen	Cambio	dbSNP v.126		Het ^b	Bases de datos ^c			dbSNP v.137	
			Posición ^a	Tipo		S	E	N	Posición ^a	Tipo
rs2156310	<i>TFF1</i>	C>T	c.-2	5'UTR	0.323		X	X	c.-2	5'UTR
rs225358	<i>TFF1</i>	A>G	c.85+8	Intrónico	0.463		X	X	c.85+8	Intrónico
rs13051704	<i>TFF1</i>	C>G	c.85+140	Intrónico	0.224		X	X	c.85+140	Intrónico
rs4920094	<i>TFF1</i>	C>T	c.85+187	Intrónico	0.490		X	X	c.85+187	Intrónico
rs225357	<i>TFF1</i>	G>A	c.85+207	Intrónico	N.I.		X	X	c.85+207	Intrónico
rs2839488	<i>TFF1</i>	C>G	c.85+344	Intrónico	0.408	X	X	X	c.85+344	Intrónico
rs225356	<i>TFF1</i>	C>T	c.85+452	Intrónico	N.I..		X	X	c.85+452	Intrónico
rs1537117	<i>TFF1</i>	G>C	c.85+620	Intrónico	N.I.		X	X	c.85+620	Intrónico
rs130447632	<i>TFF1</i>	C>T	c.85+1210	Intrónico	N.I..		X	X	c.85+1210	Intrónico
rs17767167	<i>TFF1</i>	G>A	c.86-1484	Intrónico	0.028		X	X	c.86-1484	Intrónico
rs225355	<i>TFF1</i>	G>A	c.86-966	Intrónico	0.452		X	X	c.86-966	Intrónico
rs11909644	<i>TFF1</i>	A>G	c.86-954	Intrónico	N.I..		X	X	c.86-954	Intrónico
rs225354	<i>TFF1</i>	T>C	c.86-685	Intrónico	0.464			X	c.86-685	Intrónico
rs7281220	<i>TFF1</i>	G>C	c.238+18	Intrónico	0.038		X	X	c.238+18	Intrónico
rs225353	<i>TFF1</i>	C>G	c.238+309	Intrónico	0.471		X	X	c.238+309	Intrónico
rs28476620	<i>TFF1</i>	T>C	c.239-21	Intrónico	N.D.		X	X	c.239-21	Intrónico
rs13276	<i>TFF1</i>	G>A	c.*41	3'UTR	N.I.	X	X	X	c.*41	3'UTR
rs13052596	<i>TFF2</i>	G>T	c.-450	Genómico 5'	N.I.		X	X	c.-450	Genómico 5'
rs7277409	<i>TFF2</i>	C>T	c.7	R3W	N.I.		X	X	c.7	R3W
rs5844113	<i>TFF2</i>	delA	c.79+59	Intrónico	N.I..		X	X	c.79+59	Intrónico
rs11366298	<i>TFF2</i>	delA	c.79+61	Intrónico	N.I.		X	X	c.79+61	Intrónico
rs11330381	<i>TFF2</i>	delT	c.79+95	Intrónico	N.I.		X	X	c.79+95	Intrónico
rs1079380	<i>TFF2</i>	C>T	c.80-144	Intrónico	0.5		X	X	c.80-144	Intrónico
rs13047301	<i>TFF2</i>	C>A	c.230-582	Intrónico	N.I.		X	X	c.230-582	Intrónico
rs2839486	<i>TFF2</i>	G>C	c.230-456	Intrónico	0.321		X	X	c.230-456	Intrónico
rs365821	<i>TFF2</i>	G>A	c.230-142	Intrónico	0.000		X	X	c.230-142	Intrónico
rs225337	<i>TFF2</i>	A>T	c.376+95	Intrónico	N.I.		X	X	c.376+95	Intrónico
rs225336	<i>TFF2</i>	T>A	c.376+181	Intrónico	0.483		X	X	c.376+181	Intrónico
rs225335	<i>TFF2</i>	C>T	c.376+231	Intrónico	0.499		X	X	c.376+231	Intrónico
rs225334	<i>TFF2</i>	T>C	c.*82	3'UTR	0.476		X	X	c.*82	3'UTR
rs4920084	<i>TFF2</i>	C>T	c.*160	3'UTR	0.014		X	X	c.*160	3'UTR
rs225362	<i>TFF3</i>	T>C	c.-1671	Genómico 5'	0.173	X	X	X	c.-1609	Genómico 5'
rs11701143	<i>TFF3</i>	A>G	c.106	T36A	N.I.		X	X	c.-3	5'UTR
rs11546960	<i>TFF3</i>	C>T	c.191	A64V	N.I.	X	X	X	c.83	A28V
rs225361	<i>TFF3</i>	T>C	c.232+419	Intrónico	0.463		X	X	c.124+419	Intrónico
rs225360	<i>TFF3</i>	C>T	c.233-190	Intrónico	0.460		X	X	c.125-190	Intrónico
rs11546961	<i>TFF3</i>	C>T	c.284	P95L	N.I.	X		X	c.176	P59L
rs2236705	<i>TFF3</i>	G>T	c.379+52	Intrónico	0.204	X	X	X	c.272-449	Intrónico
rs11465201	<i>MUC1</i>	C>A	c.-547	Genómico 5'	0.011			X	c.-547	Genómico 5'
rs11465202	<i>MUC1</i>	C>G	c.-489	Genómico 5'	0.017			X	c.-489	Genómico 5'
rs11465203	<i>MUC1</i>	C>T	c.-323	Genómico 5'	0.026			X	c.-323	Genómico 5'
rs11465204	<i>MUC1</i>	G>A	c.-71	Genómico 5'	0.025		X	X	c.-71	5'UTR
rs4072037	<i>MUC1</i>	A>G	c.93	T31T	0.452	X	X	X	c.93	T31T
rs4971063	<i>MUC1</i>	C>G	c.186+560	Intrónico	N.I.			X	c.186+560	Intrónico
rs12742111	<i>MUC1</i>	C>A	c.187-617	Intrónico	N.I.			X	c.187-617	Intrónico
rs12728535	<i>MUC1</i>	A>C	c.187-502	Intrónico	N.I.			X	c.187-502	Intrónico
rs12743084	<i>MUC1</i>	G>C	c.187-461	Intrónico	N.I.			X	c.187-461	Intrónico
rs12737964	<i>MUC1</i>	C>A	c.187-403	Intrónico	N.I.			X	c.187-403	Intrónico

Polimorfismo	Gen	Cambio	dbSNP v.126		Het ^b	Bases de datos ^c			dbSNP v.137	
			Posición ^a	Tipo		S	E	N	Posición ^a	Tipo
rs12737963	<i>MUC1</i>	C>A	c.187-401	Intrónico	N.I.		X		c.187-401	Intrónico
rs1131772	<i>MUC1</i>	G>C	c.187-579	Intrónico	N.I.		X		c.187-579	Intrónico
rs11465205	<i>MUC1</i>	G>A	c.187-234	Intrónico	0.013		X		c.187-234	Intrónico
rs35819649	MUC1	T>A	c.196	S66T	0.025		X		c.196	S66T
rs11465206	<i>MUC1</i>	T>C	c.256-17	Intrónico	0.012		X		c.256-17	Intrónico
rs1611770	MUC1	G>A	c.376	V126M	0.024		X		c.376	V126M
rs41264913	MUC1	T>A	c.393	N131K	N.I.		X		c.393	N131K
rs35409840	<i>MUC1</i>	InsC	c.448+14	Intrónico	N.I.		X		c.448+14	Intrónico
rs11465207	MUC1	G>A	c.452	N151S	0.007		X		c.452	N151S
rs1611771	<i>MUC1</i>	G>A	c.570+36	Intrónico	0.011		X		c.570+36	Intrónico
rs11465208	<i>MUC1</i>	DelC	c.720+205	Intrónico	0.045		X		c.720+205	Intrónico
rs28414349	<i>MUC1</i>	T>G	c.721-503	Intrónico	N.I.		X		c.721-503	Intrónico
rs12566541	<i>MUC1</i>	A>C	c.721-227	Intrónico	N.I.		X		c.721-227	Intrónico
rs11465209	MUC1	C>T	c.756	Y252Y	0.011		X		c.756	Y252Y
rs1611772	MUC1	C>T	c.*6	3'UTR	0.011		X		c.*6	3'UTR
rs1611773	MUC1	C>T	c.*79	3'UTR	0.011		X		c.*79	3'UTR
rs11465210	MUC1	G>A	c.*197	3'UTR	0.011		X		c.*197	3'UTR
rs11590286	<i>MUC1</i>	G>A	c.*530	Genómico 3'	N.I.		X		c.*530	Genómico 3'
rs11465211	<i>MUC1</i>	G>A	c.*543	Genómico 3'	0.011		X		c.*543	Genómico 3'
rs1611774	<i>MUC1</i>	C>A	c.*758	Genómico 3'	0.103		X		c.*758	Genómico 3'
rs2070803	<i>MUC1</i>	A>G	c.*896	Genómico 3'	0.500		X		c.*896	Genómico 3'
rs4971101	<i>MUC1</i>	G>A	c.*976	Genómico 3'	0.500		X		c.*976	Genómico 3'
rs11465212	<i>MUC1</i>	G>A	c.*1166	Genómico 3'	0.023		X		c.*1166	Genómico 3'
rs11465213	<i>MUC1</i>	G>A	c.*1276	Genómico 3'	0.012		X		c.*1276	Genómico 3'
rs3186176	<i>MUC1</i>	C>A	c.*1428	Genómico 3'	N.I.		X		c.*1428	Genómico 3'

Tabla A-2. Principales características de la población de pacientes con lesiones gástricas (N=387).

Características	Total		Estables		Regresión		Progresión		Valor p ^a
	n	%	n	%	n	%	n	%	
Sexo									
Hombre	178	46.0	74	44.3	46	41.4	58	53.2	0.182
Mujer	209	54.0	93	55.7	65	58.6	51	46.8	
Edad									
Media ± SD									
<40	113	29.2	46	27.5	36	32.4	31	28.4	0.312
40-50	85	22.0	37	22.2	26	23.4	22	20.2	
50-60	122	31.5	53	31.7	38	34.2	31	28.4	
>60	67	17.3	31	18.6	11	9.9	25	22.9	
Consumo									
No	196	51.7	76	46.3	55	49.5	65	62.5	0.031
Si	183	48.3	88	53.7	56	50.5	39	37.5	
Historia									
No	317	84.5	145	87.9	89	82.4	83	81.4	0.277
Si	58	15.5	20	12.1	19	17.6	19	18.6	
Consumo									
Nunca	219	57.0	97	58.4	64	58.2	58	53.7	0.727
Habitual	103	26.8	41	24.7	32	29.1	30	27.8	
Ex-fumador	62	16.2	28	16.9	14	12.7	20	18.5	
Infección									
No	88	23.0	42	25.3	19	17.1	27	25.5	0.219
Si	295	77.0	124	74.7	92	82.9	79	74.5	
Tipo de lesión^b									
Normal	7	1.8	3	1.8	0	0.0	4	3.7	2.8x10 ⁻⁴
NAG	44	11.4	20	12.0	2	1.8	22	20.2	
MAG	190	49.1	84	50.3	60	54.1	46	42.2	
IM Completa	78	20.2	29	17.4	23	20.7	26	23.9	
IM Incompleta	68	17.6	31	18.6	26	23.4	11	10.1	
Total	387	100	167	100	111	100	109	100	

a, Valor p de las diferencias entre los grupos de evolución para cada variable basado en el test χ^2 ; b, Tipo de lesión diagnosticada en el reclutamiento. NAG, gastritis no atrófica; MAG, gastritis atrófica multifocal; IM, metaplasia intestinal

Tabla A-3. Resultados del genotipado de los polimorfismos seleccionados para el estudio de seguimiento de lesiones precursoras de cáncer gástrico.

Polimorfismo	Gen	Total		No cambio		Progresión		Regresión		HWE
		MAF	N	MAF	N	MAF	N	MAF	N	
rs3814316	TRIM46	0,32	383	0,31	164	0,35	108	0,31	111	0.1287
rs9426886	TRIM46	0.42	387	0.42	167	0.45	109	0.39	111	0.2117
rs4072037 ^a	MUC1	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs1045253	MUC1	0.31	372	0.29	161	0.33	104	0.33	107	0.7168
rs4076950	MUC6	0.38	380	0.40	164	0.36	107	0.37	109	0.5150
rs7481521	MUC6	0.37	354	0.36	156	0.40	101	0.35	97	0.1391
rs11604757 ^a	MUC6	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs11246384	MUC6	0.09	385	0.10	165	0.07	109	0.09	111	0.3418
rs6597947	MUC6	0.07	379	0.05	163	0.09	106	0.07	110	0.0194
rs9794921	MUC6	0.13	366	0.11	158	0.16	102	0.15	106	0.2527
rs10902073	MUC2	0.38	380	0.39	164	0.43	107	0.32	109	0.5150
rs10794281	MUC2	0.44	359	0.45	157	0.48	99	0.38	103	0.5221
rs2856082	MUC2	0.12	383	0.11	164	0.13	108	0.11	111	0.6278

Polimorfismo	Gen	Total		No cambio		Progresión		Regresión		HWE
		MAF	N	MAF	N	MAF	N	MAF	N	
rs2071174	<i>MUC2</i>	0.29	383	0.29	165	0.25	108	0.34	110	0.5365
rs7396030	<i>MUC2</i>	0.20	385	0.21	166	0.19	108	0.18	111	0.4211
rs11245936	<i>MUC2</i>	0.06	374	0.07	161	0.06	105	0.07	108	0.1846
rs7944723	<i>MUC2</i>	0.19	375	0.21	166	0.19	106	0.17	103	0.8689
rs10902089 ^a	<i>MUC2</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs6421972	<i>MUC2</i>	0.39	381	0.40	166	0.34	105	0.40	110	1
rs10794293	<i>MUC2</i>	0.33	387	0.34	167	0.27	109	0.36	111	0.9081
rs11245954	<i>MUC2</i>	0.07	382	0.08	165	0.06	108	0.05	109	1
rs7480563	<i>MUC2</i>	0.48	367	0.50	158	0.43	102	0.50	107	0.0756
rs7126405	<i>MUC2</i>	0.09	387	0.10	167	0.08	109	0.08	111	1
rs3924453	<i>MUC2</i>	0.29	387	0.31	167	0.25	109	0.30	111	0.7110
rs4077759	<i>MUC2</i>	0.37	379	0.40	162	0.31	106	0.38	111	0.5810
rs8133510	<i>TFF3</i>	0.22	386	0.22	166	0.24	109	0.21	111	0.3759
c.*41G>A	<i>TFF3</i>	0.01	426	0.01	179	0.01	116	0.00	131	0.0245
rs2236705	<i>TFF3</i>	0.18	387	0.17	167	0.16	109	0.20	111	0.8613
V80M	<i>TFF3</i>	0.01	382	0.01	164	0.02	108	0.00 ^e	110	0.0465
rs11701143	<i>TFF3</i>	0.04	413	0.02	176	0.06	108	0.03	129	1
rs225363	<i>TFF3</i>	0.41	368	0.45	157	0.37	103	0.39	108	0.0231
rs9981660 ^a	<i>TFF3</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs692356	<i>TFF3</i>	0.40	384	0.44	165	0.36	108	0.39	111	0.0709
rs4920084	<i>TFF2</i>	0.02	426	0.02	179	0.02	116	0.03	131	1
rs225334	<i>TFF2</i>	0.38	437	0.36	181	0.40	123	0.40	133	1
c.*67_68insCTT	<i>TFF2</i>	0.06	410	0.04	173	0.08	110	0.06	127	0.6272
VNTR	<i>TFF2</i>	0.17	323	0.17	130	0.17	85	0.18	108	0.3270
V66L	<i>TFF2</i>	0.02	383	0.02	164	0.01	108	0.02	111	1
rs1079380	<i>TFF2</i>	0.47	387	0.51	167	0.41	109	0.47	111	0.1526
Q7P	<i>TFF2</i>	0.01	384	0.02	166	0.01	107	0.00	111	0.0463
rs7277409 ^b	<i>TFF2</i>	1	381	1	165	1	106	1	110	NA
rs13052596	<i>TFF2</i>	0.30	444	0.30	185	0.32	126	0.30	133	1
rs3814896	<i>TFF2</i>	0.21	372	0.20	160	0.25	104	0.20	108	0.4378
c.*62C>T ^b	<i>TFF1</i>	1	379	1	162	1	106	1	111	NA
rs13276	<i>TFF1</i>	0.05	375	0.05	165	0.05	105	0.05	105	1
rs225353	<i>TFF1</i>	0.46	373	0.43	161	0.48	106	0.50	106	0.6770
rs225357	<i>TFF1</i>	0.40	354	0.43	152	0.38	98	0.38	104	0.0028
rs4920094 ^a	<i>TFF1</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs13051704	<i>TFF1</i>	0.11	385	0.11	166	0.12	108	0.09	111	1
rs225358	<i>TFF1</i>	0.32	384	0.30	164	0.34	109	0.32	111	0.4827
rs34795821	<i>TFF1</i>	0.01	377	0.01	164	0.01	104	0.01	109	1
rs2156310	<i>TFF1</i>	0.14	434	0.13	183	0.14	121	0.15	130	0.8387
c.-29C>T	<i>TFF1</i>	0.01	385	0.01	167	0.01	108	0.00 ^e	110	0.0462
rs9976977	<i>TFF1</i>	0.38	385	0.35	166	0.39	109	0.43	110	0.1965
rs424694	<i>TFF1</i>	0.40	363	0.42	158	0.39	100	0.39	105	0.1011
rs13047838	<i>TMPRSS3</i>	0.28	366	0.24	158	0.28	104	0.32	104	0.4325
rs1800872	<i>IL10</i>	0.24	451	0.26	191	0.20	132	0.24	128	0.8961
rs1800896	<i>IL10</i>	0.37	453	0.36	191	0.39	133	0.38	129	0.3170
rs2856836	<i>IL1A</i>	0.31	400	0.31	168	0.30	125	0.29	107	0.5591
rs3783546	<i>IL1A</i>	0.24	450	0.23	190	0.25	131	0.26	129	0.7025
rs17561	<i>IL1A</i>	0.31	451	0.31	189	0.31	133	0.31	129	0.7435
rs1800587	<i>IL1A</i>	0.31	451	0.30	190	0.31	132	0.30	129	0.9118
rs1143634	<i>IL1B</i>	0.25	451	0.26	190	0.26	133	0.23	128	0.9012
rs1143627	<i>IL1B</i>	0.34	452	0.33	191	0.34	133	0.36	128	0.6026
rs16944	<i>IL1B</i>	0.33	453	0.32	191	0.32	133	0.36	129	1

Polimorfismo	Gen	Total		No cambio		Progresión		Regresión		HWE
		MAF	N	MAF	N	MAF	N	MAF	N	
rs1143623	<i>IL1B</i>	0.23	450	0.21	190	0.24	133	0.25	127	0.5076
rs4251961	<i>IL1RN</i>	0.37	431	0.36	183	0.41	129	0.35	119	0.1832
rs2637988	<i>IL1RN</i>	0.39	452	0.38	191	0.36	133	0.43	128	0.2804
rs4251985	<i>IL1RN</i>	0.29	450	0.28	189	0.27	132	0.31	129	0.4213
rs4073	<i>IL8</i>	0.47	428	0.47	179	0.43	130	0.49	119	0.6991
rs2243250	<i>IL4</i>	0.15	453	0.15	191	0.15	133	0.16	129	0.7176
rs2070874	<i>IL4</i>	0.15	453	0.15	191	0.14	133	0.17	129	0.8539
rs2227284	<i>IL4</i>	0.26	453	0.25	191	0.27	133	0.27	129	0.6282
rs3181226	<i>IL12B</i>	0.20	447	0.19	188	0.21	133	0.20	126	0.8832
rs3181224	<i>IL12B</i>	0.19	450	0.19	188	0.18	133	0.19	129	0.6455
rs1368439	<i>IL12B</i>	0.13	452	0.14	190	0.11	133	0.15	129	0.2167
rs3212227	<i>IL12B</i>	0.21	449	0.18	189	0.24	133	0.21	127	0.3187
rs1800797	<i>IL6</i>	0.36	439	0.36	187	0.40	131	0.33	121	0.3035
rs2057768	<i>IL4R</i>	0.27	451	0.27	189	0.29	133	0.26	128	0.3426
rs1805010	<i>IL4R</i>	0.43	445	0.43	189	0.43	131	0.43	125	0.9232
rs1805011	<i>IL4R</i>	0.15	453	0.15	191	0.15	133	0.14	129	0.7060
rs1805013	<i>IL4R</i>	0.03	453	0.04	191	0.03	133	0.03	129	0.3478
rs1801275	<i>IL4R</i>	0.21	453	0.20	191	0.21	133	0.22	129	0.7775
rs28362491	<i>NFKB1</i>	0.35	361	0.29	158	0.38	99	0.40	104	0.1055
rs5743708	<i>TLR2</i>	0.02	382	0.01	163	0.00	111	0.00	108	1
rs2569190	<i>CD14</i>	0.48	373	0.52	159	0.49	109	0.41	105	0.5342
rs5744455	<i>CD14</i>	0.19	377	0.23	162	0.14	108	0.18	107	0.0633
rs909253	<i>LTA</i>	0.26	451	0.27	191	0.26	132	0.25	128	1
rs1800629	<i>TNF</i>	0.13	451	0.13	191	0.11	133	0.16	127	0.5415
rs361525	<i>TNF</i>	0.08	450	0.08	191	0.08	133	0.08	126	0.0598
rs2234711	<i>IFNGR1</i>	0.46	384	0.47	167	0.42	111	0.48	106	0.4107
rs6472812	<i>LY96</i>	0.02	387	0.01	167	0.02	111	0.03	109	1
rs11466004	<i>LY96</i>	0.01	382	0.01	164	0.00	111	0.01	107	1
rs2976392	<i>PSCA</i>	0.41	378	0.39	164	0.42	109	0.41	105	0.5940
rs4986790	<i>TLR4</i>	0.06	374	0.07	160	0.03	108	0.08	106	1
rs4986791	<i>TLR4</i>	0.07	387	0.09	167	0.04	111	0.08	109	0.7086
rs2430561	<i>IFNG</i>	0.48	376	0.48	161	0.50	108	0.47	107	0.2556
rs2301756	<i>PTPN11</i>	0.10	387	0.12	167	0.10	111	0.07	109	0.2553
rs2066842	<i>NOD2</i>	0.27	384	0.26	166	0.28	110	0.27	108	0.7949
rs2066844 ^a	<i>NOD2</i>	0.04	358	0.04	157	0.04	97	0.03	104	1
rs16260 ^a	<i>CDH1</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs1125557	<i>CDH1</i>	0.40	375	0.35	165	0.45	107	0.42	103	0.7469
rs2297518	<i>NOS2A</i>	0.18	380	0.19	164	0.18	109	0.16	107	0.4784
rs25487	<i>XRCC1</i>	0.37	445	0.38	190	0.35	133	0.39	122	1
rs1799782	<i>XRCC1</i>	0.05	424	0.05	176	0.03	131	0.06	117	1
rs13181 ^c	<i>ERCC2</i>	0.36	412	0.36	176	0.33	127	0.39	109	0.3931
rs1799793	<i>ERCC2</i>	0.32	436	0.30	181	0.32	130	0.36	125	0.3834

a, Polimorfismos descartados debido a problemas de genotipado en la plataforma sequenom; b, Polimorfismos descartados por ser monomórficos; c, Polimorfismos descartados por no pasar el el control de calidad. e, MAF <0.01. HWE, *Hardy Weinberg Equilibrium*

Polimorfismo	Gen	Progresión ^a					Regresión ^a				
		Cod	Dom	Rec	Sob	Adi	Cod	Dom	Rec	Sob	Adi
rs9546481	IL1A	0.841	0.713	0.749	0.573	0.892	0.243	0.527	0.204	0.156	0.920
rs3783546	IL1A	0.193	0.469	0.156	0.169	0.937	0.392	0.987	0.189	0.533	0.600
rs4251985	IL1RN	0.637	0.974	0.365	0.607	0.667	0.923	0.763	0.730	0.909	0.702
rs2637988	IL1RN	0.456	0.483	0.223	0.822	0.261	0.356	0.664	0.151	0.564	0.304
rs4251961	IL1RN	0.408	0.205	0.430	0.469	0.186	0.186	0.404	0.070	0.688	0.128
rs1143627	IL1B	0.957	0.947	0.800	0.818	0.942	0.898	0.652	0.802	0.770	0.651
rs1143623	IL1B	0.924	0.706	0.978	0.692	0.759	0.259	0.733	0.101	0.714	0.380
rs16944	IL1B	0.942	0.744	0.833	0.849	0.733	0.988	0.877	0.982	0.887	0.900
rs1143634	IL1B	0.851	0.603	0.707	0.732	0.571	0.971	0.975	0.813	0.929	0.904
rs2070874	IL4	0.912	0.672	0.976	0.672	0.699	0.426	0.956	0.198	0.750	0.703
rs2227284	IL4	0.978	0.926	0.881	0.858	0.992	0.875	0.666	0.890	0.606	0.773
rs2243250	IL4	0.999	0.970	0.976	0.977	0.966	0.348	0.691	0.198	0.439	0.975
rs1805011	IL4R	0.908	0.661	0.920	0.676	0.673	0.910	0.794	0.781	0.720	0.877
rs1801275	IL4R	0.506	0.997	0.267	0.591	0.675	0.637	0.779	0.430	0.519	0.957
rs1805013	IL4R	0.720	0.744	0.434	0.837	0.669	0.699	0.815	0.401	0.922	0.724
rs2057768	IL4R	0.823	0.584	0.925	0.534	0.705	0.612	0.351	0.999	0.336	0.478
rs1805010	IL4R	0.229	0.390	0.256	0.094	0.953	0.126	0.315	0.200	0.046	0.950
rs1800797	IL6	0.337	0.360	0.162	0.957	0.179	0.290	0.426	0.122	0.753	0.177
rs4073	IL8	0.628	0.337	0.694	0.586	0.393	0.352	0.277	0.206	0.969	0.153
rs1800896	IL10	0.567	0.345	0.428	0.718	0.287	0.861	0.674	0.855	0.585	0.843
rs1800872	IL10	0.872	0.884	0.656	0.727	0.967	0.144	0.054	0.874	0.057	0.099
rs3181224	IL12B	0.243	0.768	0.120	0.433	0.873	0.956	0.960	0.782	0.884	0.970
rs3212227	IL12B	0.481	0.427	0.535	0.261	0.681	0.324	0.169	0.312	0.352	0.133
rs1368439	IL12B	0.536	0.264	0.727	0.305	0.287	0.215	0.081	0.546	0.110	0.086
rs3181226	IL12B	0.653	0.893	0.360	0.808	0.668	0.770	0.600	0.541	0.761	0.513
rs361525	TNF	0.944	-	-	-	-	0.944	-	-	-	-
rs1800629	TNF	0.008	0.587	0.002	0.725	0.191	0.087	0.280	0.038	0.514	0.159
rs909253	LTA	0.598	0.332	0.963	0.327	0.438	0.901	0.657	0.975	0.660	0.711
rs1799782	XRCC	0.233	0.215	0.180	0.302	0.163	0.317	0.169	0.430	0.198	0.152
rs25487	XRCC	0.682	0.531	0.436	0.955	0.398	0.624	0.388	0.486	0.714	0.332
rs13181	ERCC	0.834	0.598	0.659	0.823	0.547	0.716	0.432	0.651	0.622	0.422
rs1799793	ERCC	0.609	0.358	0.507	0.624	0.321	0.994	0.918	0.995	0.920	0.937
rs5743708	TLR2	0.179	-	-	-	-	0.151	-	-	-	-
rs2569190	CD14	0.048	0.082	0.025	0.813	0.017	0.566	0.907	0.341	0.363	0.626
rs5744455	CD14	0.651	0.544	0.391	0.825	0.417	0.173	0.102	0.162	0.290	0.065
rs6472812	LY96	0.759	-	-	-	-	0.908	-	-	-	-
rs11466004	LY96	0.551	-	-	-	-	0.354	-	-	-	-
rs2976392	PSCA	0.790	0.895	0.497	0.704	0.652	0.378	0.896	0.179	0.364	0.416
rs4986790	TLR4	0.309	0.246	0.398	0.198	0.315	0.069	0.024	0.437	0.030	0.022
rs4986791	TLR4	0.585	0.667	0.375	0.585	0.758	0.084	0.031	0.466	0.036	0.028
rs2430561	IFNG	0.633	0.655	0.524	0.345	0.922	0.932	0.822	0.717	0.909	0.723
rs2301756	PTPN1	0.058	0.216	0.024	0.517	0.100	0.740	0.941	0.466	0.761	0.896
rs2066842	NOD2	0.673	0.862	0.376	0.789	0.624	0.775	0.516	0.634	0.675	0.475
rs2066844	NOD2	0.527	-	-	-	-	0.946	-	-	-	-
rs1125557	CDH1	0.464	0.223	0.858	0.302	0.348	0.111	0.221	0.045	0.752	0.054
rs2297518	NOS2	0.603	0.315	0.764	0.360	0.341	0.851	0.576	0.971	0.575	0.618
rs2234711	IFNGR	0.284	0.210	0.630	0.124	0.613	0.504	0.267	0.456	0.676	0.255
rs28362491	NFKB1	0.291	0.156	0.248	0.562	0.116	0.371	0.301	0.202	0.911	0.173

a, Valores p obtenidos mediante análisis de regresión logística no condicional ajustado por sexo, edad y consumo de AINES y tipo de lesión en el reclutamiento. En negrita y sombreado, los valores p significativos (p<0.05). Cod, codominante; Dom,

Tabla A-5. Asociación de los SNPs de *MUC2* rs10794293, rs3924453, rs4077759, rs10794281 y rs2071174 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por el consumo de AINES.

SNP	Modelo	Genotipo	Progresión			Consumo AINES			No consumo AINES			P-Inter.			
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p				
rs10794293	Cod	CC	117/57	1.00		55/18	1.00		64/39	1.00					
		CT	125/40	0.66	(0.40 - 1.10)	70/16	0.65	(0.30 - 1.43)	55/24	0.63	(0.32 - 1.26)				
		TT	33/7	0.43	(0.17 - 1.08)	.0905	21/5	0.69	(0.21 - 2.22)	.5465	12/2	0.17 (0.03 - 0.88)	.0408	.4073	
	Dom	CC	117/57	1.00		53/18	1.00		64/39	1.00					
		CT-TT	158/47	0.61 (0.38 - 1.00)	.0476	91/21	0.66	(0.32 - 1.38)	.2731	67/26	0.54	(0.28 - 1.04)	.0632	.7980	
		CC-CT	242/97	1.00		123/34	1.00		119/63	1.00					
	Adi	TT	33/7	0.52	(0.21 - 1.28)	.1382	21/5	0.87	(0.29 - 2.60)	.7949	12/2	0.21 (0.04 - 1.04)	.0305	.1683	
		CC,CT,TT	275/104	0.66 (0.45 - 0.96)	.0284	144/39	0.77	(0.44 - 1.34)	.3507	131/65	0.52 (0.30 - 0.91)	.0180	.4247		
rs3924453	Cod	GG	135/60	1.00		69/25	1.00		66/35	1.00					
		GA	112/40	0.77	(0.47 - 1.27)	63/12	0.50	(0.23 - 1.10)	49/28	1.03	(0.52 - 2.03)				
		AA	28/4	0.30	(0.10 - 0.93)	.0597	12/2	0.42	(0.08 - 2.10)	.1567	16/2	0.24	(0.05 - 1.17)	.1149	.2395
	Dom	GG	135/60	1.00		69/25	1.00		66/35	1.00					
		GA-AA	140/44	0.67	(0.42 - 1.00)	.1074	75/14	0.49	(0.23 - 1.03)	.0557	65/30	0.83 (0.43 - 1.59)	.0377	.2220	
		GG-GA	247/100	1.00		132/37	1.00		115/63	1.00					
	Adi	AA	28/4	0.34 (0.11 - 1.01)	.0320	12/2	0.56	(0.11 - 2.73)	.4479	16/2	0.24	(0.05 - 1.13)	.5697	.4675	
		GG,GA,AA	275/104	0.66 (0.44 - 0.97)	.0317	81/27	0.56	(0.30 - 1.05)	.0624	131/65	0.72	(0.43 - 1.20)	.1962	.4647	
rs4077759	Cod	TT	105/47	1.00		57/20	1.00		48/27	1.00					
		TC	119/48	0.86	(0.52 - 1.44)	60/14	0.61	(0.27 - 1.35)	59/34	1.07	(0.53 - 2.16)				
		CC	46/7	0.35 (0.14 - 0.85)	.0437	25/4	0.42	(0.13 - 1.41)	.2454	21/3	0.28	(0.07 - 1.09)	.0830	.3943	
	Dom	TT	105/47	1.00		57/20	1.00		48/27	1.00					
		TC-CC	165/55	0.72	(0.44 - 1.18)	.1897	85/18	0.55	(0.26 - 1.16)	.1156	80/37	0.86	(0.44 - 1.69)	.6628	.3532
		TT-TC	224/95	1.00		117/34	1.00		107/61	1.00					
	Adi	CC	46/7	0.37 (0.16 - 0.88)	.0148	25/4	0.53	(0.17 - 1.68)	.4334	21/3	0.27 (0.08 - 0.99)	.0263	.4372		
		TT,TC,CC	270/102	0.68 (0.47 - 0.98)	.0338	142/38	0.64	(0.37 - 1.10)	.0953	128/64	0.71	(0.42 - 1.17)	.1707	.7441	
rs10794281	Cod	TT	89/24	1.00		36/10	1.00		53/14	1.00					
		TC	119/47	1.64	(0.90 - 2.96)	68/15	0.91	(0.36 - 2.28)	51/32	2.36	(1.07 - 5.19)				
		CC	49/23	1.79	(0.88 - 3.64)	.1685	26/6	0.92	(0.29 - 2.91)	.9791	23/17	2.61	(1.03 - 6.58)	.0509	.3074
	Dom	TT	89/24	1.00		36/10	1.00		53/14	1.00					
		TC-CC	168/70	1.68 (0.96 - 2.94)	.0621	94/21	0.98	(0.39 - 2.17)	.8380	74/49	2.44 (1.16 - 5.11)	.0151	.1241		
		TT-TC	298/71	1.00		104/25	1.00	(0.36 - 2.67)	.9613	104/46	1.00				
	Adi	CC	49/23	1.33	(0.73 - 2.42)	.3545	26/6	0.94		23/17	1.55	(0.71 - 3.35)	.2715	.5193	
		TT,TC,CC	257/94	1.35	(0.96 - 1.91)	.0854	130/31	0.95	(0.54 - 1.69)	.8714	127/63	1.64 (1.04 - 2.58)	.0303	.1856	
rs2071174	Cod	TT	132/59	1.00		72/51	1.00		60/39	1.00					
		TC	112/38	0.72	(0.43 - 1.19)	54/16	1.01	(0.47 - 2.18)	58/22	0.54	(0.27 - 1.07)				
		CC	28/6	0.48	(0.18 - 1.28)	.1973	16/3	0.72	(0.19 - 2.80)	.8788	12/3	0.29	(0.07 - 1.21)	.0698	.4303
	Dom	TT	132/59	1.00		72/20	1.00		60/39	1.00					
		TC-CC	140/44	0.67	(0.42 - 1.09)	.1064	70/19	0.95	(0.46 - 1.96)	.8930	70/25	0.49 (0.25 - 0.95)	.0320	.2098	
		TT-TC	244/97	1.00		126/36	1.00		118/61	1.00					
	Adi	CC	28/6	0.56	(0.21 - 1.44)	.3456	16/3	0.72	(0.19 - 2.67)	.6121	12/3	0.38	(0.10 - 1.54)	.1581	.5502
		TT,TC,CC	272/103	0.71	(0.48 - 1.04)	.0721	142/39	0.91	(0.53 - 1.59)	.7451	130/64	0.54 (0.31 - 0.93)	.0210	.2092	

a, Análisis de regresión logística no condicional ajustado por sexo, edad y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla A-6. Asociación de los SNPs de MUC2 rs10794293 y rs4077759 con la progresión de lesiones precursoras de cáncer gástrico y estratificación por el por sexo.

SNP	Modelo	Genotipo	Progresión			Hombres			Mujeres			P-Inter.		
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p			
rs10794293	Cod	CC	117/57	1.00		52/31	1.00		65/26	1.00				
		CT	125/40	0.66	(0.40-1.10)	49/21	0.76	(0.36 - 1.59)	76/19	0.59	(0.29 - 1.20)			
		TT	33/7	0.43	(0.17-1.08)	.0905	17/2	0.17	(0.04 - 0.86)	.0476	16/5	0.85	(0.26 - 2.80)	.3387
	Dom	CC	117/57	1.00		52/31	1.00		65/26	1.00				
		CT-TT	158/47	0.61	(0.38-1.00)	.0476	66/23	0.59	(0.29 - 1.21)	.1474	92/24	0.63	(0.32 - 1.24)	.1777
	Rec	CC-CT	242/97	1.00		101/52	1.00		141/45	1.00				
		TT	33/7	0.52	(0.21-1.28)	.1382	17/2	0.20	(0.04 - 0.95)	.0185	16/5	1.12	(0.36 - 3.50)	.8418
Adi	CC,CT,TT	275/104	0.66	(0.45-0.96)	.0284	118/54	0.55	(0.32 - 0.97)	.0333	157/50	0.77	(0.45 - 1.32)	.3344	.5359
rs4077759	Cod	TT	105/47	1.00		52/25	1.00		53/22	1.00				
		TC	119/48	0.86	(0.52-1.44)	49/26	0.91	(0.43 - 1.90)	70/22	0.74	(0.35 - 1.56)			
		CC	46/7	0.35	(0.14-0.85)	.0437	14/3	0.40	(0.10 - 1.65)	.4010	32/4	0.28	(0.09 - 0.92)	.0744
	Dom	TT	105/47	1.00		52/25	1.00		53/22	1.00				
		TC-CC	165/55	0.72	(0.44-1.18)	.1897	63/29	0.80	(0.39 - 1.63)	.5392	102/26	0.58	(0.29 - 1.17)	.1315
	Rec	TT-TC	224/95	1.00		101/51	1.00		123/44	1.00				
		CC	46/7	0.37	(0.16-0.88)	.0148	14/3	0.42	(0.11 - 1.64)	.1844	32/4	0.33	(0.11 - 1.01)	.0322
Adi	TT,TC,CC	270/102	0.68	(0.47-0.98)	.0338	115/54	1.06	(0.42 - 1.29)	.2851	155/22	0.59	(0.36 - 0.97)	.0343	.3763

a, Análisis de regresión logística no condicional ajustado por edad , consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla A-7. Asociación de los SNPs de *MUC2* rs10794281, rs2071174 y 7944723 con la regresión de lesiones precursoras de cáncer gástrico y estratificación por la infección por *H.pylori*.

SNP	Modelo	Genotipo	Regresión			Hombres			Mujeres			P-Inter.			
			Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p	Controles/ Casos	OR (CI 95%) ^a	Valor p				
rs10794281	Cod	TT	72/41	1.00		34/17	1.00		38/24	1.00					
		TC	121/45	0.60	(0.35 - 1.04)	54/21	0.62	(0.27 - 1.42)	67/24	0.57	(0.27 - 1.18)				
		CC	55/17	0.53	(0.27 - 1.06)	.1037	31/6	0.32	(0.10 - 0.98)	.1139	24/11	0.73	(0.28 - 1.88)	.3184	.4387
	Dom	TT	72/41	1.00		34/17	1.00		38/24	1.00					
		TC-CC	176/62	0.58	(0.35 - 0.96)	.0361	85/27	0.51	(0.10 - 0.98)	.0939	91/35	0.61	(0.31 - 1.21)	.1586	.8916
	Rec	TT-TC	193/86	1.00		88/38	1.00		105/48	1.00					
		CC	55/17	0.71	(0.38 - 1.33)	.2800	31/6	0.42	(0.15 - 1.16)	.0806	24/11	1.02	(0.44 - 2.37)	.9576	.2124
	Adi	TT,TC,CC	248/103	0.71	(0.50 - 1.00)	.0463	119/44	0.57	(0.34 - 0.98)	.8671	129/59	0.80	(0.50 - 1.29)	.3552	.4544
	rs2071174	Cod	TT	138/53	1.00		67/21	1.00		71/32	1.00				
TC			110/40	0.93	(0.56 - 1.54)	54/17	1.09	(0.50 - 2.39)	56/23	0.81	(0.41 - 1.61)				
CC			17/17	2.76	(1.24 - 6.13)	.0275	4/8	7.71	(1.89 - 31.4)	.0097	13/9	1.68	(0.61 - 4.66)	.4039	.1663
Dom		TT	138/53	1.00		67/21	1.00		71/32	1.00					
		TC-CC	127/57	1.17	(0.73 - 1.86)	.5186	58/25	1.54	(0.74 - 3.19)	.2412	69/32	0.96	(0.51 - 1.81)	.9105	.4807
Rec		TT-TC	248/93	1.00		121/38	1.00		127/55	1.00					
		CC	17/17	2.85	(1.32 - 6.14)	.0077	4/8	7.41	(1.90 - 28.9)	.0024	13/9	1.84	(0.69 - 4.88)	.2269	.0602
Adi		TT,TC,CC	265/110	1.36	(0.95 - 1.93)	.0893	125/46	1.91	(1.08 - 3.38)	.0247	140/64	1.12	(0.71 - 1.78)	.6316	.1902
rs7944723		Cod	CC	169/69	1.00		78/30	1.00		91/39	1.00				
	CG		83/33	0.91	(0.54 - 1.51)	37/13	0.97	(0.43 - 2.20)	46/20	0.90	(0.46 - 1.77)				
	GG		12/1	0.15	(0.02 - 1.22)	.0803	8/0	Na	Na	.0222	4/1	0.43	(0.05 - 4.14)	.7255	.3613
	Dom	CC	169/69	1.00		78/30	1.00		91/39	1.00					
		CG-GG	95/34	0.79	(0.48 - 1.31)	.3667	45/13	0.72	(0.32 - 1.58)	.4026	50/21	0.86	(0.44 - 1.67)	.6521	.7342
	Rec	CC-CG	252/102	1.00		115/43	1.00		137/59	1.00					
		GG	12/1	0.16	(0.02 - 1.25)	.0268	8/0	Na	Na	.0058	4/1	0.45	(0.05 - 4.25)	.4585	.1534
	Adi	CC,CG,GG	264/103	0.72	(0.47 - 1.12)	.1416	123/43	0.60	(0.31 - 1.16)	.1156	141/60	0.83	(0.46 - 1.50)	.5345	.5058

a, Análisis de regresión logística no condicional ajustado por sexo, edad, consumo de AINES y tipo de lesión en el reclutamiento. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla A-8. Descriptiva de la edad, consumo de alcohol y dieta de la población caso-control de cancer gástrico.

Características	Total			Controles			Casos			Valor-p
	N	Media (SD)	Mediana (P25-P75)	N	Media (SD)	Mediana (P25-P75)	N	Media (SD)	Mediana (P25-P75)	
Edad	1649	58.4 (7.74)	59.6 (53.7-63.7)	1284	58.4 (7.69)	59.6 (53.6-63.7)	365	58.4 (7.93)	59.5 (53.7-63.8)	
Dieta										
Energía (kcal/d)	1636	2138 (682.68)	2067 (1657-2524)	1272	2137 (696.87)	2066 (1556-2519)	364	2142 (631.46)	2074 (1672-2524)	0.8542
Vegetales (g/d)	1636	185.1 (134.96)	145.5 (94.3-235.4)	1272	185.7 (133.40)	145.6 (94.6-237.7)	364	183.2 (140.46)	145.4 (91.2-232.0)	0.6586
Frutas, frutos secos y seeds (g/d)	1636	231.2 (183.65)	186.7 (106.1-307.9)	1272	235.6 (182.66)	194.4 (110.5-310.6)	364	215.7 (182.00)	165.9 (91.0-302.5)	0.0412
Carne roja (g/d)	1636	49.1 (37.32)	41.2 (21.1-69.7)	1272	48.6 (37.79)	40.8 (20.6-68.7)	364	50.8 (35.62)	43.4 (24.0-71.3)	0.2769
Carne procesada (g/d)	1636	36.3 (34.55)	27.5 (12.6-50.5)	1272	35.5 (33.31)	26.6 (12.1-49.2)	364	39.4 (38.46)	30.1 (14.1-53.1)	0.0096
Consumo alcohol	1636	15.4 (22.90)	7.1 (1.0-22.2)	1272	15.0 (22.75)	6.9 (1.0-21.4)	364	16.9 (23.8)	7.6 (1.0-24.3)	0.1671

Tabla A-9. Resultados del genotipado de los polimorfismos en los genes *MUC* y *TFF* en la población caso-control de cáncer gástrico.

Polimorfismo	Gen	Total		Controles		Casos		HWE	Promedio Genotipado
		MAF	N	MAF	N	MAF	N		
rs3814316	<i>TRIM46</i>	0.35	1645	0.33	1281	0.38	364	0.9002	99.8%
rs9426886	<i>TRIM46</i>	0.41	1634	0.41	1270	0.41	364	0.2947	99.1%
rs4072037	<i>MUC1</i>	0.47	1634	0.48	1273	0.46	361	0.1774	99.1%
rs1045253	<i>MUC1</i>	0.29	1638	0.30	1278	0.29	360	0.0655	99.3%
rs4076950 ^a	<i>MUC6</i>	NA	NA	NA	NA	NA	NA	NA	NA
rs7481521	<i>MUC6</i>	0.44	1573	0.43	1230	0.47	343	0.2222	95.4%
rs11604757	<i>MUC6</i>	0.13	1508	0.13	1181	0.13	327	0.8010	91.4%
rs11246384	<i>MUC6</i>	0.08	1559	0.08	1219	0.10	340	0.6931	94.5%
rs6597947	<i>MUC6</i>	0.11	1631	0.10	1273	0.11	358	0.1740	98.9%
rs9794921	<i>MUC6</i>	0.22	1599	0.21	1243	0.23	356	0.7998	97.0%
rs10902073	<i>MUC2</i>	0.36	1642	0.36	1278	0.36	364	0.5838	99.6%
rs10794281	<i>MUC2</i>	0.43	1632	0.42	1269	0.43	363	0.455	99.0%
rs2856082	<i>MUC2</i>	0.12	1524	0.12	1190	0.13	334	0.1010	92.4%
rs2071174	<i>MUC2</i>	0.34	1572	0.33	1231	0.34	341	0.0841	95.3%
rs11245936	<i>MUC2</i>	0.08	1625	0.08	1267	0.08	358	0.2741	98.5%
rs7944723	<i>MUC2</i>	0.78	1603	0.77	1246	0.80	357	0.1952	97.2%
rs10902089	<i>MUC2</i>	0.47	1570	0.47	1228	0.48	342	0.1361	95.2%
rs6421972	<i>MUC2</i>	0.45	1556	0.45	1216	0.44	340	0.4517	94.4%
rs10794293 ^a	<i>MUC2</i>	NA	NA	NA	NA	NA	NA	NA	NA
rs11245954	<i>MUC2</i>	0.07	1577	0.08	1232	0.07	345	0.0613	95.6%
rs7480563	<i>MUC2</i>	0.49	1628	0.59	1269	0.50	359	0.3398	98.7%
rs7126405	<i>MUC2</i>	0.10	1498	0.11	1169	0.09	329	0.0686	90.8%
rs3924453	<i>MUC2</i>	0.28	1638	0.28	1276	0.28	362	0.9443	99.3%
rs4077759 ^a	<i>MUC2</i>	NA	NA	NA	NA	NA	NA	NA	NA
rs8133510	<i>TFF3</i>	0.19	1593	0.19	1243	0.17	350	1	96.6%
c.*41G>A	<i>TFF3</i>	0.004	1006	0.003	1280	0.004	274	1	77.6%
rs2236705	<i>TFF3</i>	0.16	1644	0.16	1279	0.17	365	0.1410	99.7%
V80M	<i>TFF3</i>	0.007	1648	0.008	1283	0.004	365	0.0946	99.9%
rs11701143 ^a	<i>TFF3</i>	NA	NA	NA	NA	NA	NA	NA	NA
rs225363	<i>TFF3</i>	0.42	1636	0.42	1275	0.43	361	0.6880	99.2%
rs9981660 ^a	<i>TFF3</i>	NA	NA	NA	NA	NA	NA	NA	NA
rs692356	<i>TFF3</i>	0.41	1446	0.41	1116	0.42	330	0.0002	87.7%
rs11203197	<i>TFF3</i>	0.03	1615	0.03	1260	0.03	355	0.0725	97.9%
rs12483635	<i>TFF3</i>	0.05	1561	0.05	1216	0.04	345	0.1787	94.7%
rs225334	<i>TFF2</i>	0.37	1586	0.37	1241	0.36	345	0.0072	96.2%
c.*67_68insCTT	<i>TFF2</i>	0.04	1586	0.04	1241	0.04	345	0.1661	96.2%
V66L	<i>TFF2</i>	0.001	1649	0.001	1284	0.001	365	1	100%
rs1079380	<i>TFF2</i>	0.49	1562	0.50	1218	0.50	344	0.3904	94.7%
rs7277409 ^b	<i>TFF2</i>	1	1649	1	1284	1	365	1	100%
rs3814896	<i>TFF2</i>	0.21	1642	0.21	1279	0.22	363	0.0358	99.6%
rs13276	<i>TFF1</i>	0.06	1648	0.05	1283	0.06	365	0.0942	99.9%
rs225353	<i>TFF1</i>	0.43	1636	0.43	1274	0.41	362	0.1295	99.2%
rs225356	<i>TFF1</i>	0.46	1446	0.47	1124	0.46	322	0.1860	87.7%
rs2839488	<i>TFF1</i>	0.39	1623	0.39	1264	0.40	359	1	98.4%
rs13051704	<i>TFF1</i>	0.10	1645	0.10	1281	0.09	364	0.1988	99.8%
rs225358 ^a	<i>TFF1</i>	NA	NA	NA	NA	NA	NA	NA	NA
c.-29C>T ^a	<i>TFF1</i>	NA	NA	NA	NA	NA	NA	NA	NA
rs9976977	<i>TFF1</i>	0.34	1646	0.34	1281	0.36	365	1	99.8%
rs424694	<i>TFF1</i>	0.55	1648	0.55	1283	0.54	365	0.9551	99.9%
rs13047838	<i>TMPRSS3</i>	0.21	1649	0.21	1284	0.21	365	0.6183	100%

a, Polimorfismos descartados debido a problemas de genotipado en la plataforma Illumina; b, Polimorfismos descartados por tener $MAF < 0.0005$. HWE, *Hardy Weinberg Equilibrium*

Tabla A-10. Valores p de significación de la asociación entre polimorfismos en genes *MUC* y *TFF* con cáncer gástrico y su subtipo histológico bajo los diferentes modelos de herencia.

Polimorfismo	Gen	Cáncer gástrico					Difuso					Intestinal				
		Cod	Dom	Rec	Sob	Adi	Cod	Dom	Rec	Sob	Adi	Cod	Dom	Rec	Sob	Adi
rs3814316	<i>MUC1</i>	0.025	0.089	0.103	0.945	0.012	0.504	0.619	0.244	0.777	0.348	0.026	0.058	0.015	0.841	0.001
rs9426886	<i>MUC1</i>	0.103	0.383	0.034	0.427	0.077	0.855	0.622	0.915	0.588	0.776	0.050	0.542	0.015	0.195	0.075
rs4072037	<i>MUC1</i>	0.014	0.021	0.373	0.005	0.338	0.772	0.472	0.833	0.641	0.557	0.042	0.062	0.351	0.013	0.514
rs1045253	<i>MUC1</i>	0.890	0.632	0.833	0.717	0.648	0.119	0.039	0.500	0.091	0.064	0.827	0.799	0.662	0.595	0.996
rs7481521	<i>MUC6</i>	0.220	0.154	0.154	0.867	0.082	0.263	0.238	0.141	0.921	0.108	0.504	0.249	0.558	0.547	0.279
rs11604757	<i>MUC6</i>	0.735	0.805	0.435	0.984	0.668	0.593	0.758	0.374	0.570	0.957	0.613	0.486	0.402	0.643	0.397
rs11246384	<i>MUC6</i>	0.081	0.261	0.033	0.566	0.119	0.825	0.642	0.628	0.714	0.590	0.800	0.870	0.544	0.764	0.975
rs6597947	<i>MUC6</i>	0.806	0.541	0.698	0.612	0.513	0.590	0.376	0.457	0.510	0.318	0.841	0.864	0.558	0.992	0.763
rs9794921	<i>MUC6</i>	0.765	0.474	0.730	0.563	0.469	0.184	0.072	0.379	0.152	0.070	0.736	0.885	0.481	0.656	0.903
rs10902073	<i>MUC2</i>	0.107	0.281	0.160	0.049	0.914	0.898	0.650	0.812	0.777	0.657	0.048	0.261	0.076	0.027	0.990
rs10794281	<i>MUC2</i>	0.247	0.144	0.768	0.111	0.417	0.427	0.210	0.447	0.574	0.212	0.314	0.248	0.569	0.130	0.650
rs2856082	<i>MUC2</i>	0.513	0.439	0.521	0.315	0.613	0.564	0.331	0.850	0.287	0.422	0.780	0.481	0.832	0.512	0.493
rs2071174	<i>MUC2</i>	0.040	0.103	0.181	0.014	0.567	0.550	0.506	0.539	0.292	0.839	0.342	0.143	0.665	0.245	0.207
rs11245936	<i>MUC2</i>	0.907	0.664	0.857	0.692	0.658	0.227	0.306	0.111	0.574	0.172	0.056	0.038	0.125	0.072	0.026
rs7944723	<i>MUC2</i>	0.438	0.224	0.446	0.372	0.199	0.151	0.160	0.437	0.058	0.421	0.219	0.481	0.084	0.979	0.235
rs10902089	<i>MUC2</i>	0.343	0.237	0.677	0.158	0.615	0.659	0.366	0.849	0.516	0.498	0.771	0.474	0.737	0.716	0.518
rs6421972	<i>MUC2</i>	0.856	0.714	0.603	0.934	0.593	0.451	0.252	0.377	0.713	0.210	0.990	0.996	0.895	0.918	0.937
rs11245954	<i>MUC2</i>	0.375	0.646	0.165	0.893	0.468	0.977	0.832	0.982	0.832	0.841	0.289	0.437	0.138	0.623	0.321
rs7480563	<i>MUC2</i>	0.683	0.382	0.785	0.598	0.484	0.978	0.972	0.853	0.848	0.927	0.114	0.157	0.054	0.621	0.040
rs7126405	<i>MUC2</i>	0.216	0.113	0.266	0.197	0.086	0.461	0.420	0.507	0.275	0.617	0.081	0.121	0.060	0.252	0.071
rs3924453	<i>MUC2</i>	0.952	0.821	0.888	0.759	0.905	0.560	0.931	0.325	0.508	0.715	0.697	0.879	0.398	0.781	0.644
rs8133510	<i>TFF3</i>	0.022	0.033	0.295	0.007	0.153	0.226	0.154	0.626	0.088	0.305	0.144	0.109	0.534	0.053	0.254
c.*41G>A	<i>TFF3</i>	0.187	-	-	-	-	0.004	-	-	-	-	0.264	-	-	-	-
rs2236705	<i>TFF3</i>	0.040	0.118	0.120	0.032	0.363	0.839	0.881	0.608	0.728	0.973	0.051	0.056	0.305	0.021	0.175
V80M	<i>TFF3</i>	0.186	-	-	-	-	0.954	-	-	-	-	0.347	-	-	-	-
rs225363	<i>TFF3</i>	0.405	0.755	0.260	0.241	0.677	0.924	0.693	0.873	0.803	0.722	0.030	0.044	0.382	0.008	0.372
rs692356	<i>TFF3</i>	0.394	0.843	0.248	0.248	0.612	0.790	0.510	0.663	0.772	0.505	0.027	0.081	0.288	0.007	0.573
rs11203197	<i>TFF3</i>	0.952	0.822	0.867	0.788	0.861	0.345	0.387	0.381	0.238	0.583	0.320	0.188	0.556	0.163	0.228
rs12483635	<i>TFF2</i>	0.278	0.784	0.110	0.977	0.622	0.592	0.957	0.308	0.910	0.837	0.600	0.861	0.337	0.743	0.980
TFF2V66L	<i>TFF2</i>	0.683	-	-	-	-	0.597	-	-	-	-	0.418	-	-	-	-
rs1079380	<i>TFF2</i>	0.887	0.807	0.631	0.843	0.656	0.913	0.715	0.746	0.963	0.671	0.612	0.376	0.903	0.374	0.627
rs7277409	<i>TFF2</i>	0.492	-	-	-	-	0.716	-	-	-	-	0.635	-	-	-	-
rs3814896	<i>TFF2</i>	0.613	0.490	0.640	0.349	0.700	0.852	0.602	0.709	0.723	0.572	0.281	0.132	0.970	0.115	0.230

Polimorfismo	Gen	Cáncer gástrico					Difuso					Intestinal				
		Cod	Dom	Rec	Sob	Adi	Cod	Dom	Rec	Sob	Adi	Cod	Dom	Rec	Sob	Adi
c.*68 69insCTT	TFF2	0.857	-	-	-	-	0.740	-	-	-	-	0.249	-	-	-	-
rs225334	TFF2	0.962	0.782	0.895	0.857	0.795	0.918	0.864	0.680	0.894	0.741	0.540	0.308	0.437	0.646	0.268
rs13276	TFF1	0.791	0.501	0.984	0.494	0.529	0.927	0.849	0.775	0.784	0.915	0.069	0.021	0.633	0.025	0.025
rs225353	TFF1	0.318	0.176	0.880	0.159	0.413	0.996	0.951	0.969	0.930	0.986	0.378	0.249	0.748	0.176	0.552
rs225356	TFF1	0.021	0.207	0.053	0.006	0.743	0.567	0.523	0.537	0.289	0.944	0.001	0.027	0.044	2×10^{-4}	0.799
rs2839488	TFF1	0.902	0.723	0.702	0.950	0.654	0.993	0.907	0.950	0.947	0.909	0.022	0.385	0.006	0.182	0.033
rs13051704	TFF1	0.948	0.781	0.816	0.827	0.755	0.969	0.947	0.827	0.890	0.999	0.772	0.677	0.630	0.578	0.792
rs9976977	TFF1	0.382	0.487	0.173	0.834	0.241	0.579	0.711	0.296	0.742	0.432	0.013	0.081	0.005	0.778	0.007
rs424694	TFF1	0.042	0.068	0.325	0.012	0.537	0.144	0.120	0.567	0.055	0.495	0.025	0.049	0.298	0.007	0.487
rs13047838	TFF1	0.153	0.326	0.191	0.104	0.746	0.325	0.881	0.166	0.422	0.686	0.256	0.415	0.107	0.956	0.193

a, Valores p obtenidos mediante análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. En negrita, los valores p significativos ($p < 0.05$). Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante y Adi, log-aditivo.

Tabla A-11. Valores p de significación de la asociación entre polimorfismos en genes *MUC* y *TFF* con los subtipos anatómicos de cáncer gástrico bajo los diferentes modelos de herencia.

Polimorfismo	Gen	Cardias					No cardias				
		Cod	Dom	Rec	Sob	Adi	Cod	Dom	Rec	Sob	Adi
rs3814316	MUC1	0.086	0.055	0.099	0.486	0.026	0.404	0.542	0.183	0.784	0.271
rs9426886	MUC1	0.272	0.190	0.187	0.852	0.107	0.454	0.883	0.219	0.440	0.441
rs4072037	MUC1	0.076	0.036	0.938	0.044	0.176	0.184	0.281	0.299	0.066	0.954
rs1045253	MUC1	0.510	0.629	0.249	0.864	0.386	0.782	0.488	0.898	0.527	0.559
rs7481521	MUC6	0.724	0.529	0.796	0.430	0.790	0.182	0.105	0.169	0.700	0.066
rs11604757	MUC6	0.679	0.410	0.629	0.485	0.382	0.974	0.820	0.962	0.827	0.828
rs11246384	MUC6	0.730	0.674	0.564	0.570	0.788	0.105	0.119	0.084	0.251	0.063
rs6597947	MUC6	0.207	0.330	0.249	0.156	0.602	0.349	0.201	0.712	0.155	0.294
rs9794921	MUC6	0.318	0.132	0.563	0.198	0.140	0.194	0.071	0.514	0.121	0.082
rs10902073	MUC2	0.411	0.195	0.916	0.230	0.327	0.421	0.976	0.214	0.403	0.567
rs10794281	MUC2	0.646	0.359	0.903	0.445	0.497	0.776	0.479	0.755	0.673	0.521
rs2856082	MUC2	0.793	0.523	0.956	0.496	0.585	0.097	0.038	0.948	0.031	0.069
rs2071174	MUC2	0.140	0.486	0.121	0.094	0.832	0.150	0.094	0.671	0.053	0.306
rs11245936	MUC2	0.344	0.144	0.732	0.160	0.161	0.901	0.842	0.657	0.933	0.771
rs7944723	MUC2	0.776	0.982	0.490	0.773	0.787	0.206	0.076	0.531	0.124	0.092
rs10902089	MUC2	0.756	0.761	0.587	0.473	0.909	0.200	0.073	0.600	0.252	0.156
rs6421972	MUC2	0.624	0.524	0.622	0.332	0.902	0.076	0.023	0.436	0.129	0.053
rs11245954	MUC2	0.263	0.960	0.111	0.675	0.782	0.934	0.907	0.751	0.841	0.971
rs7480563	MUC2	0.421	0.194	0.521	0.560	0.239	0.710	0.408	0.797	0.622	0.509
rs7126405	MUC2	0.151	0.348	0.063	0.616	0.213	0.635	0.342	0.758	0.377	0.349
rs3924453	MUC2	0.945	0.826	0.859	0.750	0.919	0.640	0.747	0.346	0.864	0.520
rs8133510	TFF3	0.156	0.110	0.560	0.059	0.231	0.545	0.694	0.382	0.431	0.984
rs2236705	TFF3	0.757	0.960	0.480	0.760	0.865	0.042	0.070	0.208	0.021	0.234
V80M	TFF3	0.050	-	-	-	-	0.697	-	-	-	-
rs225363	TFF3	0.077	0.912	0.039	0.070	0.256	0.638	0.688	0.346	0.752	0.436
rs692356	TFF3	0.043	0.985	0.023	0.047	0.188	0.536	0.590	0.267	0.736	0.345
rs11203197	TFF3	0.803	0.886	0.509	0.965	0.818	0.789	0.726	0.509	0.851	0.632
c.*41G>A	TFF3	0.358	-	-	-	-	0.871	-	-	-	-
rs12483635	TFF2	0.193	0.084	0.429	0.104	0.076	0.300	0.526	0.193	0.387	0.695
V66L	TFF2	0.144	-	-	-	-	0.433	-	-	-	-
rs1079380	TFF2	0.394	0.586	0.174	0.452	0.234	0.200	0.684	0.123	0.106	0.514
rs7277409	TFF2	0.713	-	-	-	-	0.627	-	-	-	-
rs3814896	TFF2	0.887	0.977	0.636	0.842	0.840	0.318	0.160	0.908	0.132	0.275
c.*68_69insCTT	TFF2	0.202	-	-	-	-	0.756	-	-	-	-
rs225334	TFF2	0.828	0.651	0.825	0.539	0.838	0.932	0.709	0.926	0.763	0.760
rs13276	TFF1	0.348	0.150	0.599	0.182	0.149	0.910	0.683	0.956	0.666	0.715
rs225353	TFF1	0.394	0.176	0.739	0.291	0.272	0.898	0.927	0.693	0.688	0.872
rs225356	TFF1	0.184	0.768	0.107	0.100	0.415	0.532	0.331	0.816	0.285	0.605
rs2839488	TFF1	0.887	0.630	0.814	0.758	0.643	0.272	0.111	0.467	0.337	0.139
rs13051704	TFF1	0.969	0.863	0.830	0.906	0.834	0.926	0.810	0.812	0.745	0.882
rs9976977	TFF1	0.218	0.937	0.096	0.303	0.381	0.645	0.350	0.728	0.487	0.395
rs424694	TFF1	0.108	0.145	0.370	0.036	0.683	0.602	0.418	0.776	0.328	0.715
rs13047838	TFF1	0.469	0.265	0.864	0.219	0.381	0.705	0.957	0.432	0.675	0.802

a, Valores p obtenidos mediante análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. En negrita, los valores p significativos (p<0.05). Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante y Adi, log-aditivo.

Tabla A-12. SNPs en genes *TFF* que mostraron asociación significativa con cáncer gástrico estratificando por sexo.

SNP (gen)	Modelo	Genotipo	Cáncer gástrico				Hombres				Mujeres				P-Interact.
			Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	
rs225356 (<i>TFF1</i>)	Cod	CC	306/99	1.00			170/55	1.00			136/44	1.00			
		CT	601/144	0.74	(0.55-0.99)		355/91	0.78	(0.53 - 1.15)		246/53	0.67	(0.43 - 1.06)		
		TT	217/79	1.11	(0.79-1.57)	.0205	140/44	0.96	(0.61 - 1.53)	.3799	77/35	1.40	(0.82 - 2.38)	.0145	.2845
	Dom	CT	306/99	1.00			170/55	1.00			136/44	1.00			
		CT-TT	818/223	0.84	(0.64-1.10)	.2071	495/135	0.83	(0.58 - 1.20)	.3243	323/88	0.85	(0.56 - 1.29)	.4366	.9989
	Rec	CC-CT	907/243	1.00			525/146	1.00			382/97	1.00			
		TT	217/79	1.35	(1.00-1.81)	.0525	140/44	1.13	(0.77 - 1.67)	.5362	77/35	1.77	(1.11 - 2.82)	.0186	.1507
	Sob	CC-TT	523/178	1.00			310/99	1.00			213/79	1.00			
		CT	601/144	0.71	(0.55-0.91)	.0065	355/91	0.80	(0.57 - 1.10)	.1671	246/53	0.59	(0.39 - 0.88)	.0085	.2307
	Adi	CC,CT,TT	1124/322	1.03	(0.86-1.23)	.7431	665/190	0.97	(0.76 - 1.23)	.7939	459/132	1.13	(0.85 - 1.49)	.4003	.4386
rs13047838 (<i>TFF1</i>)	Cod	GG	792/235	1.00			478/143	1.00			314/92	1.00			
		GA	437/108	0.83	(0.64 - 1.07)		247/62	0.84	(0.60 - 1.18)		190/46	0.82	(0.54 - 1.23)		
		AA	55/22	1.33	(0.79 - 2.24)	.1528	34/9	0.88	(0.41 - 1.89)	.5891	21/13	2.10	(0.99 - 4.44)	.0659	.2406
	Dom	GG	792/235	1.00			478/143	1.00			314/92	1.00			
		GA-AA	492/130	0.88	(0.69 - 1.13)	.3260	281/71	0.85	(0.61 - 1.17)	.3070	211/59	0.94	(0.64 - 1.37)	.7475	.6358
	Rec	GG-GA	1229/343	1.00			725/205	1.00			504/138	1.00			
		AA	55/22	1.42	(0.85 - 2.37)	.1910	34/9	0.93	(0.44 - 1.99)	.8571	21/13	2.27	(1.09 - 4.73)	.0343	.0874
	Sob	GG-AA	847/257	1.00			512/152	1.00			335/105	1.00			
		GA	437/108	0.81	(0.63 - 1.05)	.1039	247/62	0.85	(0.61 - 1.18)	.3289	190/46	0.76	(0.51 - 1.13)	.1699	.2757
	Adi	GG,GA,AA	1284/365	0.97	(0.79 - 1.19)	.7457	759/214	0.88	(0.67 - 1.16)	.3569	525/151	1.10	(0.80 - 1.50)	.5674	.7124
rs3814896 (<i>TFF2</i>)	Cod	AA	808/222	1.00			495/123	1.00			313/99	1.00			
		AG	401/123	1.12	(0.87 - 1.44)		220/80	1.47	(1.06 - 2.03)		181/43	0.75	(0.50 - 1.13)		
		GG	70/18	0.92	(0.53 - 1.58)	.6134	39/9	0.93	(0.44 - 1.97)	.0597	31/9	0.90	(0.41 - 1.97)	.3893	.0394
	Dom	AA	808/222	1.00			495/123	1.00			313/99	1.00			
		AG-GG	471/141	1.09	(0.86 - 1.39)	.4904	259/89	1.39	(1.01 - 1.90)	.0415	212/52	0.78	(0.53 - 1.14)	.1894	.0215
	Rec	AA-AG	1209/345	1.00			715/203	1.00			494/142	1.00			
		GG	70/18	0.88	(0.52 - 1.50)	.6404	39/9	0.81	(0.38 - 1.70)	.5619	31/9	0.99	(0.45 - 2.14)	.9730	.7038
	Sob	AA-GG	878/240	1.00			534/132	1.00			344/108	1.00			
		AG	401/123	1.13	(0.88 - 1.44)	.3488	220/80	1.48	(1.07 - 2.04)	.0180	181/43	0.76	(0.51 - 1.14)	.1784	.0112
	Adi	AA,AG,GG	1279/363	1.04	(0.85 - 1.26)	.6997	754/212	1.21	(0.94 - 1.56)	.1446	525/151	0.84	(0.62 - 1.15)	.2822	.0809

SNP (gen)	Modelo	Genotipo	Cáncer gástrico				Hombres				Mujeres				P-Interact.
			Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	
rs8133510 (TFF3)	Cod	GG	811/249	1.00			484/150	1.00			327/99	1.00			
		GA	387/84	0.70	(0.53 - 0.92)		230/43	0.60	(0.41 - 0.87)		157/41	0.86	(0.57 - 1.30)		
		AA	45/17	1.23	(0.69 - 2.20)	.0216	22/11	1.61	(0.76 - 3.41)	.0065	23/6	0.86	(0.34 - 2.19)	.7488	.2221
	Dom	GG	811/249	1.00			484/150	1.00			327/99	1.00			
		GA-AA	432/101	0.76	(0.58 - 0.98)	.0328	252/54	0.69	(0.48 - 0.97)	.0308	180/47	0.86	(0.58 - 1.28)	.4469	.4065
	Rec	GG-GA	1198/333	1.00			714/193	1.00			484/140	1.00			
		AA	45/17	1.37	(0.77 - 2.43)	.2947	22/11	1.86	(0.88 - 3.93)	.1146	23/6	0.90	(0.36 - 2.28)	.8240	.2312
	Sob	GG-AA	856/266	1.00			506/161	1.00			350/105	1.00			
		GA	387/84	0.69	(0.53 - 0.91)	.0073	230/43	0.58	(0.40 - 0.85)	.0033	157/41	0.87	(0.57 - 1.31)	.4896	.1648
	Adi	GG,GA,AA	1243/350	0.85	(0.68 - 1.06)	.1527	736/204	0.83	(0.61 - 1.11)	.2010	507/146	0.89	(0.64 - 1.24)	.4752	.2307
rs2236705 (TFF3)	Cod	CC	914/245	1.00			544/138	1.00			370/107	1.00			
		CA	326/114	1.30	(1.01 - 1.68)		190/72	1.50	(1.08 - 2.09)		136/42	1.06	(0.70 - 1.59)		
		AA	39/6	0.57	(0.24 - 1.36)	.0405	22/4	0.70	(0.24 - 2.08)	.0396	17/2	0.42	(0.09 - 1.85)	.4116	.3912
	Dom	CC	914/245	1.00			544/138	1.00			370/107	1.00			
		CA-AA	365/120	1.22	(0.95 - 1.57)	.1175	212/76	1.42	(1.03 - 1.96)	.0365	153/44	0.99	(0.66 - 1.47)	.9444	.1747
	Rec	CC-CA	1240/359	1.00			734/210	1.00			506/149	1.00			
		AA	39/6	0.53	(0.22 - 1.26)	.1198	22/4	0.62	(0.21 - 1.84)	.3674	17/2	0.41	(0.09 - 1.81)	.1910	.6433
	Sob	CC-AA	953/251	1.00			566/142	1.00			387/109	1.00			
		CA	326/114	1.33	(1.03 - 1.71)	.0321	190/72	1.52	(1.09 - 2.11)	.0141	136/42	1.08	(0.72 - 1.63)	.7076	.2172
	Adi	CC,CA,AA	1279/365	1.11	(0.89 - 1.38)	.3630	756/214	1.25	(0.95 - 1.66)	.1184	523/151	0.92	(0.65 - 1.32)	.6590	.1882

a, Análisis de regresión logística no condicional ajustado por edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla A-13 SNPs de *TFF1* y región adyacente que mostraron asociación con cáncer gástrico de tipo intestinal estratificando por sexo.

SNP	Modelo	Genotipo	Intestinal				Hombres				Mujeres				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs13276	Cod	GG	1150/105	1.00			678/71	1.00			472/34	1.00			
		GA	126/20	1.89	(1.12-3.19)		78/11	1.46	(0.74-2.92)		48/9	2.95	(1.29-6.74)		
		AA	7/1	1.91	(0.23-15.89)	.0687	3/1	4.23	(0.40-44.29)	.3369	4/0	Na	Na	0.0409	.1739
	Dom	GG	1150/105	1.00			678/71	1.00			472/34	1.00			
		GA-AA	133/21	1.89	(1.13-3.16)	.0206	81/12	1.55	(0.80-3.02)	.2139	52/9	2.67	(1.18-6.07)	0.0279	.2930
	Rec	GG-GA	1276/125	1.00			756/82	1.00			520/43	1.00			
		AA	7/1	1.73	(0.21-14.39)	.6334	3/1	4.02	(0.39-41.86)	.3005	4/0	Na	Na	0.4037	.1892
	Sob	GG-AA	1157/106	1.00			681/72	1.00			476/34	1.00			
		GA	126/20	1.88	(1.11-3.17)	.0246	78/11	1.45	(0.73-2.88)	.3080	48/9	2.98	(1.30-6.81)	0.0161	.4633
	Adi	GG,GA,A	1283/126	1.77	(1.11-2.82)	.0246	759/83	1.58	(0.85-2.91)	.1633	524/43	2.08	(1.00-4.33)	0.0667	.1740
rs225356	Cod	GG	306/41	1.00			170/24	1.00			136/17	1.00			
		GA	601/38	0.47	(0.29-0.75)		355/27	0.52	(0.29-0.94)		246/11	0.39	(0.17-0.86)		
		AA	217/31	1.04	(0.62-1.72)	.0009	140/22	1.13	(0.60-2.12)	.0192	77/9	0.89	(0.37-2.16)	0.0441	.7579
	Dom	GG	306/41	1.00			170/24	1.00			136/17	1.00			
		GA-AA	818/69	0.62	(0.41-0.94)	.0273	495/49	0.69	(0.41-1.17)	.1753	323/20	0.52	(0.26-1.03)	0.0649	.4409
	Rec	GG-GA	907/79	1.00			525/51	1.00			382/28	1.00			
		AA	217/31	1.60	(1.02-2.51)	.0445	140/22	1.67	(0.98-2.88)	.0687	77/9	1.43	(0.63-3.26)	0.4052	.7652
	Sob	GG-AA	523/72	1.00			310/46	1.00			213/26	1.00			
		GA	601/38	0.46	(0.31-0.70)	.0002	355/27	0.49	(0.30-0.82)	.0053	246/11	0.40	(0.19-0.85)	0.0130	.5823
	Adi	GG,GA,A	1124/110	0.96	(0.72-1.28)	.7992	665/73	1.05	(0.73-1.49)	.8063	459/37	0.82	(0.50-1.34)	0.4235	.3891
rs2839488	Cod	GG	467/41	1.00			294/72	1.00			173/14	1.00			
		GC	603/51	0.97	(0.63-1.50)		345/108	1.28	(0.91-1.79)		258/13	0.61	(0.28-1.35)		
		CC	194/32	1.78	(1.13-3.08)	.0222	109/30	1.14	(0.70-1.84)	.3594	85/16	2.17	(0.98-4.79)	0.0075	.1183
	Dom	GG	467/41	1.00			294/72	1.00			173/14	1.00			
		GC-CC	797/83	1.19	(0.80-1.77)	.3850	454/138	1.24	(0.90-1.72)	.1805	343/29	1.00	(0.51-1.97)	0.9973	.6473
	Rec	GG-GC	1070/92	1.00			639/180	1.00			431/27	1.00			
		CC	194/92	1.89	(1.22-2.93)	.0058	109/30	0.99	(0.64-1.53)	.9561	85/16	2.85	(1.44-5.63)	0.0039	.1087
	Sob	GG-CC	661/73	1.00			403/102	1.00			258/30	1.00			
		GC	603/51	0.77	(0.53-1.13)	.1824	345/108	1.23	(0.91-1.68)	.1824	258/13	0.44	(0.22-0.86)	0.0132	.0408
	Adi	GG,GC,C	1264/124	1.33	(1.02-1.73)	.0333	748/210	1.11	(0.89-1.39)	.3606	516/43	1.49	(0.95-2.34)	0.0829	.4500

SNP	Modelo	Genotipo	Intestinal				Hombres				Mujeres				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs9976977	Cod	GG	563/45	1.00			336/30	1.00			227/15	1.00			
		GA	573/55	1.19	(0.79-1.80)		340/38	1.28	(0.77-2.12)		233/17	1.06	(0.51-2.20)		
		AA	145/26	2.26	(1.34-3.81)	0.0126	81/15	2.06	(2.06-1.05)	0.1199	64/11	2.53	(1.06-6.04)	0.0921	0.6912
	Dom	GG	536/45	1.00			336/30	1.00			227/15	1.00			
		GA-AA	718/81	1.40	(0.95-2.06)	0.0813	421/53	1.44	(0.89-2.31)	0.1302	297/28	1.35	(0.69-2.64)	0.3702	0.9918
	Rec	GG-GA	1136/100	1.00			676/68	1.00			460/32	1.00			
		AA	145/26	2.06	(1.28-3.30)	0.0045	81/15	1.81	(0.98-3.34)	0.0683	64/11	2.46	(1.14-5.30)	0.0293	0.4415
	Sob	GG-AA	708/71	1.00			417/45	1.00			291/26	1.00			
		GA	573/55	0.95	(0.65-1.38)	0.7783	340/38	1.06	(0.67-1.68)	0.7991	233/17	0.78	(0.41-1.49)	0.4545	0.4729
	Adi	GG,GA,A	1281/126	1.45	(1.11-1.89)	0.0068	757/83	1.40	(1.01-1.95)	0.0450	524/43	1.53	(0.97-2.41)	0.0713	0.6470
rs424694	Cod	CC	377/48	1.00			218/32	1.00			159/16	1.00			
		CT	639/47	0.58	(0.38-0.88)		386/29	0.51	(0.30-0.88)		253/18	0.70	(0.34-1.45)		
		TT	267/31	0.92	(0.57-1.50)	0.0246	155/22	0.95	(0.53-1.73)	0.0247	112/9	0.85	(0.35-2.04)	0.6348	0.6175
	Dom	CC	377/48	1.00			218/32	1.00			159/16	1.00			
		CT-TT	906/78	0.68	(0.46-0.99)	0.0488	541/51	0.64	(0.40-1.03)	0.0679	365/27	0.75	(0.38-1.45)	0.3940	0.8424
	Rec	CC-CT	1016/95	1.00			604/61	1.00			412/34	1.00			
		TT	267/31	1.26	(0.82-1.95)	0.2984	155/22	1.39	(0.82-2.36)	0.2297	112/9	1.04	(0.48-2.28)	0.9192	0.4278
	Sob	CC-TT	644/79	1.00			373/54	1.00			271/25	1.00			
		CT	639/47	0.60	(0.41-0.87)	0.0069	386/29	0.52	(0.33-0.84)	0.0066	253/18	0.75	(0.39-1.43)	0.3790	0.3780
	Adi	CC,CT,TT	1283/126	0.91	(0.70-1.19)	0.4872	759/83	0.92	(0.66-1.27)	0.6174	524/43	0.89	(0.57-1.40)	0.6122	0.7382
rs13047838	Cod	CC	792/73	1.00			478/52	1.00			314/21	1.00			
		CT	437/43	1.07	(0.72-1.60)		247/28	1.04	(0.64-1.70)		190/15	1.23	(0.60-2.50)		
		TT	55/10	1.92	(0.92-3.97)	0.2561	34/3	0.76	(0.22-2.60)	0.8800	21/7	5.04	(1.80-14.12)	0.0168	0.0319
	Dom	CC	792/73	1.00			478/52	1.00			314/21	1.00			
		CT-TT	492/53	1.17	(0.80-1.71)	0.4152	281/31	1.01	(0.63-1.62)	0.9744	211/22	1.59	(0.83-3.05)	0.1654	0.2444
	Rec	CC-CT	1229/116	1.00			725/80	1.00			504/36	1.00			
		TT	55/10	1.86	(0.91-3.81)	0.1068	34/3	0.75	(0.22-2.53)	0.6337	21/7	4.60	(1.73-12.20)	0.0051	0.0094
	Sob	CC-TT	847/83	1.00			512/55	1.00			335/28	1.00			
		CT	437/43	1.01	(0.68-1.50)	0.9563	247/28	1.06	(0.65-1.72)	0.8117	190/15	0.95	(0.48-1.86)	0.8754	0.7971
	Adi	CC,CT,TT	1284/126	1.23	(0.90-1.67)	0.1935	759/83	0.97	(0.65-1.45)	0.8885	525/43	1.85	(1.12-3.07)	0.0187	0.0371

a, Análisis de regresión logística no condicional ajustado por edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla A-16. SNPs de *TFF3* que mostraron asociación significativa con cáncer gástrico del cardias estratificando por sexo.

SNP	Modelo	Genotipo	Cardias				Hombres				Mujeres				P-Interact.
			Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	Controles/casos	OR	(CI 95%) ^a	Valor p	
rs8133510	Cod	GG	811/76	1.00			484/62	1.00			327/14	1.00			
		GA	387/22	0.64	(0.39 - 1.05)		230/13	0.47	(0.25 - 0.88)		157/9	1.21	(0.49 - 2.99)		
		AA	45/4	1.23	(0.42 - 3.59)	.1563	22/4	1.75	(0.56 - 5.40)	.0209	23/0	Na	Na	.2939	.0409
	Dom	GG	811/76	1.00			484/62	1.00			327/14	1.00			
		GA-AA	432/26	0.69	(0.43 - 1.10)	.1096	252/17	0.57	(0.32 - 1.00)	.0424	180/9	1.04	(0.43 - 2.55)	.9308	.2055
	Rec	GG-GA	1198/98	1.00			714/75	1.00			484/23	1.00			
		AA	45/4	1.39	(0.48 - 4.04)	.5597	22/4	2.12	(0.69 - 6.50)	.2236	23/0	Na	Na	.1316	.0761
	Sob	GG-AA	856/80	1.00			506/66	1.00			350/14	1.00			
		GA	387/22	0.63	(0.38 - 1.03)	.0586	230/13	0.46	(0.25 - 0.86)	.0087	157/9	1.31	(0.53 - 3.22)	.5595	.0497
	Adi	GG,GA,AA	1243/102	0.78	(0.52 - 1.18)	.2310	736/79	0.72	(0.44 - 1.18)	.1780	507/23	0.88	(0.41 - 1.90)	.7390	.5542
rs225363	Cod	GG	421/34	1.00			238/28	1.00			183/6	1.00			
		CG	630/43	0.80	(0.50 - 1.29)		377/31	0.67	(0.39 - 1.16)		253/12	1.37	(0.49 - 3.81)		
		CC	224/29	1.45	(0.85 - 2.47)	.0775	137/24	1.42	(0.79 - 2.59)	.0365	87/5	1.72	(0.49 - 6.03)	.6835	.3912
	Dom	GG	421/34	1.00			238/28	1.00			183/6	1.00			
		GC-CC	854/72	0.98	(0.63 - 1.51)	.9125	514/55	0.87	(0.54 - 1.42)	.5840	340/17	1.46	(0.55 - 3.85)	.4362	.3283
	Rec	GG-GC	1051/77	1.00			615/59	1.00			436/18	1.00			
		CC	224/29	1.65	(1.04 - 2.62)	.0385	137/24	1.79	(1.07 - 3.01)	.0323	87/5	1.41	(0.49 - 4.07)	.5346	.6272
	Sob	GG-CC	645/63	1.00			375/52	1.00			270/11	1.00			
		GC	630/43	0.69	(0.46 - 1.03)	.0701	377/31	0.58	(0.36 - 0.93)	.0216	253/12	1.11	(0.47 - 2.63)	.8072	.1682
	Adi	GG,GC,CC	1275/106	1.18	(0.89 - 1.57)	.2565	752/83	1.16	(0.84 - 1.61)	.3681	523/23	1.32	(0.71 - 2.44)	.3861	.7459
rs692356	Cod	AA	422/34	1.00			238/28	1.00			184/6	1.00			
		AG	472/31	0.76	(0.45 - 1.27)		296/23	0.63	(0.35 - 1.13)		176/8	1.43	(0.47 - 4.38)		
		GG	222/30	1.51	(0.89 - 2.57)	.0433	134/24	1.42	(0.78 - 2.58)	.0323	88/6	2.03	(0.61 - 6.78)	.5157	.4688
	Dom	AA	422/34	1.00			238/28	1.00			184/6	1.00			
		AG-GG	694/61	1.00	(0.64 - 1.57)	.9846	430/47	0.88	(0.53 - 1.45)	.6128	264/14	1.64	(0.60 - 4.51)	.3241	.2993
	Rec	AA-AG	894/65	1.00			534/51	1.00			360/14	1.00			
		GG	222/30	1.75	(1.09 - 2.79)	.0232	134/24	1.80	(1.06 - 3.05)	.0347	88/6	1.68	(0.60 - 4.69)	.3373	.8830
	Sob	AA-GG	644/64	1.00			372/52	1.00			272/12	1.00			
		AG	472/31	0.64	(0.40 - 1.00)	.0466	296/23	0.54	(0.32 - 0.91)	.0183	176/8	1.07	(0.41 - 2.77)	.8890	.2476
	Adi	AA,AG,GG	1116/95	1.21	(0.91 - 1.61)	.1883	668/75	1.16	(0.84 - 1.61)	.3680	448/20	1.42	(0.78 - 2.60)	.2498	.5943

a, Análisis de regresión logística no condicional ajustado edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla A-17. SNPs de *TFFs* que mostraron asociación con cáncer gástrico no cardias, estratificando por la infección por *H.pylori*.

SNP (gen)	Modelo	Genotipo	Hp+				Hp-				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs2839488 (<i>TFF1</i>)	Cod	GG	276/50	1.00			187/2	1.00			
		GC	368/80	1.22	(0.83-1.81)		226/11	5.34	(1.14-25.01)		
		CC	122/32	1.45	(0.87-2.39)	.3319	72/1	1.85	(0.16-21.12)	.0406	.0987
	Dom	GG	276/50	1.00			187/2	1.00			
		GC-CC	490/112	1.28	(0.88-1.85)	.1883	298/12	4.59	(0.99-21.21)	.0239	.1682
	Rec	GG-GC	644/130	1.00			413/13	1.00			
		CC	122/32	1.28	(0.82-2.00)	.2777	72/1	0.57	(0.07-4.52)	.5688	.2015
	Sob	GG-CC	398/82	1.00			259/3	1.00			
		GC	368/80	1.08	(0.76-1.52)	.6707	226/11	4.51	(1.21-16.78)	.0129	.0324
	Adi	GG,GC,C	766/162	1.21	(0.94-1.54)	.1384	485/14	1.68	(0.76-3.69)	.2017	.8364
rs13051704 (<i>TFF1</i>)	Cod	GG	631/136	1.00			407/9	1.00			
		GC	133/24	0.82	(0.51-1.33)		81/5	2.76	(0.87-8.82)		
		CC	13/2	0.59	(0.12-3.00)	.5852	3/1	94.8	(2.94-Na)	.0202	.0331
	Dom	GG	631/136	1.00			407/9	1.00			
		GC-CC	146/26	0.80	(0.50-1.27)	.3373	84/6	2.76	(1.12-10.12)	.0392	.0208
	Rec	GG-GC	764/160	1.00			488/14	1.00			
		CC	13/2	0.61	(0.12-3.09)	.5317	3/1	72.5	(2.47-Na)	.0234	.0781
	Sob	GG-CC	644/138	1.00			410/10	1.00			
		GC	133/24	0.83	(0.51-1.34)	.4290	81/5	2.51	(0.80-7.85)	.1316	.0725
	Adi	GG,GC,C	777/162	0.81	(0.53-1.23)	.3043	491/15	3.87	(1.42-10.54)	.0126	.0098
rs9976977 (<i>TFF1</i>)	Cod	GG	327/67	1.00			231/4	1.00			
		GA	361/73	1.00	(0.69-1.44)		205/10	3.56	(1.04-12.21)		
		AA	89/23	1.13	(0.66-1.95)	.8880	56/1	1.13	(0.12-10.73)	.0847	.1252
	Dom	GG	327/67	1.00			231/4	1.00			
		GA-AA	450/96	1.03	(0.72-1.45)	.8890	261/11	2.93	(0.89-9.62)	.0609	.1436
	Rec	GG-GA	688/140	1.00			436/56	1.00			
		AA	89/23	1.14	(0.68-1.89)	.6263	56/1	0.57	(0.07-4.57)	.5692	.3564
	Sob	GG-AA	416/90	1.00			287/5	1.00			
		GA	361/73	0.97	(0.68-1.37)	.8509	205/10	3.48	(1.10-11.03)	.0265	.0427
	Adi	GG,GA,AA	777/163	1.05	(0.81-1.35)	.7332	492/15	1.51	(0.72-3.15)	.2791	.5547

SNP (gen)	Modelo	Genotipo	Hp+			Hp-			P Interact	
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR		(CI 95%) ^a
rs3814896 (TFF2)	Cod	AA	493/95	1.00			309/6	1.00		
		AG	239/59	1.28	(0.89 - 1.85)		156/7	2.33	(0.75 - 7.23)	
		GG	46/8	0.82	(0.35 - 1.89)	.3216	23/2	5.46	(0.99 - 30.13)	.1246
	Dom	AA	493/95	1.00			309/6	1.00		
		AG-GG	285/67	1.21	(0.85 - 1.73)	.2912	179/9	2.67	(0.91 - 7.83)	.0693
	Rec	AA-AG	732/154	1.00			465/13	1.00		
		GG	46/8	0.74	(0.32 - 1.70)	.4702	23/2	3.77	(0.76 - 18.64)	.1522
	Sob	AA-GG	539/103	1.00			332/8	1.00		
		AG	239/59	1.30	(0.91 - 1.87)	.1536	156/7	1.84	(0.64 - 5.28)	.2659
	Adi	AA,AG,GG	778/162	1.09	(0.82 - 1.45)	.5562	488/15	2.33	(1.06 - 5.12)	.0413
rs225334 (TFF2)	Cod	TT	309/63	1.00			206/5	1.00		
		TC	322/68	1.05	(0.72 - 1.55)		202/5	0.95	(0.27 - 3.40)	
		CC	125/21	0.86	(0.50 - 1.50)	.7725	65/5	3.40	(0.91 - 12.77)	.1422
	Dom	TT	309/63	1.00			206/5	1.00		
		TC-CC	447/89	1.00	(0.70 - 1.44)	.9882	267/10	1.48	(0.49 - 4.48)	.4777
	Rec	TT-TC	631/131	1.00			408/10	1.00		
		CC	125/21	0.84	(0.50 - 1.41)	.5053	65/5	3.48	(1.09 - 11.11)	.0484
	Sob	TT-CC	434/84	1.00			271/10	1.00		
		TC	322/68	1.09	(0.76 - 1.57)	.6210	202/5	0.62	(0.21 - 1.88)	.3909
	Adi	TT,TC,CC	756/152	0.96	(0.75 - 1.23)	.7449	473/15	1.81	(0.87 - 3.78)	.1142
rs2236705 (TFF3)	Cod	CC	550/110	1.00			356/6	1.00		
		CA	200/50	1.33	(0.91 - 1.95)		121/9	4.51	(1.54 - 13.17)	
		AA	26/3	0.57	(0.17 - 1.95)	.1834	13/0	NA	NA	.0165
	Dom	CC	550/110	1.00			356/6	1.00		
		CA-AA	226/53	1.24	(0.86 - 1.80)	.2586	134/9	4.21	(1.44 - 12.30)	.0081
	Rec	CC-CA	750/160	1.00			477/15	1.00		
		AA	26/3	0.53	(0.16 - 1.78)	.2632	13/0	NA	NA	.4633
	Sob	CC-AA	576/113	1.00			369/6	1.00		
		CA	200/50	1.36	(0.93 - 1.99)	.1150	121/9	4.62	(1.58 - 13.50)	.0049
	Adi	CC,CA,AA	776/163	1.11	(0.81 - 1.52)	.5328	490/15	2.78	(1.17 - 6.61)	.0269

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Sob, sobredominante; Adi, aditivo.

Tabla A-18. SNPs en genes *MUC* que mostraron asociación con cáncer gástrico estratificando por sexo.

SNP (gen)	Modelo	Genotipo	Cáncer gástrico				Hombres				Mujeres				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs3814316 (<i>MUC1</i>)	Cod	CC	566/143	1.00			319/79	1.00			247 /64	1.00			
		CT	573/162	1.13	(0.87-1.46)		353/101	1.15	(0.83 - 1.61)		220 /61	1.08	(0.72 - 1.62)		
		TT	142/59	1.66	(1.16-2.38)	.0246	84/33	1.59	(0.99 - 2.57)	.1707	58 /26	1.76	(1.02 - 3.04)	.1328	.8897
	Dom	CC	566/143	1.00			319/79	1.00			247 /64	1.00			
		CT-TT	715/221	1.23	(0.97-1.57)	.0894	437/134	1.24	(0.90 - 1.70)	.1883	278 /87	1.22	(0.84 - 1.77)	.2978	.9933
	Rec	CC-CT	1139/305	1.00			672/180	1.00			467 /125	1.00			
		TT	142/59	1.56	(1.12-2.17)	.0104	84/33	1.47	(0.95 - 2.28)	.0930	58 /26	1.69	(0.84 - 1.77)	.0483	.6473
	Adi	CC,CT,TT	1281/364	1.25	(1.05-1.48)	.0120	756/213	1.23	(0.98 - 1.55)	.0713	525 /151	1.27	(0.97 - 1.65)	.0833	.8474
rs4072037 (<i>MUC1</i>)	Cod	AA	337/118	1.00			202/77	1.00			135/41	1.00			
		AG	660/157	0.68	(0.51-0.89)		394/84	0.56	(0.39 - 0.79)		266/73	0.90	(0.58 - 1.39)		
		GG	276/86	0.98	(0.65-1.23)	.0141	157/50	0.84	(0.56 - 1.28)	.0036	119/36	0.99	(0.59 - 1.67)	.8468	.2397
	Dom	AA	337/118	1.00			202/77	1.00			135/41	1.00			
		AG-GG	936/243	0.74	(0.57-0.95)	.0211	551/134	0.64	(0.46 - 0.88)	.0075	385/109	0.93	(0.61 - 1.40)	.7156	.1740
	Rec	AA-AG	997/275	1.00			596/161	1.00			401/114	1.00			
		GG	276/86	1.14	(0.86-1.50)	.3733	157/50	1.19	(0.83 - 1.72)	.3432	119/36	1.07	(0.69 - 1.65)	.7156	.6848
	Adi	AA,AG,GG	1273/361	0.92	(0.78-1.09)	.3382	753/211	0.88	(0.70 - 1.09)	.2334	520/150	0.99	(0.76 - 1.29)	.9636	.4887
rs2071174 (<i>MUC2</i>)	Cod	TT	559/139	1.00			326/86	1.00			233/53	1.00			
		CT	521/169	1.32	(1.02 - 1.71)		307/93	1.16	(0.83 - 1.63)		214/76	1.62	(1.08 - 2.42)		
		CC	151/33	0.89	(0.58 - 1.35)	.0403	94/20	0.81	(0.47 - 1.40)	.3617	57/13	1.04	(0.52 - 2.04)	.0473	.4411
	Dom	TT	559/139	1.00			326/86	1.00			233/53	1.00			
		TC-CC	672/202	1.23	(0.96 - 1.57)	.1029	401/113	1.08	(0.79 - 1.49)	.6255	271/89	1.50	(1.02 - 2.21)	.0392	.1946
	Rec	TT-TC	1080/308	1.00			633/179	1.00			447/129	1.00			
		TC	521/169	1.36	(1.07 - 1.73)	.0135	307/93	1.22	(0.89 - 1.67)	.2273	214/76	1.61	(1.10 - 2.35)	.0136	.2540
	Adi	TT,CT,CC	1231/341	1.05	(0.88 - 1.26)	.5674	727/199	0.98	(0.78 - 1.24)	.8640	504/142	1.19	(0.90 - 1.58)	.2245	.2893

a, Análisis de regresión logística no condicional ajustado por edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla A-19. SNPs en genes *MUC* que mostraron asociación con cáncer gástrico del tipo intestinal estratificando por sexo.

SNP (gen)	Modelo	Genotipo	Controles/ casos	Intestinal			Controles/ casos	Hombres			Controles/ casos	Mujeres			P Interact.
				OR	(CI 95%) ^a	Valor p		OR	(CI 95%) ^a	Valor p		OR	(CI 95%) ^a	Valor p	
rs3814316 (<i>MUC1</i>)	Cod	CC	566/46	1.00			319/26	1.00			247/20	1.00			
		CT	573/57	1.28	(0.85-1.94)		353/42	1.53	(0.91 - 2.57)		220/15	0.94	(0.46 - 1.93)		
		TT	142/23	2.17	(1.26-3.75)	.0261	84/15	2.33	(1.16 - 4.65)	.0507	58/8	2.17	(0.88 - 5.35)	.2120	.5258
	Dom	CC	566/46	1.00			319/26	1.00			247/20	1.00			
		CT-TT	715/80	1.45	(0.98-2.13)	.0583	437/57	1.68	(1.02 - 2.74)	.0360	278/23	1.18	(0.62 - 2.24)	.6229	.3440
	Rec	CC-CT	1139/103	1.00			672/68	1.00			467/35	1.00			
		TT	142/23	1.90	(1.16-3.11)	.0150	84/15	1.81	(0.98 - 3.34)	.0683	58/8	2.23	(0.96 - 5.18)	.0794	.8245
Adi	CC,CT,TT	1281/126	1.44	(1.10-1.88)	.0092	756/83	1.53	(1.09 - 2.14)	.0146	525/43	1.34	(0.84 - 2.13)	.2196	.5571	
rs9426886 (<i>MUC1</i>)	Cod	AA	440/41	1.00			258/25	1.00			182/16	1.00			
		AT	631/55	0.95	(0.62-1.46)		374/38	1.06	(0.62 - 1.81)		257/17	0.83	(0.40 - 1.72)		
		TT	199/30	1.72	(1.04-2.87)	.0503	118/20	1.83	(0.97 - 3.46)	.1427	81/10	1.65	(0.70 - 3.90)	.2945	.8657
	Dom	AA	440/41	1.00			258/25	1.00			182/16	1.00			
		AT-TT	830/85	1.13	(0.76-1.68)	.5419	492/58	1.24	(0.76 - 2.04)	.3866	338/27	1.02	(0.53 - 1.98)	.9552	.6196
	Rec	AA-AT	1071/96	1.00			632/63	1.00			439/33	1.00			
		TT	199/30	1.77	(1.14-2.76)	.0149	118/20	1.77	(1.02 - 3.06)	.0500	81/10	1.83	(0.85 - 3.94)	.1382	.9740
Adi	AA,AT,TT	1270/126	1.27	(0.98-1.67)	.0753	750/83	1.33	(0.96 - 1.85)	.0916	520/43	1.22	(0.77 - 1.93)	.3931	.7072	
rs4072037 (<i>MUC1</i>)	Cod	AA	337/43	1.00			202/32	1.00			135/11	1.00			
		AG	660/51	0.60	(0.39-0.92)		394/32	0.50	(0.30 - 0.84)		266/19	0.82	(0.37 - 1.79)		
		GG	276/32	0.90	(0.55-1.47)	.0419	157/19	0.75	(0.41 - 1.38)	.0328	119/13	1.18	(0.50 - 2.79)	.6308	.5050
	Dom	AA	337/43	1.00			202/32	1.00			135/11	1.00			
		AG-GG	936/83	0.68	(0.46-1.01)	.0620	551/51	0.57	(0.35 - 0.92)	.0232	385/32	0.93	(0.45 - 1.92)	.8495	.2302
	Rec	AA-AG	997/94	1.00			596/64	1.00			401/30	1.00			
		GG	276/32	1.23	(0.80-1.88)	.3512	157/19	1.12	(0.65 - 1.94)	.6784	119/13	1.35	(0.67 - 2.71)	.4117	.7171
Adi	AA,AG,GG	1273/126	0.92	(0.70-1.19)	.5142	753/83	0.81	(0.58 - 1.13)	.2074	520/43	1.10	(0.70 - 1.72)	.6915	.2761	
rs10902073 (<i>MUC2</i>)	Cod	CC	532/46	1.00			318/29	1.00			214/17	1.00			
		CA	578/70	1.40	(0.95 - 2.08)		336/48	1.58	(0.97 - 2.57)		242/22	1.15	(0.59 - 2.25)		
		AA	168/10	0.69	(0.34 - 1.40)	.0480	100/6	0.65	(0.26 - 1.62)	.0411	68/4	0.76	(0.24 - 2.39)	.7406	.6436
	Dom	CC	532/46	1.00			318/29	1.00			214/17	1.00			
		CA-AA	746/80	1.24	(0.85 - 1.82)	.2613	436/54	1.37	(0.85 - 2.20)	.1931	310/26	1.07	(0.56 - 2.04)	.8430	.5362
	Rec	CC-CA	1110/116	1.00			654/77	1.00			456/39	1.00			
		AA	168/10	0.57	(0.29 - 1.11)	.0756	100/6	0.50	(0.21 - 1.19)	.0872	68/4	0.71	(0.24 - 2.07)	.5116	.6064
Adi	CC,CA,AA	1278/126	1.00	(0.76 - 1.31)	.9903	754/83	1.03	(0.73 - 1.44)	.8784	524/43	0.96	(0.60 - 1.54)	.8688	.8296	
rs11245936 (<i>MUC2</i>)	Cod	GG	1065/110	1.00			625/75	1.00			440/35	1.00			
		GA	190/12	0.58	(0.31 - 1.07)		115/6	0.41	(0.17 - 0.97)		75/6	0.94	(0.38 - 2.37)		
		AA	12/0	Na	Na	.0559	10/0	Na	Na	.0272	2/0	Na	Na	.8265	.4257
	Dom	GG	1065/122	1.00			625/75	1.00			440/35	1.00			
		GA-AA	202/12	0.54	(0.29 - 1.01)	.0379	125/6	0.38	(0.16 0.89)	.0125	77/6	0.92	(0.37 - 2.30)	.8552	.1659
	Rec	GG-GA	1255/122	1.00			740/81	1.00			515/41	1.00			
		AA	12/0	Na	Na	.1252	10/0	Na	Na	.1446	2/0	Na	Na	.5451	1
Adi	GG,GA,AA	1267/122	0.54	(0.30 - 0.97)	.0261	750/81	0.38	(0.17 - 0.88)	.0090	517/41	0.89	(0.37 - 2.18)	.8002	.1739	

a, Análisis de regresión logística no condicional ajustado por edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla A-20. SNPs en *MUC2* que mostraron asociación significativa con cáncer gástrico del cardias en la estratificación por sexo.

SNP	Modelo	Genotipo	Controles/ casos	Cardias			Controles/ casos	Hombres			Controles/ casos	Mujeres			P Interact.
				OR	(CI 95%) ^a	Valor p		OR	(CI 95%) ^a	Valor p		OR	(CI 95%) ^a	Valor p	
rs2071174	Cod	CC	559/41	1.00			326/32	1.00			233/9	1.00			
		CT	521/50	1.32	(0.85 - 2.05)		307/40	1.31	(0.80 - 2.16)		214/10	1.37	(0.52 - 3.59)		
		TT	151/8	0.66	(0.30 - 1.46)	.1396	94/4	0.38	(0.13 - 1.11)	.0296	57/4	2.81	(0.77 - 10.22)	.3271	.1057
	Dom	CC	559/41	1.00			326/32	1.00			233/9	1.00			
		CT-TT	672/58	1.16	(0.76 - 1.78)	.4862	401/44	1.08	(0.66 - 1.75)	.7691	271/14	1.60	(0.65 - 3.94)	.3030	.4986
	Rec	CC-CT	1080/91	1.00			633/72	1.00			447/19	1.00			
		TT	151/8	0.57	(0.27 - 1.22)	.1209	94/4	0.33	(0.12 - 0.93)	.0151	57/4	2.39	(0.74 - 7.72)	.1759	.0337
Adi	CC,CT,TT	1231/99	0.97	(0.71 - 1.31)	.8324	727/76	0.85	(0.60 - 1.22)	.3818	504/23	1.61	(0.84 - 3.07)	.1530	.1486	
rs11245936	Cod	GG	1065/93	1.00			625/74	1.00			440/19	1.00			
		GA	190/10	0.63	(0.32 - 1.24)		115/5	0.40	(0.16 - 1.02)		75/5	1.24	(0.43 - 3.59)		
		AA	12/1	0.67	(0.08 - 5.31)	.3438	10/1	0.74	(0.09 - 5.96)	.0975	2/0	NA	NA	.7817	.2200
	Dom	GG	1065/93	1.00			625/74	1.00			440/19	1.00			
		GA-AA	202/11	0.63	(0.33 - 1.21)	.1443	125/6	0.43	(0.18 - 1.03)	.0359	77/5	1.20	(0.42 - 3.47)	.7380	.1129
	Rec	GG-GA	1255/103	1.00			740/79	1.00			515/24	1.00			
		AA	12/1	0.71	(0.09 - 5.62)	.7316	10/1	0.81	(0.10 - 6.55)	.8428	2/0	NA	NA	.5632	.6717
Adi	GG,GA,AA	1267/104	0.67	(0.37 - 1.21)	.1609	750/80	0.51	(0.24 - 1.10)	.0572	517/24	1.13	(0.41 - 3.14)	.8131	.1737	
rs7480563	Cod	GG	341/23	1.00			216/15	1.00			125/8	1.00			
		GA	617/52	1.35	(0.80 - 2.26)		351/40	1.68	(0.90 - 3.13)		266/12	0.76	(0.29 - 1.98)		
		AA	311/28	1.42	(0.79 - 2.55)	.4210	181/24	1.98	(1.00 - 3.92)	.1106	130/4	0.44	(0.12 - 1.64)	.4423	.1092
	Dom	GG	341/23	1.00			216/15	1.00			125/8	1.00			
		GA-AA	928/80	1.37	(0.84 - 2.23)	.1941	532/64	1.78	(0.99 - 3.21)	.0441	396/16	0.66	(0.26 - 1.63)	.3736	.0710
	Rec	GG-GA	958/75	1.00			567/55	1.00			391/20	1.00			
		AA	311/28	1.17	(0.73 - 1.85)	.5208	181/24	1.40	(0.84 - 2.35)	.0440	130/4	0.53	(0.16 - 1.69)	.2504	.1061
Adi	GG,GA,AA	1269/103	1.19	(0.89 - 1.58)	.2390	748/79	1.39	(1.00 - 1.92)	.0489	521/24	0.68	(0.37 - 1.26)	.2152	.0398	

a, Análisis de regresión logística no condicional ajustado por edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla A-21. SNPs en genes *MUC* que mostraron asociación significativa con cáncer gástrico no cardias en la estratificación por la infección por *H.pylori*.

SNP (gen)	Modelo	Genotipo	Hp+				Hp-				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs2071174 (<i>MUC2</i>)	Cod	TT	341/58	1.00			209/6	1.00			
		TC	327/75	1.35	(0.92 - 1.99)		192/8	1.51	(0.50 4.57)		
		CC	78/18	1.43	(0.79 - 2.60)	.2355	71/0	NA	NA	.0617	.0760
	Dom	TT	341/58	1.00			209/6	1.00			
		TC-CC	405/93	1.37	(0.95 - 1.97)	.0911	263/8	1.04	(0.35 3.11)	.9396	.6448
	Rec	TT-TC	668/133	1.00			401/14	1.00			
		CC	78/18	1.22	(0.70 - 2.13)	.4943	71/0	NA	NA	.0250	.0238
Adi	TT,TC,CC	746/151	1.24	(0.95 - 1.62)	.1131	472/14	0.72	(0.33 1.61)	.4178	.2132	
rs2856082 (<i>MUC2</i>)	Cod	CC	570/102	1.00			344/13	1.00			
		CG	135/42	1.85	(1.22 - 2.80)		106/1	0.22	(0.03 1.76)		
		GG	13/3	1.41	(0.38 - 5.18)	.0171	10/0	NA	NA	.1579	.0453
	Dom	CC	570/102	1.00			344/13	1.00			
		CG-GG	148/45	1.81	(1.21 - 2.71)	.0048	116/1	0.20	(0.03 1.60)	.0611	.0139
	Rec	CC-CG	705/144	1.00			450/14	1.00			
		GG	13/3	1.21	(0.33 - 4.42)	.7790	10/0	NA	NA	.4135	.4225
Adi	CC,CG,GG	718/147	1.61	(1.13 - 2.30)	.0010	460/14	0.21	(0.03 1.61)	.0561	.0160	
rs6421972 (<i>MUC2</i>)	Cod	GG	220/58	1.00			147/6	1.00			
		GA	366/62	0.63	(0.42 - 0.94)		220/6	0.70	(0.22 2.28)		
		AA	146/30	0.77	(0.47 - 1.27)	.0808	104/1	0.24	(0.03 2.05)	.3219	.3902
	Dom	GG	220/58	1.00			147/6	1.00			
		GA-AA	512/92	0.67	(0.46 - 0.97)	.0369	324/7	0.55	(0.18 1.70)	.3054	.6383
	Rec	GG-GA	586/120	1.00			367/12	1.00			
		AA	146/30	1.01	(0.64 - 1.58)	.9729	104/1	0.29	(0.04 2.28)	.1653	.1678
Adi	GG,GA,AA	732/150	0.84	(0.65 - 1.08)	.1714	471/13	0.56	(0.24 1.29)	.1595	.2956	
rs11246384 (<i>MUC6</i>)	Cod	CC	627/122	1.00			397/9	1.00			
		CT	108/26	1.25	(0.77 - 2.02)		68/4	2.68	(0.78 - 9.24)		
		TT	3/2	3.63	(0.57 -	.3097	3/1	19.7	(1.43 -	.0730	.3927
	Dom	CC	627/122	1.00			397/9	1.00			
		CT-TT	111/28	1.31	(0.82 - 2.09)	.2713	71/5	3.24	(1.02 -	.0602	.1717
	Rec	CC-CT	735/148	1.00			465/13	1.00			
		TT	3/2	3.46	(0.55 -	.2123	3/1	15.6	(1.18 -	.0793	.4017
Adi	CC,CT,TT	738/150	1.35	(0.87 - 2.09)	.1912	468/14	3.32	(1.24 - 8.88)	.0270	.1220	

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla A-22. SNPs en genes *MUC* que mostraron asociación significativa con cáncer gástrico no cardias en los infectados por *H.pylori* cuando se estratificó por la presencia de CagA.

SNP	Modelo	Genotipo	CagA+				CagA-				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs7481521	Cod	AA	190/40	1.00			56/2	1.00			
		AG	266/67	1.26	(0.81 - 1.96)		91/6	1.57	(0.29 - 8.52)		
		GG	100/33	1.76	(1.02 - 3.03)	.1278	41/4	2.44	(0.40 - 14.71)	.6069	.8939
	Dom	AA	190/40	1.00			56/2	1.00			
		AG-GG	366/100	1.38	(0.91 - 2.10)	.1225	132/10	1.85	(0.38 - 9.10)	.4237	.6170
	Rec	AA-AG	456/107	1.00			147/8	1.00			
		GG	100/33	1.53	(0.96 - 2.44)	.0800	41/4	1.80	(0.47 - 6.84)	.3984	.8598
	Adi	AA,AG,GG	556/140	1.32	(1.01 - 1.73)	.0444	188/12	0.96	(0.64 - 3.77)	.3176	.6842
	rs9794921	Cod	GG	376/85	1.00			107/5	1.00		
GT			167/56	1.49	(1.00 - 2.21)		73/6	1.68	(0.45 - 6.21)		
TT			22/8	1.70	(0.68 - 4.27)	.1084	10/1	2.26	(0.19 - 27.03)	.6771	.9773
Dom		GG	376/85	1.00			107/5	1.00			
		GT-TT	189/64	1.51	(1.03 - 2.21)	.0367	83/7	1.73	(0.48 - 6.21)	.3950	.8271
Rec		GG-GT	543/141	1.00			180/11	1.00			
		TT	22/8	1.47	(0.59 - 3.63)	.4200	10/1	1.69	(0.16 - 17.72)	.6764	.9906
Adi		GG,GT,TT	565/149	1.41	(1.02 - 1.94)	.0398	190/12	1.58	(0.57 - 4.39)	.3838	.8825
rs2856082		Cod	CC	433/97	1.00			137/5	1.00		
	CG		100/36	1.77	(1.13 - 2.79)		35/6	5.59	(1.44 - 21.75)		
	GG		8/3	1.89	(0.46 - 7.74)	.0416	5/0	Na	Na	.0283	.2396
	Dom	CC	433/97	1.00			137/5	1.00			
		CG-GG	108/39	1.78	(1.15 - 2.77)	.0117	40/6	4.27	(1.16 - 15.77)	.0305	.2006
	Rec	CC-CG	533/133	1.00			172/11	1.00			
		GG	8/3	1.65	(0.41 - 6.70)	.4980	5/0	Na	NA	.3275	.3791
	Adi	CC,CG,GG	541/136	1.65	(1.11 - 2.44)	.0146	177/11	2.29	(0.83 - 6.36)	.1242	.4231
	rs2071174	Cod	TT	260/57	1.00			81/1	1.00		
TC			243/64	1.21	(0.81 - 1.83)		84/11	11.5	(1.41 - 93.03)		
CC			55/18	1.61	(0.86 - 3.00)	.3096	23/0	NA	NA	.0023	.0057
Dom		TT	260/57	1.00			81/1	1.00			
		TC-CC	298/82	1.28	(0.87 - 1.89)	.2103	107/11	8.97	(1.11 - 72.51)	.0078	.0322
Rec		TT-TC	503/121	1.00			165/12	1.00			
		CC	55/18	1.45	(0.81 - 2.61)	.2218	23/0	Na	Na	.0796	.0461
Adi		TT,TC,CC	558/139	1.25	(0.94 - 1.67)	.1291	188/12	1.69	(0.70 - 4.10)	.2469	.5654

SNP	Modelo	Genotipo	CagA+				CagA-				P Interact
			Controles/ casos	OR	(CI 95%) ^a	Valor p	Controles/ casos	OR	(CI 95%) ^a	Valor p	
rs7944723	Cod	CC	326/100	1.00			124/8	1.00			
		CG	203/40	0.64	(0.42 - 0.98)		53/4	1.47	(0.39 - 5.49)		
		GG	35/9	0.81	(0.37 - 1.77)	.1094	10/0	Na	Na	.5225	.3832
	Dom	CC	326/100	1.00			124/8	1.00			
		CG-GG	238/49	0.94	(0.45 - 0.99)	.0416	63/4	1.24	(0.34 - 4.61)	.7449	.5532
	Rec	CC-CG	529/140	1.00			177/12	1.00			
		GG	35/9	0.66	(0.43 - 2.03)	.8702	10/0	Na	Na	.3209	.2807
	Adi	CC,CG,GG	564/149	0.76	(0.55 - 1.05)	.0878	187/12	1.01	(0.33 - 3.06)	.9901	.8574
	rs6421972	Cod	GG	157/53	1.00			63/5	1.00		
GA			277/57	0.60	(0.39 - 0.92)		89/5	0.88	(0.23 - 3.38)		
AA			113/28	0.72	(0.42 - 1.22)	.0674	33/2	1.04	(0.17 - 6.16)	.9751	.9928
Dom		GG	157/53	1.00			63/5	1.00			
		GA-AA	390/85	0.63	(0.42 - 0.94)	.0266	122/7	0.92	(0.26 - 3.23)	.8958	.9123
Rec		GG-GA	434/110	1.00			152/10	1.00			
		AA	113/28	0.97	(0.60 - 1.56)	.8937	33/2	1.11	(0.22 - 5.64)	.8980	.9918
Adi		GG,GA,AA	547/138	0.81	(0.61 - 1.06)	.1235	185/12	0.99	(0.41 - 2.39)	.9817	.9697
rs7126405		Cod	AA	424/117	1.00			146/ 7	1.00		
	AG		96/13	0.50	(0.27 - 0.93)		32/ 5	3.55	(0.99 -12.79)		
	GG		7/2	1.29	(0.24 - 6.81)	.0621	1/0	Na	Na	.1487	.0364
	Dom	AA	424/117	1.00			146/7	1.00			
		AG-GG	103/15	0.54	(0.30 - 0.98)	.0325	33/5	3.34	(0.93 -11.98)	.0741	.0164
	Rec	AA-AG	520/130	1.00			178/12	1.00			
		GG	7/2	1.42	(0.27 - 7.51)	.6893	1/0	Na	Na	.5802	.6455
	Adi	AA,AG,GG	527/132	0.62	(0.36 - 1.06)	.0671	179/12	2.65	(0.81 - 8.62)	.1182	.0325

a, Análisis de regresión logística no condicional ajustado por sexo, edad y país de procedencia. Cod, codominante; Dom, dominante; Rec, recesivo; Adi, aditivo.

Tabla A-23. Haplotipos en genes *TFFs* y asociación con los subtipos anatómicos de cancer gástrico.

Gen	Haplotipo	Frec.	Cardias			No cardias		
			OR	(IC 95%) ^d	Valor p	OR	(IC 95%) ^d	Valor p
<i>TFF1</i>^a	CTGGGACG	0.320	1.00			1.00		
	CCGGGGGG	0.126	0.86	(0.52-1.41)	.5447	0.86	(0.58-1.28)	.4567
	TCAGCGGG	0.091	0.71	(0.38-1.31)	.2690	1.36	(0.71-2.60)	.3539
	CTGGGAGG	0.079	1.19	(0.67-2.12)	.5501	1.13	(0.70-1.82)	.6263
	CCAGCGCG	0.060	1.59	(0.88-2.87)	.1234	1.33	(0.80-2.20)	.2658
	TCACCGCG	0.057	1.07	(0.58-1.99)	.8275	0.86	(0.49-1.51)	.6018
	CCAGCGGA	0.046	1.74	(0.95-3.18)	.0727	1.15	(0.67-1.99)	.6116
	TCAGCGCG	0.032	0.79	(0.29-2.15)	.6439	1.36	(0.71-2.60)	.3539
	CCGGGACG	0.029	1.39	(0.62-3.11)	.4245	0.55	(0.20-1.54)	.2581
	CCGGCGGG	0.022	0.87	(0.28-2.70)	.8115	1.65	(0.79-3.46)	.1824
	CTGCCGCG	0.021	0.76	(0.21-2.78)	.6764	1.07	(0.49-2.33)	.8738
	TCAGGACG	0.018	1.76	(0.70-4.40)	.2269	0.58	(0.19-1.80)	.3493
	CTGGCGGG	0.017	Na	Na	Na	0.56	(0.16-1.91)	.3554
	CCAGCGGG	0.011	0.55	(0.07-4.16)	.5596	1.12	(0.29-4.33)	.8701
	CCGCCGCG	0.011	0.46	(0.06-3.57)	.4607	1.44	(0.52-4.01)	.4826
<i>TFF2</i>^b	AG1TA	0.388	1.00					
	GA1CA	0.182	0.90	(0.61-1.34)	.6151	1.14	(0.83-1.57)	.4159
	AA1TA	0.170	0.78	(0.48-1.26)	.3082	1.00	(0.68-1.47)	.9865
	AA1CA	0.104	0.74	(0.42-1.32)	.3076	0.89	(0.54-1.44)	.6289
	AG1CA	0.080	0.93	(0.52-1.66)	.8150	1.18	(0.70-1.99)	.5226
	GA1TA	0.030	0.59	(0.21-1.69)	.3287	1.50	(0.77-2.92)	.2328
	AG2TG	0.022	0.71	(0.24-2.14)	.5435	0.72	(0.21-2.42)	.5949
	AA2TG	0.022	0.16	(0.01-2.05)	.1594	2.05	(0.96-4.34)	.0622
<i>TFF3</i>^c	AGCGCGG	0.403	1.00			1.00		
	AAGGCGG	0.212	0.93	(0.64-1.35)	.6995	0.97	(0.71-1.32)	.8399
	AAGCGGA	0.166	0.74	(0.47-1.16)	.1878	1.07	(0.78-1.48)	.6648
	AAGGAGG	0.156	0.91	(0.60-1.40)	.6756	1.24	(0.90-1.70)	.1954
	GAGGCGG	0.027	0.85	(0.33-2.20)	.7416	1.17	(0.60-2.29)	.6394
	AGCGCGA	0.015	0.74	(0.18-3.14)	.6861	0.97	(0.31-2.99)	.9522

a, Los haplotipos de *TFF1* incluyen los SNPs en el siguiente orden: rs13047838, rs424694, rs9976977, rs13051704, rs2839488, rs225356; rs225353, rs13276; b, Los haplotipos de *TFF2* incluyen los SNPs en el siguiente orden: rs3814896, rs1079380, c.*67_68insCTT (alelo sin inserción=1, InsCTT=2), rs225334 y rs12483635; c, Los haplotipos de *TFF3* incluyen los SNPs en el siguiente orden: rs8133510, c.*41G>A, rs2236705, V80M, rs225363, rs692356 y rs11203197; d, Análisis de regresión logística ajustado por sexo, edad y país de referencia.

Tabla A-24. Haplotipos en genes *MUC* y asociación con los subtipos histológicos de cancer gástrico.

Gen	Haplotipo ^a	Frec.	Intestinal			Difuso		
			OR	(IC 95%) ^d	valor-p	OR	(IC 95%) ^d	valor-p
<i>MUC1</i> ^a	CAGC	0.463						
	TTAT	0.178	1.31	(0.92-1.86)	.1371	1.34	(0.94-1.91)	.1071
	TTAC	0.164	1.28	(0.88-1.86)	.1920	0.97	(0.64-1.45)	.8727
	CAAT	0.102	0.69	(0.41-1.15)	.1533	1.18	(0.78-1.81)	.4339
	CTAC	0.068	0.78	(0.43-1.43)	.4277	0.88	(0.49-1.57)	.6539
	CAAC	0.011	0.59	(0.09-4.07)	.5961	1.34	(0.32-5.51)	.6863
<i>MUC2 5'</i> ^b	ACCTG	0.313						
	CTCTG	0.307	1.03	(0.72-1.48)	.8551	1.04	(0.50-2.15)	.9189
	CTCCG	0.183	1.33	(0.92-1.92)	.1357	0.85	(0.58-1.26)	.4300
	CTGCG	0.042	0.31	(0.06-1.71)	.1800	0.85	(0.41-1.75)	.6564
	CCGCA	0.038	0.83	(0.38-1.83)	.6505	1.58	(0.89-2.80)	.1161
	CTCTA	0.035	0.27	(0.07-0.99)	.0476	1.04	(0.50-2.15)	.9189
	ACCCG	0.027	1.09	(0.40-2.97)	.8693	0.95	(0.35-2.60)	.9220
	CCGCG	0.026	1.55	(0.78-3.08)	.2155	1.10	(0.50-2.45)	.8116
	ACGCG	0.014	1.79	(0.47-6.75)	.3903	Na	Na	Na
	<i>MUC2 3'</i> ^c	CAGAAAG	0.292					
CGAAGAA		0.168	0.78	(0.52-1.17)	.2284	1.14	(0.76-1.70)	.5191
GGAAAAG		0.154	0.76	(0.50-1.15)	.1880	1.07	(0.71-1.62)	.7442
CAGAGAA		0.078	0.62	(0.34-1.13)	.1208	1.10	(0.64-1.89)	.7393
CAGAGAG		0.072	0.71	(0.39-1.29)	.2588	1.41	(0.84-2.36)	.1948
CGGGGGG		0.067	0.57	(0.31-1.06)	.0744	1.02	(0.59-1.78)	.9386
CGAAGAG		0.048	0.90	(0.48-1.69)	.7504	0.42	(0.16-1.10)	.0768
GGAAGAG		0.022	Na	Na	Na	0.42	(0.10-1.78)	.2412
GGAAAGG		0.017	0.42	(0.10-1.71)	.2240	0.93	(0.34-2.52)	.8859
CAGAGGG		0.016	Na	Na	Na	0.88	(0.31-2.51)	.8051
CGGAGAG		0.014	1.47	(0.51-4.21)	.4721	1.99	(0.69-5.70)	.2010
GGAAGAA		0.014	1.32	(0.48-3.68)	.5911	0.52	(0.07-3.77)	.5169
GGAAAAA		0.010	0.56	(0.12-2.56)	.4554	0.21	(0.00-12.25)	.4519
<i>MUC6</i> ^d	ACCGG	0.428	1.00					
	GCCGG	0.146	1.31	(0.90-1.92)	.1603	1.14	(0.75-1.73)	.5287
	ATCGG	0.132	0.92	(0.57-1.47)	.7178	1.15	(0.74-1.78)	.5284
	GCCGT	0.106	1.09	(0.69-1.72)	.7073	1.48	(0.97-2.28)	.0711
	GCCTT	0.105	0.96	(0.60-1.53)	.8697	1.45	(0.95-2.21)	.0865
	GCTGG	0.082	1.02	(0.59-1.76)	.9503	1.29	(0.77-2.14)	.3337

a, Los haplotipos de *MUC1* incluyen los SNPs en el siguiente orden: rs3814316, rs9426886, rs4072037, rs1045253, rs2839488, rs225356; rs225353, rs13276; b, Los haplotipos de la región 5' *MUC2* incluyen los SNPs en el siguiente orden: rs10902073, rs10794281, rs2856082, rs2071174 y rs11245936 ; c, Los haplotipos la región 3' *MUC2* incluyen los SNPs en el siguiente orden: rs7944723, rs10902089, rs6421972, rs11245954, rs7480563, rs7126405, rs3924453; d, Los haplotipos la región 3' *MUC6* incluyen los SNPs en el siguiente orden: 7481521, rs11604757, rs11246384, rs6597947 y rs9794921; e, Análisis de regresión logística ajustado por sexo, edad y país de referencia.

Apéndice II: Figuras

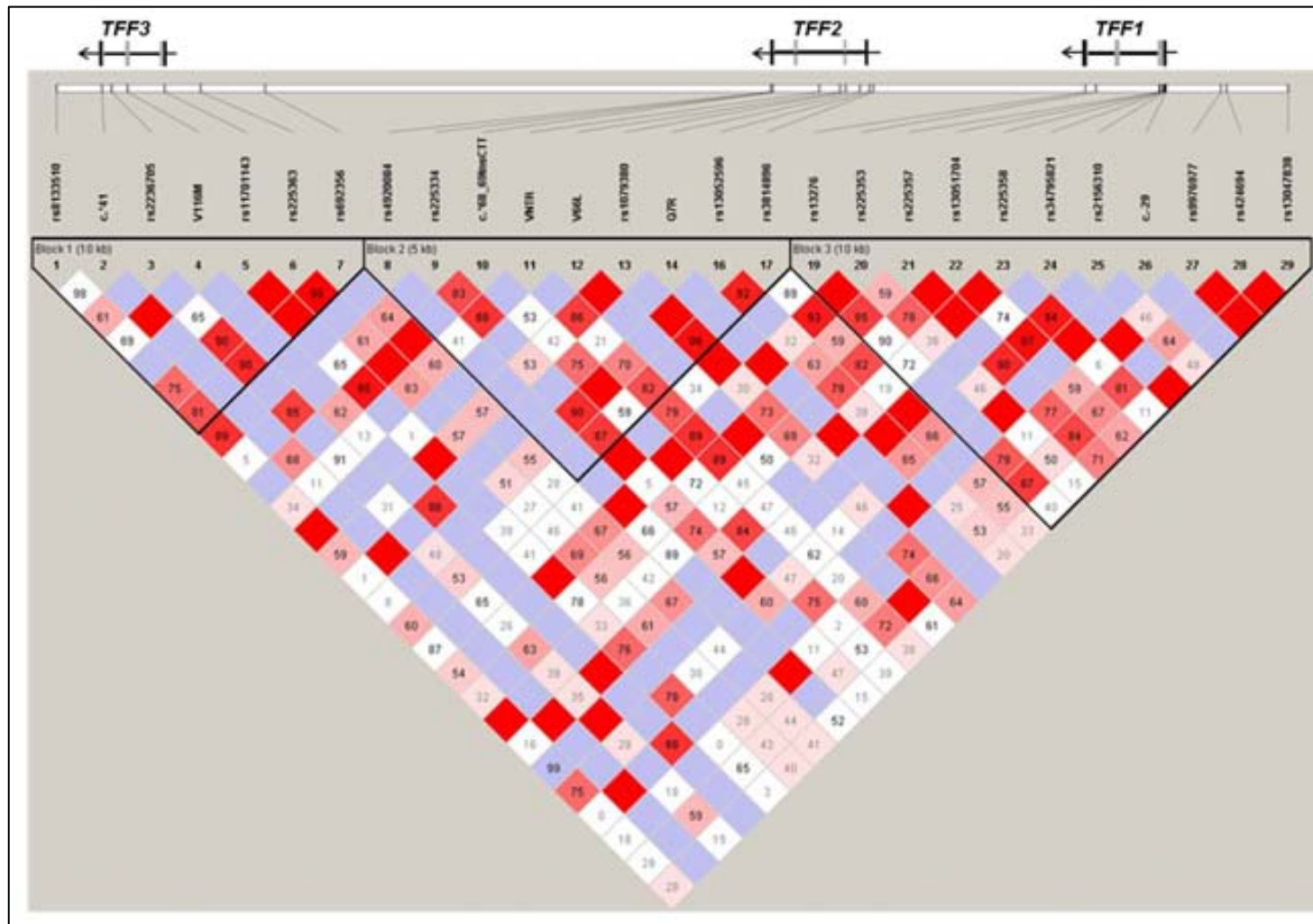


Figura A-1A. Representación gráfica del coeficiente D' de LD en la región 21q22.3 donde se localizan *TFF1*, *TFF2* y *TFF3*, calculada a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico. En la parte superior se localizan e indica la estructura de cada uno de los genes. Los valores D' de 1 no se muestran. Esquema de color: cuadrados en blanco, $D' < 1$ y $\text{LOD} < 2$; en azul, $D' = 1$ y $\text{LOD} < 2$; en tonos rosados, $D' < 1$ y $\text{LOD} \geq 2$ y en rojo fuerte, $D' = 1$ y $\text{LOD} \geq 2$. B. Representación gráfica del coeficiente r^2 . Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

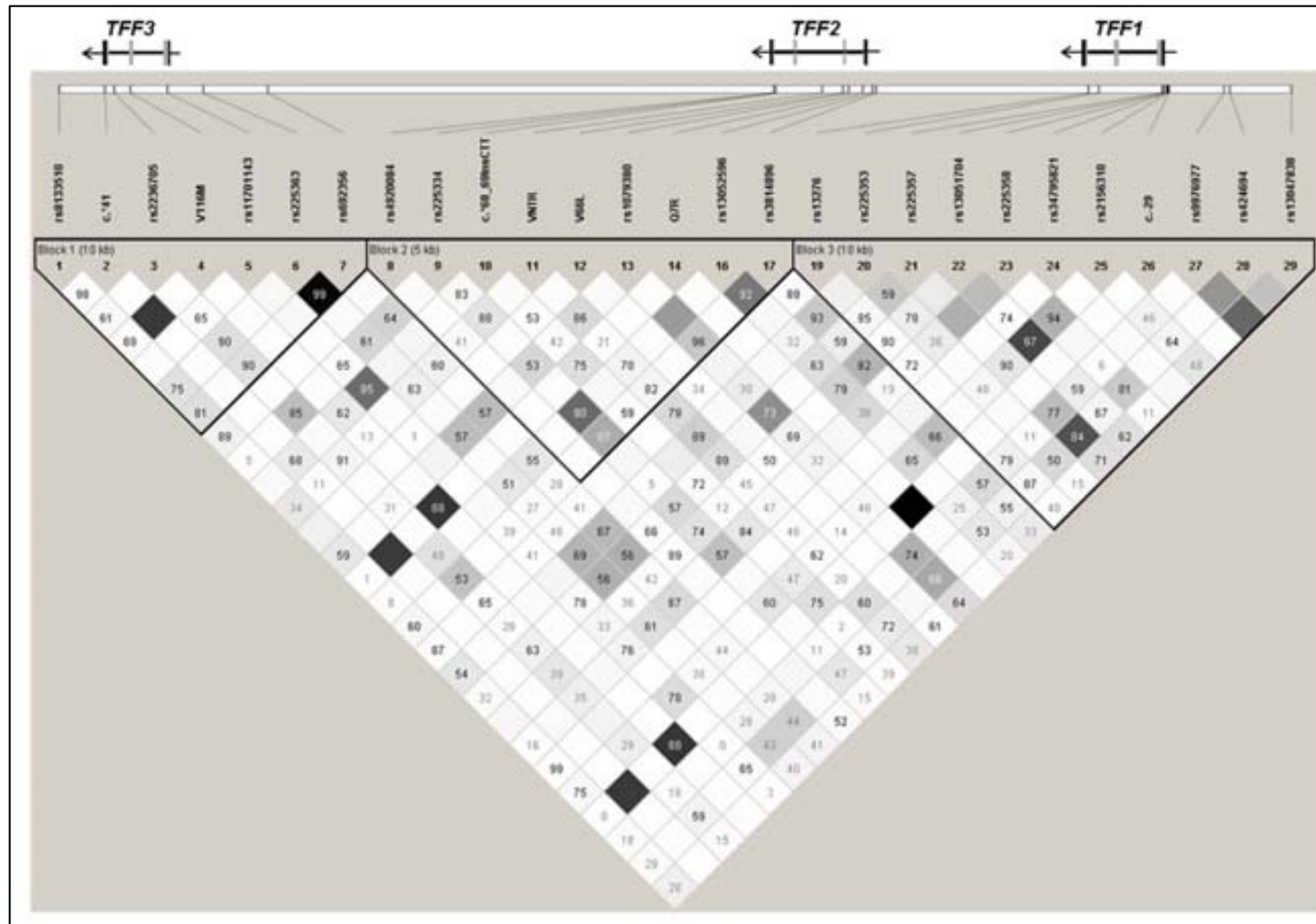


Figura A-1B. Representación gráfica del coeficiente r^2 de LD en la región 21q22.3 donde se localizan *TFF1*, *TFF2* y *TFF3*, calculada a partir de los datos de genotipado en pacientes con lesiones precursoras de cáncer gástrico. En la parte superior se localizan e indica la estructura de cada uno de los genes. Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

rs10794281	rs2856082	rs2071174	rs7396030	rs11245936	rs7944723	rs6421972	rs10794293	rs11245954	rs7480563	rs7126405	rs3924453	rs4077759															
1.0	1.0	0.796	0.609	0.763	0.844	0.710	0.628	1.0	0.755	0.435	0.263	0.060	0.265	0.222	0.724	0.452	0.672	0.481	0.538	0.381	0.025	0.507	0.668	0.489	0.527	D'	rs10902073
0.773	0.683	0.052	0.019	0.149	0.142	0.078	0.039	0.042	0.021	0.076	0.042	0.003	0.018	0.015	0.119	0.009	0.013	0.132	0.127	0.009	0.000	0.065	0.077	0.087	0.070	r ²	rs10794281
	0.334	0.636	0.292	0.158	0.017	0.234	0.267	0.286	0.374	0.277	0.072	0.406	0.280	0.662	0.279	0.285	0.320	0.296	0.211	0.401	0.363	0.396	0.237	0.177	D'	rs10794281	
	0.019	0.115	0.027	0.007	0.000	0.030	0.007	0.017	0.045	0.037	0.003	0.061	0.030	0.125	0.007	0.013	0.073	0.056	0.006	0.037	0.043	0.040	0.026	0.011	r ²	rs2856082	
		0.941	1.0	0.877	1.0	0.407	0.402	0.626	1.0	0.747	0.781	0.593	0.111	0.370	0.334	0.636	0.643	0.388	0.286	0.042	0.213	0.480	0.576	D'	rs2856082		
		0.281	0.258	0.419	0.525	0.089	0.117	0.013	0.024	0.048	0.065	0.023	0.001	0.074	0.061	0.060	0.049	0.110	0.067	0.001	0.013	0.052	0.068	r ²	rs2071174		
			0.322	0.447	0.198	0.456	0.912	0.765	0.245	0.395	0.332	0.479	0.163	0.309	0.794	0.776	0.200	0.248	0.413	0.561	0.584	0.590	D'	rs2071174			
			0.063	0.098	0.007	0.039	0.083	0.065	0.040	0.122	0.094	0.184	0.005	0.014	0.279	0.275	0.010	0.013	0.168	0.273	0.242	0.278	r ²	rs7396030			
				1.0	1.0	0.744	1.0	0.825	0.930	0.728	0.788	0.857	0.808	0.577	0.669	0.772	0.660	0.128	0.135	0.558	0.523	D'	rs7396030				
				0.295	0.380	0.033	0.049	0.108	0.176	0.063	0.071	0.219	0.188	0.089	0.101	0.238	0.186	0.010	0.010	0.130	0.108	r ²	rs11245936				
					1.0	1.0	0.732	0.803	0.587	0.577	0.729	0.760	0.798	0.622	0.717	0.622	0.269	0.535	0.852	0.590	D'	rs11245936					
					0.017	0.021	0.024	0.050	0.012	0.016	0.519	0.439	0.049	0.345	0.383	0.345	0.002	0.015	0.082	0.052	r ²	rs7944723					
						0.986	1.0	0.403	0.632	1.0	1.0	0.664	0.556	0.422	0.243	0.271	0.551	0.446	0.639	D'	rs7944723						
						0.386	0.318	0.081	0.171	0.018	0.018	0.102	0.078	0.004	0.002	0.007	0.035	0.029	0.064	r ²	rs6421972						
							0.887	0.950	1.0	1.0	0.112	0.312	0.721	0.631	0.402	0.509	0.175	0.260	D'	rs6421972							
							0.611	0.674	0.046	0.059	0.008	0.057	0.033	0.035	0.104	0.179	0.028	0.067	r ²	rs10794293							
								1.0	1.0	0.303	0.401	0.688	1.0	0.458	0.498	0.283	0.383	D'	rs10794293								
								0.035	0.042	0.049	0.088	0.023	0.055	0.176	0.208	0.068	0.130	r ²	rs11245954								
									0.865	1.0	0.956	1.0	1.0	1.0	0.900	1.0	D'	rs11245954									
									0.061	0.065	0.703	0.676	0.029	0.040	0.094	0.113	r ²	rs7480563									
										0.716	0.849	0.807	1.0	0.848	1.0	D'	rs7480563										
										0.055	0.069	0.285	0.397	0.449	0.573	r ²	rs7126405										
											1.0	1.0	0.832	0.707	D'	rs7126405											
											0.040	0.059	0.113	0.084	r ²	rs3924453											
												0.992	1.0	D'	rs3924453												
												0.687	0.694	r ²													

Figura A-2. Representación gráfica de los coeficientes D' y r² de LD entre los 14 SNPs localizados en el gen MUC2 y regiones adyacentes, en 11p15. En la primera fila se muestran los valores D' y en la segunda fila, en negrita, los valores r². En estilo normal se muestran los valores de LD obtenidos en la población de pacientes con lesiones gástricas y en cursiva se muestran los valores obtenidos a partir de los datos de HapMap en población CEU.

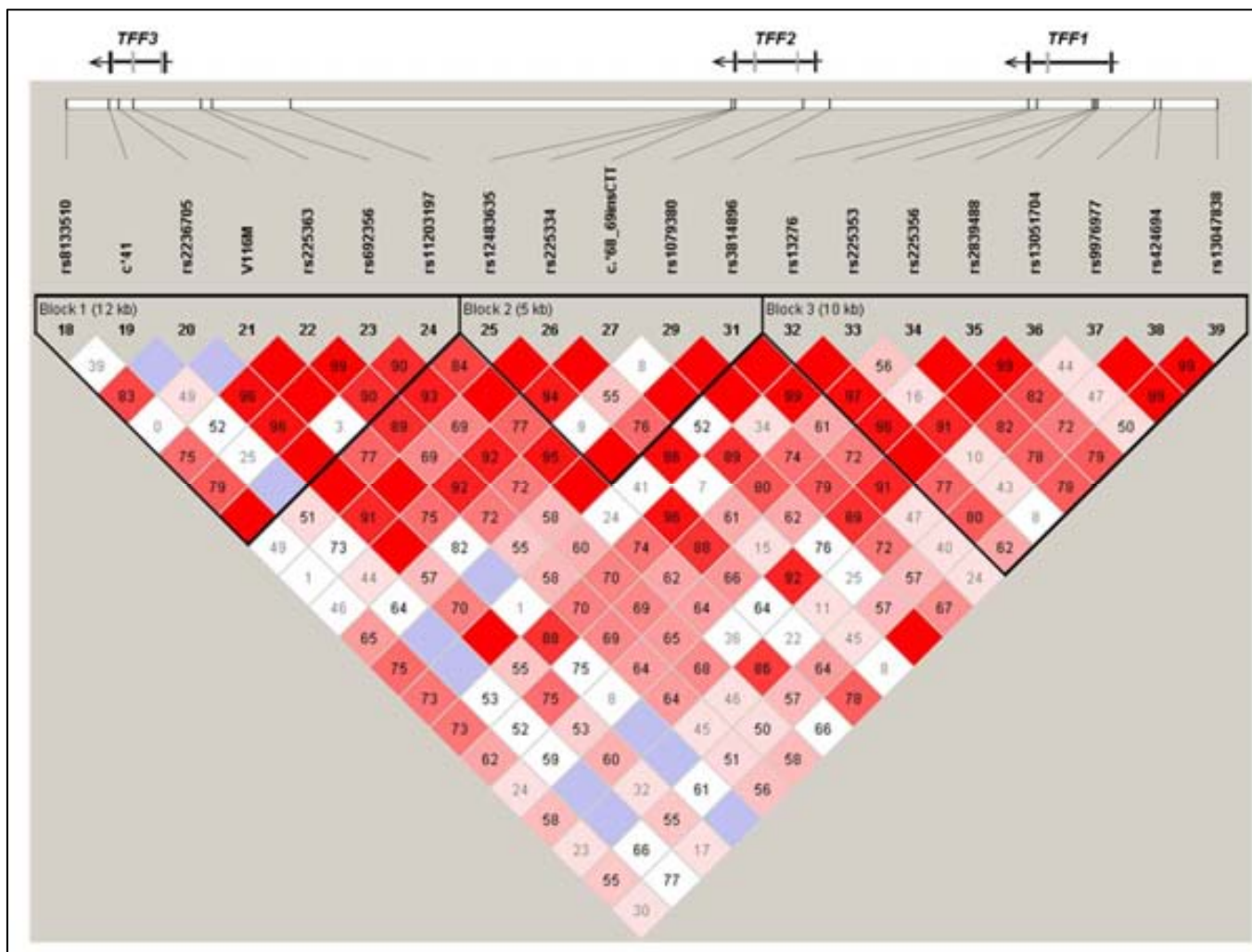


Figura A-3A. Representación gráfica del coeficiente D' de LD en la región 21q22.3 donde se localizan *TFF1*, *TFF2* y *TFF3*, calculada a partir de los datos de genotipado en la población caso-control de cáncer gástrico. En todos los casos se indica la estructura de cada uno de los genes. Los valores D' de 1 no se muestran. Esquema de color: cuadrados en blanco, $D' < 1$ y $\text{LOD} < 2$; en azul, $D' = 1$ y $\text{LOD} < 2$; en tonos rosa y rojo, $D' < 1$ y $\text{LOD} \geq 2$ y en rojo fuerte, $D' = 1$ y $\text{LOD} \geq 2$.

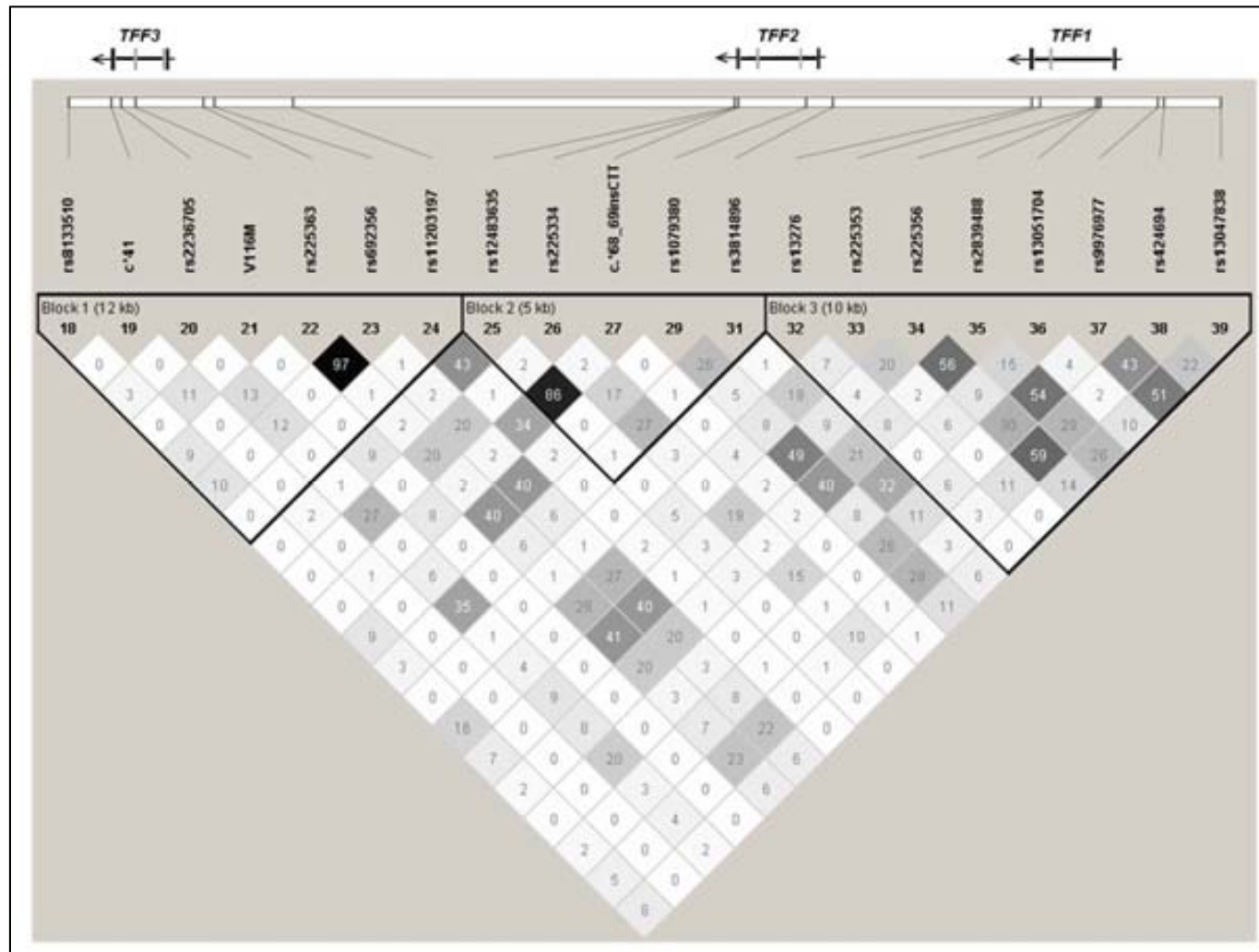


Figura A-3B. Representación gráfica del coeficiente r^2 de LD en la región 21q22.3 donde se localizan *TFF1*, *TFF2* y *TFF3*, calculada a partir de los datos de genotipado en la población caso-control de cáncer gástrico. En todos los casos se indica la estructura de cada uno de los genes. Esquema de color: en blanco, $r^2 = 0$; en tonos grises, $0 < r^2 < 1$ y en negro, $r^2 = 1$.

rs10794281	rs2856082	rs2071174	rs11245936	rs7944723	rs10902089	rs6421972	rs11245954	rs7480563	rs7126405	rs3924453													
1.0	1.0	0.626	0.609	0.671	0.844	0.911	0.755	0.291	0.263	0.129	0.224	0.048	0.265	0.566	0.672	0.430	0.538	0.327	0.025	0.525	0.668	D'	rs10902073
0.750	0.683	0.031	0.019	0.128	0.142	0.042	0.021	0.043	0.042	0.011	0.021	0.001	0.018	0.014	0.013	0.107	0.127	0.007	0.000	0.059	0.077	r ²	
	0.372	0.636	0.247	0.158	0.134	0.286	0.209	0.277	0.085	0.116	0.177	0.406	0.225	0.285	0.262	0.296	0.151	0.401	0.457	0.396		D'	rs10794281
	0.026	0.115	0.023	0.007	0.002	0.017	0.017	0.037	0.006	0.008	0.019	0.061	0.005	0.013	0.053	0.056	0.004	0.037	0.059	0.040		r ²	
		0.980	1.0	0.431	0.402	0.720	1.0	0.096	0.094	0.766	0.781	0.370	0.334	0.622	0.643	0.325	0.286	0.001	0.213		D'	rs2856082	
		0.269	0.258	0.115	0.117	0.021	0.024	0.001	0.001	0.066	0.065	0.075	0.061	0.053	0.049	0.086	0.067	0.000	0.013		r ²		
				0.380	0.456	0.685	0.765	0.511	0.598	0.330	0.395	0.296	0.309	0.715	0.776	0.164	0.248	0.426	0.561		D'	rs2071174	
				0.026	0.039	0.068	0.065	0.116	0.160	0.067	0.122	0.014	0.014	0.248	0.275	0.006	0.013	0.137	0.273		r ²		
						0.877	1.0	0.702	0.774	0.887	0.803	0.674	0.760	0.837	0.622	0.585	0.622	0.534	0.535		D'	rs11245936	
						0.020	0.021	0.039	0.047	0.056	0.050	0.396	0.439	0.062	0.345	0.258	0.345	0.010	0.015		r ²		
								0.986	1.0	0.986	1.0	1.0	1.0	0.659	0.556	0.168	0.243	0.607	0.551		D'	rs7944723	
								0.244	0.242	0.347	0.318	0.023	0.018	0.129	0.078	0.001	0.002	0.040	0.035		r ²		
										0.995	1.0	0.932	1.0	0.246	0.412	0.600	0.817	0.330	0.529		D'	rs10902089	
										0.700	0.715	0.061	0.081	0.056	0.164	0.037	0.072	0.037	0.125		r ²		
												0.962	1.0	0.085	0.312	0.599	0.631	0.435	0.509		D'	rs6421972	
												0.058	0.059	0.006	0.057	0.033	0.035	0.090	0.179		r ²		
														1.0	1.0	1.0	1.0	1.0	1.0		D'	rs11245954	
														0.076	0.065	0.645	0.676	0.031	0.040		r ²		
																0.582	0.849	0.891	1.0		D'	rs7480563	
																0.038	0.069	0.293	0.397		r ²		
																		1.0	1.0		D'	rs7126405	
																		0.045	0.059		r ²		

Figura A-4. Representación gráfica de los coeficientes D' y r² de LD entre los 12 SNPs localizados en el gen *MUC2* y regiones adyacentes, en 11p15. En la primera fila se muestran los valores D' y en la segunda fila, en negrita, los valores r². En estilo normal se muestran los valores de LD obtenidos en la población caso-control de cáncer gástrico y en cursiva se muestran los valores obtenidos a partir de los datos de HapMap en población CEU.

Apéndice IV: Publicaciones

En este apéndice se incluye la publicación obtenida como consecuencia directa del trabajo realizado en esta tesis:

“Genetic variation in *MUC1*, *MUC2* and *MUC6* genes and evolution of gastric cancer precursor lesions in a long-term follow-up in a high-risk area in Spain”

Autores: **Fátima Marín**, Catalina Bonet, Xavier Muñoz, Nadia García, María Luisa Pardo, Juan Maria Ruiz-Liso, Pablo Alonso, Gabriel Capella, José Miguel Sanz-Anquela, Carlos A. González y Núria Sala.

Carcinogenesis vol.33 no.5 pp.1072-1080, 2012

Además, se incluye un artículo metodológico donde se describe la puesta a punto de la técnica *High Resolution Melting* para el análisis de los polimorfismos nulos de *GSTT1* y *GSTM1*. Esta técnica, desarrollada por la doctoranda durante el periodo de ejecución de esta tesis, se utilizó en la misma para el genotipado de un SNP y un polimorfismo de inserción y para la búsqueda de variantes polimórficas en *MUC1*

“Simultaneous genotyping of *GSTT1* and *GSTM1* null polymorphisms by melting curve analysis in presence of SYBR Green I”

Autores: **Fátima Marín**, Nadia García, Xavier Muñoz, Gabriel Capellà, Carlos A. González, Antonio Agudo y Núria Sala.

Journal of Molecular Diagnostics, vol.12, no. 3, pp.300-304, 2010

Finalmente, se ha incluido la publicación resultado de la colaboración con otros miembros del grupo de investigación en el que se ha realizado esta tesis, donde se aplicó la técnica anteriormente mencionada.

“Aromatic DNA adducts and polymorphisms in metabolic genes in healthy adults: findings from the EPIC-Spain cohort”

Autores: Antonio Agudo, Marco Peluso, Núria Sala, Gabriel Capellà, Armelle Munnia, Sara Piro, **Fátima Marín**, Raquel Ibáñez¹, Pilar Amiano, M.José Tormo, Eva Ardanaz, Aurelio Barricarte, M.Dolores Chirlaque, Miren Dorransoro, Nerea Larrañaga, Carmen Martínez, Carmen Navarro, J.Ramón Quirós, M.José Sánchez y Carlos A.González

Carcinogenesis vol.30 no.6 pp.968-976, 2009

Genetic variation in *MUC1*, *MUC2* and *MUC6* genes and evolution of gastric cancer precursor lesions in a long-term follow-up in a high-risk area in Spain

Fátima Marín^{1,2}, Catalina Bonet², Xavier Muñoz^{1,2},
Nadia García^{1,2}, María Luisa Pardo³, Juan Maria Ruiz-Liso³,
Pablo Alonso⁴, Gabriel Capellà¹, José Miguel Sanz-Anquela⁵,
Carlos A. González² and Núria Sala^{1,2,*}

¹Translational Research Laboratory, Institut Català d'Oncologia (IDIBELL-ICO), Gran Via, km 2.7 s/n, 08907 L'Hospitalet de Llobregat, Spain, ²Unit of Nutrition, Environment and Cancer, Cancer Epidemiology program, Institut Català d'Oncologia (IDIBELL-ICO), L'Hospitalet de Llobregat, Spain,

³Pathology Department, Complejo Hospitalario de Soria, Spain,

⁴Gastroenterology Unit, Complejo Hospitalario de Soria, Spain and

⁵Department of Medical Specialities, College of Medicine, Alcala University, Hospital Principe de Asturias, Spain

*To whom correspondence should be addressed. Tel: +34 932607464;

Fax: +34 932607466;

Email: nsala@iconcologia.net

In order to assess whether inherited genetic variability in the mucin genes associates with the evolution of gastric cancer precursor lesions (GCPLs), we genotyped 22 tagSNPs in *MUC1*, *MUC6* and *MUC2* genes of 387 patients with GCPLs that had been followed up for 12.8 years. According to the diagnosis at recruitment and at the end of follow-up, the lesions did not change in 43.1% of the patients, regressed in 28.7% and progressed in 28.2%. Three SNPs in the 3'-moiety of *MUC2* were significantly associated with a decreased risk of progression of the lesions, whereas another four SNPs, located at the 5'-moiety, were found to be significantly associated either with increased [one single-nucleotide polymorphism (SNP)] or decreased (three SNPs) probability of regression. Stratified analysis indicated that significance was maintained only in those subjects positive for *Helicobacter pylori* infection and in those not consuming non-steroidal anti-inflammatory drugs, which were found protective against lesion progression. Haplotype analyses indicated the presence of two haplotypes, one in each moiety of the gene, that were significantly associated with decreased risk of progression of the lesions [odds ratio (OR) = 0.49 and 0.46; 95% confidence interval (CI) = 0.28–0.85 and 0.25–0.86, respectively]. The 5'-end haplotype was also associated with increased probability of regression (OR = 1.67; 95% CI = 1.02–2.73), altogether suggesting a protective role against progression of the precancerous lesions. No significant association was found with variants in *MUC1* and *MUC6* genes. These results indicate, for the first time, that genetic variability in *MUC2* is associated with evolution of GCPLs, especially in *H.pylori* infected patients, suggesting a role of this secreted mucin in gastric carcinogenesis.

Introduction

Mucins are high-molecular weight glycoproteins (50–80% O-linked oligosaccharides) synthesized by specialized epithelial cells as membrane-bound or secretory proteins that form higher-order structures (1,2). These glycoproteins represent the major constituents of the protective mucous layer of the gastrointestinal tract where they have a central role in maintaining epithelial homeostasis (3,4).

Normal gastric mucosa shows cell type-specific expression of secreted mucins *MUC6*, in glandular structures, and *MUC5AC*, in the surface epithelium. High levels of membrane-bound *MUC1* have

Abbreviations: CI, confidence interval; GCPL, gastric cancer precursor lesion; IM, intestinal metaplasia; LD, linkage disequilibrium; NAG, non-atrophic gastritis; NSAID, non-steroidal anti-inflammatory drug; MAG, multifocal atrophic gastritis; OR, odds ratio; SNP, single-nucleotide polymorphism; TR, tandem repeat.

also been detected in both deep glands and superficial epithelium. This specific pattern of expression is altered in the process of neoplastic transformation of the stomach, with decreased expression of *MUC6*, overexpression of *MUC1* and the novo expression of secreted *MUC2* (5–7). Alteration in the expression pattern of mucins may modify their epithelial functions, which in turn may influence the tumorigenicity and metastatic properties of cancer cells (8,9).

Gastric cancer is the result of a long multistep and multifactorial process that involves well-characterized sequential stages. The initial lesion is an inflammatory process, usually associated with *Helicobacter pylori* infection, that results in a chronic superficial gastritis, which progresses toward atrophic gastritis, intestinal metaplasia and, finally, to dysplasia and adenocarcinoma (10,11). Although *H.pylori* is the most important established risk factor for gastric cancer development, other environmental and lifestyle factors, as well as host genetic susceptibility, are involved in the cancer process (12).

MUC2 and *MUC6* genes are located at 11p15.5, in a recombination-rich region containing a 400 kb gene cluster which also contains two other mucin genes (*MUC5AC* and *MUC5B*). Their 5' region is highly similar and is clearly differentiated from the 3' region by a large central tandem repeat (TR) domain, which is in exon 30 in *MUC2* and in exon 31 in *MUC6*. In contrast, the C-terminal domain is encoded by 19 exons in *MUC2* and only 2 exons in *MUC6* (13). *MUC1* localizes in 1q21 and it is a smaller gene with its TR region in exon 2. Some studies have addressed the role of genetic polymorphisms in mucosal protection genes but most of them have been focused on the serine and threonine-rich VNTR (variable number of TRs), characteristic of the *MUC* genes (14–20). More recently, a functional single-nucleotide polymorphism (SNP) (rs4072037) in *MUC1* has been strongly associated with gastric cancer in Asian populations (21–23). However, none of these studies have assessed how these inherited genetic variants predispose to progression or regression of premalignant lesions of the stomach, although MUCs have been long implicated in the pathogenesis of gastric carcinogenesis (4,9).

In the present study, we have used comprehensive genetic markers that capture the most common variability at each locus to analyze the constitutional genetic variability of *MUC1* (gene ID: 4582), *MUC2* (gene ID: 4583) and *MUC6* (gene ID: 4588) genes and to assess its relationship with the evolution of gastric cancer precursor lesions (GCPLs) after a long-term follow-up in a Spanish region with one of the highest incidence and mortality from gastric cancer.

Methods

Study population

The study population for association with the evolution of GCPLs consisted of 478 subjects from a follow-up study described in detail previously (24). Briefly, patients with GCPLs were identified from the Pathology Department Registry of the Soria Hospital as having undergone gastroscopy with gastric biopsy between 1988 and 1994. Patients were followed up until 2005–2007 (12.8 years of mean follow-up) when a new gastroscopy with biopsy was performed. Genomic DNA, either from blood ($n = 417$) or from paraffin-embedded tissue blocks when blood sampling was not possible ($n = 36$), was obtained from 453 (95%) patients. Sixty-six samples had to be excluded from this study because their DNA did not fulfill the minimum criteria required for genotyping in the Sequenom platform. The specific study population for genetic analysis was therefore of 387 patients with GCPLs (Table I).

Gastric lesions were classified, at recruitment and at the end of follow-up, as: 1 = normal, 2 = non-atrophic gastritis, 3 = non-metaplastic multifocal atrophic gastritis, 4 = complete intestinal metaplasia (IM, including predominant complete IM), 5 = incomplete IM (including predominant incomplete IM), 6 = dysplasia and 7 = gastric cancer. Based on the diagnosis at recruitment and at the end of follow-up, it was considered that the lesions had progressed or regressed if they had respectively advanced or regressed at least one point of in the overall score (1 to 7). The lesions were considered to be stable if they maintained the same score.

Table I. Main characteristics of the analyzed population

Characteristics	Total		No change		Regression		Progression		P value ^a
	n	%	n	%	n	%	n	%	
Sex									
Men	178	46.0	74	44.3	46	41.4	58	53.2	
Women	209	54.0	93	55.7	65	58.6	51	46.8	0.182
Age									
<40	113	29.2	46	27.5	36	32.4	31	28.4	
40–50	85	22.0	37	22.2	26	23.4	22	20.2	
50–60	122	31.5	53	31.7	38	34.2	31	28.4	
>60	67	17.3	31	18.6	11	9.9	25	22.9	0.312
Use of NSAID									
No	196	51.7	76	46.3	55	49.5	65	62.5	
Yes	183	48.3	88	53.7	56	50.5	39	37.5	0.031
Family history									
No	317	84.5	145	87.9	89	82.4	83	81.4	
Yes	58	15.5	20	12.1	19	17.6	19	18.6	0.277
Smoking status									
Never	219	57.0	97	58.4	64	58.2	58	53.7	
Current	103	26.8	41	24.7	32	29.1	30	27.8	
Ex-smoker	62	16.2	28	16.9	14	12.7	20	18.5	0.727
<i>Helicobacter pylori</i> infection									
No	88	23.0	42	25.3	19	17.1	27	25.5	
Yes	295	77.0	124	74.7	92	82.9	79	74.5	0.219
Type of lesion ^b									
Normal	7	1.8	3	1.8	0	0.0	4	3.7	
NAG	44	11.4	20	12.0	2	1.8	22	20.2	
MAG	190	49.1	84	50.3	60	54.1	46	42.2	
Complete IM	68	17.6	31	18.6	26	23.4	11	10.1	
Incomplete IM	78	20.2	29	17.4	23	20.7	26	23.9	0.00028
Total	387	100	167	100	111	100	109	100	

MAG, multifocal atrophic gastritis; NAG, non-atrophic gastritis.

^aP value of the association between each variable and evolution status based on chi square.

^bType of lesion diagnosed at recruitment.

DNA extraction

Genomic DNA from blood samples was extracted from lymphocytes by use of the Flexigene DNA kit (Qiagen), according to the manufacturer's instructions. Paraffin-embedded tissue blocks were cut into 10 sections of 10 µm and deparaffinized in xylene. Genomic DNA extraction from the deparaffinized tissue was carried out using the QIAamp DNA Mini Kit (Qiagen). DNA concentrations were measured by optical density at 260 nm on a NanoDrop ND-1000 (Thermo Scientific) spectrophotometer and stored at –80°C in 96-well plates at a concentration ~100 ng/µl.

SNP selection criteria

We selected tagSNPs from the 1q21 genomic region containing *MUC1* gene and from the mucin cluster in 11p15, which contains *MUC2* and *MUC6* genes. We used HapMap information for Caucasians [phase II CEU population, release 24, where CEU refers to CEPH (Utah residents with ancestry from northern and western Europe) population] to compile a list of SNPs from a minimum of 7 kb upstream to 3 kb downstream of each of the *MUC* genes. Unfortunately, due to absence of HapMap information for most of the *MUC5AC* gene, which also locates in the same cluster region, this gene could not be analyzed. Haplotype blocks were defined by use of the confidence intervals (CIs) method of Gabriel *et al.* (25), as implemented in Haploview v4.0 software (26). Tagging SNPs were selected by use of the Tagger algorithm as implemented in Haploview. Parameters used for tagSNPs selection were minor allele frequency ≥5% in Caucasians, $r^2 \geq 0.8$ between each pair of tagged and tagging SNPs (pairwise tagging) and SNPs tagging haplotypes with a frequency ≥0.05. Validated potentially functional SNPs replaced tagger selected tagSNPs only if they were in complete linkage disequilibrium (LD) with them.

Following these criteria, we defined two haplotype blocks spanning 51 kb in the 1q21 genomic region containing *MUC1* (4.4 kb) and part of *TRIM46* gene (at 855 bp 3'-downstream of *MUC1*) and we selected four tagSNPs, two for each block. In chromosome region 11p15, we defined a 29 kb haplotype block that included *MUC6* (23.9 kb) and was tagged by six tagSNPs. Also in 11p15, in the 47 kb genomic region that includes *MUC2* (29.5 kb), we defined seven haplotype blocks and 15 tagSNPs (Figure 1). Table II shows the main characteristics and location of all selected SNPs.

SNP genotyping and data filtering

The set of 25 tagSNPs selected in *MUC1*, *MUC2* and *MUC6* genes (Table II) were genotyped using the MassArray™ iPLEX-Gold technology (Sequenom) at the National Genotyping Center of Spain (CEGEN), in a customized assay design. In addition to the internal genotyping controls included by CEGEN, 17 samples (4.4%) chosen randomly were genotyped in duplicate for all tagSNPs to assess the genotyping error rate. Concordance of their genotypes was complete in all cases and no genotyping errors were observed. From the 25 SNPs initially selected, 3 were excluded (Table II) because they could not be amplified or their genotyping signal or cluster separation was not good enough. The total number of SNPs successfully genotyped and available for analysis was therefore 22. None of these SNPs deviated from Hardy–Weinberg equilibrium in the total analyzed population (chi-square test $P > 10^{-3}$; Table II). The average call rate of the analyzed SNPs was 98% and none of them had a call rate <80%.

Helicobacter pylori infection diagnosis

Helicobacter pylori infection was detected by histopathological examination of Giemsa-stained biopsies. Infection status was also identified by reviewing the medical records (results from rapid urease test on biopsy specimen and urea breath test) and interviewing subjects about previous prevalence of *H.pylori*. Finally, DNA was isolated from paraffin-embedded tissue blocks from available gastric biopsies ($n = 312$) and *H.pylori* detection and strain characterization was carried out as described previously (27). A positive result for any of these tests was taken as a positive *H.pylori* infection in the analyzed patient.

Association analyses

The Pearson χ^2 test was used to examine differences among the GCPLs evolution groups according to the main baseline characteristics of the population. In order to capture all the variability associated with the evolution of lesions, two type of analyses were performed (i) in the analysis of lesion progression, the group of patients whose lesions progressed were compared with those whose lesions remained stable or regressed (reference group); (ii) in the analysis of lesion regression, those patients whose lesions regressed were compared with those whose lesions remained stable or progressed. Odds ratios (OR) and 95% CIs were estimated using unconditional logistic regression, as

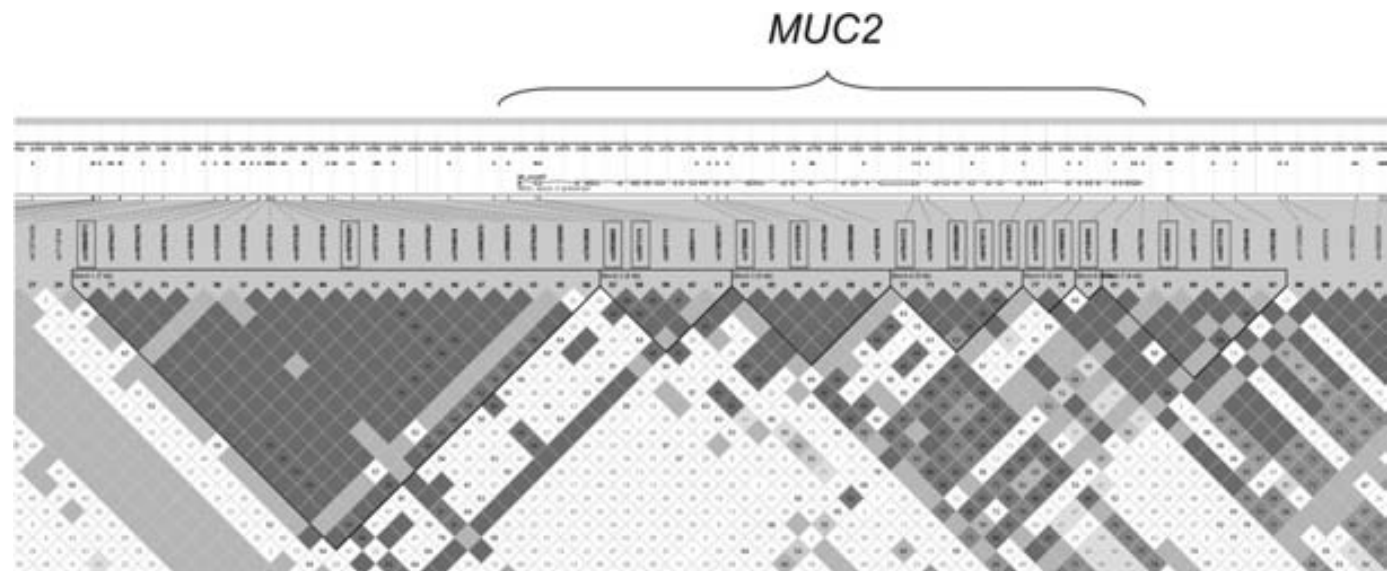


Fig. 1. LD structure in the 11p15.5 genomic region around *MUC2*, according to Haploview analysis of genotype data downloaded from HapMap (phase II CEU population, release 24 based on dbSNP version 126 and NCBI genome build 36). The LD map is shown as a confidence bounds scheme using Haploview, where black boxes indicate strong evidence of LD while white ones indicate strong evidence of recombination and the grey ones showing intermediate values. The 15 tagSNPs selected for study are indicated in square boxes and their numbering, from left (5') to right (3'), is the same as listed in Table 2, for *MUC2* SNPs, from top to bottom. The position and exon–intron structure of *MUC2*, in this region, is also indicated.

Table II. Selected SNPs in *MUC* genomic regions and genotyping results

Polymorphism	Gene	Allele ^a	Location ^b	Chromosome	Position ^c	Total		No change		Progression		Regression		HWE
						MAF	N	MAF	N	MAF	N	MAF	N	
rs3814316	<i>TRIM46</i>	C/T	T287T	1	153416342	0.32	383	0.31	164	0.35	108	0.31	111	0.1287
rs9426886	<i>TRIM46</i>	A/T	Intronic	1	153418378	0.42	387	0.42	167	0.45	109	0.39	111	0.2117
rs4072037 ^d	<i>MUC1</i>	A/G	T31T	1	153428691	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs1045253	<i>MUC1</i>	C/T	5'-upstream	1	153467859	0.31	372	0.29	161	0.33	104	0.33	107	0.7168
rs4076950	<i>MUC6</i>	C/T	Intronic	11	1005021	0.38	380	0.40	164	0.36	107	0.37	109	0.5150
rs7481521	<i>MUC6</i>	A/G	V615M	11	1017811	0.37	354	0.36	156	0.40	101	0.35	97	0.1391
rs11604757 ^d	<i>MUC6</i>	C/T	Intronic	11	1019320	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs11246384	<i>MUC6</i>	C/T	Intronic	11	1020448	0.09	385	0.10	165	0.07	109	0.09	111	0.3418
rs6597947	<i>MUC6</i>	G/T	5'-upstream	11	1027029	0.07	379	0.05	163	0.09	106	0.07	110	0.0194
rs9794921	<i>MUC6</i>	G/T	5'-upstream	11	1029867	0.13	366	0.11	158	0.16	102	0.15	106	0.2527
rs10902073	<i>MUC2</i>	C/A	5'-upstream	11	1050934	0.38	380	0.39	164	0.43	107	0.32	109	0.5150
rs10794281	<i>MUC2</i>	T/C	5'-upstream	11	1053149	0.44	359	0.45	157	0.48	99	0.38	103	0.5221
rs2856082	<i>MUC2</i>	C/G	5'-upstream	11	1061562	0.12	383	0.11	164	0.13	108	0.11	111	0.6278
rs2071174	<i>MUC2</i>	T/C	5'-upstream	11	1063712	0.29	383	0.29	165	0.25	108	0.34	110	0.5365
rs7396030	<i>MUC2</i>	C/T	Intronic	11	1073364	0.20	385	0.21	166	0.19	108	0.18	111	0.4211
rs11245936	<i>MUC2</i>	G/A	G832S	11	1074362	0.06	374	0.07	161	0.06	105	0.07	108	0.1846
rs7944723	<i>MUC2</i>	C/G	P1832P	11	1083710	0.19	375	0.21	166	0.19	106	0.17	103	0.8689
rs10902089 ^d	<i>MUC2</i>	A/G	Intronic	11	1084357	NA	NA	NA	NA	NA	NA	NA	NA	NA
rs6421972	<i>MUC2</i>	G/A	I2154T	11	1086494	0.39	381	0.40	166	0.34	105	0.40	110	1
rs10794293	<i>MUC2</i>	C/T	Intronic	11	1088939	0.33	387	0.34	167	0.27	109	0.36	111	0.9081
rs11245954	<i>MUC2</i>	A/G	V2459V	11	1091078	0.07	382	0.08	165	0.06	108	0.05	109	1
rs7480563	<i>MUC2</i>	A/G	T2524P	11	1091649	0.48	367	0.50	158	0.43	102	0.50	107	0.0756
rs7126405	<i>MUC2</i>	A/G	Q2653P	11	1093296	0.09	387	0.10	167	0.08	109	0.08	111	1
rs3924453	<i>MUC2</i>	G/A	3'-downstream	11	1095806	0.29	387	0.31	167	0.25	109	0.30	111	0.7110
rs4077759	<i>MUC2</i>	T/C	3'-downstream	11	1095976	0.37	379	0.40	162	0.31	106	0.38	111	0.5810

HWE, Hardy–Weinberg equilibrium; MAF, minor allele frequency; N, individuals tested.

^aThe more common allele in Caucasians is reported first.

^bSNP location relative to each gene in the region.

^cChromosome position corresponding to Genome Build 36.3.

^dSelected tagSNPs that failed in the genotyping platform.

implemented in SNPpass R library software (28). Final models included sex, age, non-steroidal anti-inflammatory drugs (NSAID) consumption and type of lesion at recruitment, as covariables for adjustment. All analyses were done under codominant, dominant, recessive and log-additive (per allele) inheritance models. Stratified analyses were carried out according to *H.pylori* infection and NSAID consumption. Potential interactions between the SNP

of interest and *H.pylori* status or NSAID consumption were assessed by means of the likelihood ratio test.

LD between polymorphisms in each *MUC* gene region was explored with the Haploview software and haplotype frequencies were inferred using the expectation-maximization algorithm as implemented in the haplo.stats R package. Association between progression or regression and each *MUC* haplotype

with a frequency >0.01 was measured using the most frequent haplotype among controls as reference.

Results

Evolution of the lesions was assessed based on the diagnosis at recruitment and at the end of the follow-up. We identified 167 patients (43.1%) whose lesions did not change, 111 (28.7%) regressed and 109 (28.2%) progressed. The main characteristics of the analyzed population ($n = 387$) are shown in Table I. The use of NSAIDs was protective against lesion progression (OR = 0.57, 95% CI: 0.34–0.87, $P = 0.01$).

MUC genes genotype frequencies and association with evolution of the lesions

The total number of SNPs successfully genotyped and available for analysis was 22. Of these, 3 localized in *MUC1*, 5 in *MUC6* and 14 in *MUC2* genomic regions. Table II shows their minor allele frequencies in the total population and in each evolution group.

No significant association was observed between evolution of the GCPLs and the genotyped variants in *MUC1* and *MUC6* genes (data not shown). On the other hand, as shown in Table III, three SNPs in *MUC2*, rs10794293 in intron 37 and rs3924453 and rs4077759 in 3'-downstream from the gene, associated with a decreased risk of progression of the lesions. The A allele of rs3924453 and the C allele of rs4077759 appeared as protective factors for progression of the lesions and this effect was significant under the log-additive and recessive inheritance models, as well as the codominant model for rs4077759. Both SNPs tagged a 4 kb haplotype block containing the 3' end and downstream region of *MUC2* (Figure 1). The T allele of rs10794293 in intron 37 was also associated with a decreased risk of progression under the log-additive and dominant inheritance models. This SNP tags a haplotype block also downstream from the central TR region of the gene.

As shown in Table IV, another four *MUC2* SNPs located at the central and 5' region of the gene (rs10902073, rs10794281, rs2071174 and rs7944723) were found to be significantly associated with regression of the lesions. SNPs rs10902073 and rs10794281, which tag the main *MUC2* 7 kb haplotype block at 5'-upstream of the gene, were associated with a reduced probability of regression under the log-additive and dominant models. A significant association with less regression was also observed for rs7944723, a synonymous coding SNP in central exon 30, under the recessive model. On the other hand, the C allele of rs2071174, located at 5'-upstream of *MUC2* and tagging the haplotype block that contains the promoter region and the first exons of the gene, was associated with an increased probability of regression of the lesions, under the codominant and recessive models. Altogether, our results point to a protective effect of SNPs at *MUC2* during evolution of GCPLs.

Effect modification by Helicobacter pylori

When the effect of those *MUC2* SNPs significantly associated with the evolution of the GCPLs was analyzed according to *H.pylori* infection, it was seen that in all cases the association was significant only in those subjects positive for the infection (Tables III and IV), being the interaction statistically significant for rs4077759, under the codominant and recessive models (Table III), as well as for rs10902073 and rs10794281, under the log-additive, dominant and codominant models (Table IV).

Effect modification by consumption of NSAIDs

When the effect of those *MUC2* SNPs significantly associated with the evolution of the GCPLs was analyzed according to consumption of NSAIDs, it was seen that all SNPs but rs10902073 and rs7944723 were significantly associated with the evolution of the lesions in non-NSAID users, but not in users, and that in all cases, the association was in agreement with the one observed with the non-stratified

analyses (data not shown), but no interaction was observed between these *MUC2* variants and NSAIDs consumption.

Mucin haplotypes and evolution of lesions

Measures of LD (r^2 and D') between the analyzed SNPs in the mucin genes were in good agreement with those measured in HapMap for Caucasians (data not shown). In the same way, haplotype analysis showed similar results to those observed in HapMap from Caucasians.

Regarding *MUC2* gene, haplotype analysis was performed separately in the 5' and 3' regions. Six SNPs were located from central exon 30 to the 5' region and tagged 10 different haplotypes with frequencies between 0.345 and 0.014. Eight SNPs were located in 3'-half region of the gene and these tagged another 10 haplotypes with frequencies between 0.322 and 0.018 (Table V). The haplotype CTCCCG, at the 5' region, was significantly associated with a decreased risk of progression of the lesions as well as an increased probability of regression (Table V). This haplotype carried the minor allele (C) of rs2071174 that had also been found associated with an increased probability of regression and the two common alleles of SNPs rs10902073 (C/A) and rs10794281 (T/C) whose variant alleles associated with a reduced probability of regression in the individual SNP analysis. In the 3' region, haplotype CATAGAAC also was associated with a decreased risk of progression (Table V). This haplotype carried the three alleles (T, A and C of rs10794293, rs3924453 and rs4077759, respectively) that had also been found associated with a decreased risk of progression in individual SNP analysis. The haplotype resulting from the combination of these two significant haplotypes in the 5' and 3' regions (CTCCCGCATAGAAC, with a frequency of 0.120) also showed a trend as protector for progression (OR = 0.53, 95% CI: 0.26–1.07, $P = 0.075$) when compared with the ACCTCGCGCAAAGT (frequency of 0.147), although it did not reach statistical significance (results not shown).

As expected, no significant association with progression or regression of the GCPLs was found with the six inferred *MUC1* haplotypes with frequencies between 0.01 and 0.46 and with the six inferred *MUC6* haplotypes with frequencies between 0.06 and 0.37.

Discussion

In this study, we have assessed whether genetic variability in three genes (*MUC1*, *MUC2* and *MUC6*) involved in the mucosal protection associates with the evolution, either progression or regression, of GCPLs after a mean follow-up of 12.8 years in a region with one of the highest incidence of and mortality from GC in Spain and Western Europe (24). The major novel finding is that variability in *MUC2* gene is protective regarding the evolution of GCPLs and this effect is more evident in *H.pylori*-infected patients and in those not consuming NSAIDs.

MUC2 is the major constituent of the secreted mucus barrier in the small and large intestines, where it is mainly produced by goblet cells (29,30). A number of reports have provided evidence on the role of *MUC2* in maintaining the integrity of the protective mucus layer in intestine and reduced *MUC2* expression has been related to inflammation and development of several intestinal diseases, including cancer (3,9,30,31). In contrast, *MUC2*, which is not normally expressed in stomach, is present at increased levels in intestinal metaplasia and in gastric adenocarcinoma (5–7,32,33). It has been reported that *MUC2* overexpression might be the consequence of its activation by proinflammatory cytokines expressed after *H.pylori* infection (33). *MUC2* overexpression may also reflect the stomach transformation during intestinal metaplasia into an intestinal-type epithelium in which glandular gastric cells are replaced by goblet cells. With regard to its role in gastric carcinogenesis, it has been suggested that *MUC2* might contribute to malignant phenotype by protecting against recognition by antitumor immune effectors (9). Therefore, genetic variation in the *MUC2* gene that would modify its expression or biological activity could result in an altered pathway toward gastric carcinogenesis. Only one study has recently assessed the association between

Table III. ORs and 95% CIs for SNPs in *MUC2* locus showing a nominal significant association with progression of preneoplastic gastric lesions, stratified by *Helicobacter pylori* infection status

SNP	Model	Genotype	Progression			<i>H.pylori</i> +			<i>H.pylori</i> -			<i>P</i> interaction
			Controls/cases	OR (CI 95%) ^a	<i>P</i> value	Controls/cases	OR (CI 95%) ^a	<i>P</i> value	Controls/cases	OR (CI 95%) ^a	<i>P</i> value	
rs10794293	Codominant	CC	117/57	1.00		90/45	1.00		26/10	1.00		
		CT	125/40	0.66 (0.40–1.10)		102/29	0.52 (0.29–0.94)		23/11	1.48 (0.48–4.62)		
		TT	33/7	0.43 (0.17–1.08)	0.0905	21/4	0.32 (0.10–1.06)	0.0328	12/3	0.71 (0.14–3.58)	0.6123	
	Dominant	CC	117/57	1.00		90/45	1.00		26/10	1.00		
		CT-TT	158/47	0.61 (0.38–1.00)	0.0476	123/33	0.49 (0.28–0.87)	0.0133	35/14	1.22 (0.42–3.53)	0.7117	
	Recessive	CC-CT	242/97	1.00		192/74	1.00		49/21	1.00		
		TT	33/7	0.52 (0.21–1.28)	0.1382	21/4	0.45 (0.14–1.44)	0.1538	12/3	0.59 (0.13–2.63)	0.4731	
Log-additive	CC, CT, TT	275/104	0.66 (0.45–0.96)	0.0284	213/78	0.54 (0.34–0.87)	0.0092	61/24	0.96 (0.47–1.98)	0.9131		
rs3924453	Codominant	GG	135/60	1.00		112/47	1.00		23/12	1.00		
		GA	112/40	0.77 (0.47–1.27)		82/30	0.83 (0.47–1.46)		29/9	0.70 (0.20–2.43)		
		AA	28/4	0.30 (0.10–0.93)	0.0597	19/1	0.12 (0.01–0.94)	0.0292	9/3	0.36 (0.07–1.97)	0.4600	
	Dominant	GG	135/60	1.00		112/47	1.00		23/12	1.00		
		GA-AA	140/44	0.67 (0.42–1.00)	0.1074	101/31	0.7 (0.40–1.22)	0.2053	38/12	0.56 (0.19–1.70)	0.3066	
	Recessive	GG-GA	247/100	1.00		194/77	1.00		52/21	1.00		
		AA	28/4	0.34 (0.11–1.01)	0.0320	19/1	0.13 (0.02–1.00)	0.0099	9/3	0.40 (0.08–2.11)	0.2665	
Log-additive	GG, GA, AA	275/104	0.66 (0.44–0.97)	0.0317	213/78	0.63 (0.39–1.01)	0.0488	61/24	0.62 (0.29–1.34)	0.2203		
rs4077759	Codominant	TT	105/47	1.00		89/36	1.00		16/10	1.00		
		TC	119/48	0.86 (0.52–1.44)		85/38	1.05 (0.59–1.87)		34/9	0.53 (0.14–1.97)		
		CC	46/7	0.35 (0.14–0.85)	0.0437	35/2	0.14 (0.03–0.65)	0.0045	10/5	0.56 (0.12–2.53)	0.5782	
	Dominant	TT	105/47	1.00		89/36	1.00		16/10	1.00		
		TC-CC	165/55	0.72 (0.44–1.18)	0.1897	120/40	0.80 (0.45–1.40)	0.4267	44/14	0.54 (0.17–1.72)	0.2966	
	Recessive	TT-TC	224/95	1.00		174/74	1.00		50/19	1.00		
		CC	46/7	0.37 (0.16–0.88)	0.0148	35/2	0.14 (0.03–0.62)	0.0010	10/5	0.74 (0.18–3.00)	0.6685	
Log-additive	TT, TC, CC	270/102	0.68 (0.47–0.98)	0.0338	209/76	0.64 (0.42–1.00)	0.0439	60/24	0.71 (0.34–1.51)	0.3732		

Significant *P* values are given in bold.^aUnconditional logistic regression analysis, adjusted for sex, age, NSAID consumption and lesion at recruitment.

Table IV. ORs and 95% CIs for SNPs in *MUC2* locus showing a nominal significant association with regression of preneoplastic gastric lesions stratified by *Helicobacter pylori* infection status

SNP	Model	Genotype	Regression			<i>H.pylori</i> +			<i>H.pylori</i> -			<i>P</i> interaction	
			Controls/cases	OR (CI 95%) ^a	<i>P</i> value	Controls/cases	OR (CI 95%) ^a	<i>P</i> value	Controls/cases	OR (CI 95%) ^a	<i>P</i> value		
rs10902073	Codominant	CC	88/50	1.00		57/45	1.00		29/5	1.00			
		CA	134/48	0.59 (0.36–0.97)		102/37	0.41 (0.23–0.73)		31/11	2.03 (0.58–7.04)			
	Dominant	AA	41/11	0.47 (0.22–1.04)	0.0549	36/9	0.34 (0.14–0.82)	0.0034	5/2	1.45 (0.19–11.10)	0.5249	0.0381	
		CC	88/50	1.00		57/45	1.00		29/5	1.00			
	Recessive	CA-AA	175/59	0.56 (0.35–0.91)	0.0190	138/46	0.39 (0.23–0.68)	0.0008	36/13	1.92 (0.57–6.44)	0.2791	0.0100	
		CC-CA	222/98	1.00		159/82	1.00		60/16	1.00			
rs10794281	Log-additive	AA	41/11	0.64 (0.31–1.33)	0.2191	36/9	0.56 (0.25–1.27)	0.1520	5/2	0.95 (0.15–6.17)	0.9559	0.5431	
		CC, CA, AA	263/109	0.65 (0.46–0.94)	0.0195	195/91	0.52 (0.34–0.79)	0.0016	65/18	1.43 (0.59–3.44)	0.4294	0.0272	
	Codominant	TT	72/41	1.00		45/35	1.00		26/6	1.00			
		TC	121/45	0.60 (0.35–1.04)		89/37	0.47 (0.25–0.87)		30/8	1.26 (0.35–4.52)			
	Dominant	CC	55/17	0.53 (0.27–1.06)	0.1037	48/12	0.32 (0.14–0.71)	0.0080	7/5	2.38 (0.50–11.20)	0.5459	0.0486	
		TT	72/41	1.00		45/35	1.00		26/	1.00			
	Recessive	TC-CC	176/62	0.58 (0.35–0.96)	0.0361	137/49	0.42 (0.23–0.75)	0.0032	37/6	1.53 (0.47–5.01)	0.4787	0.0375	
		TT-TC	193/86	1.00		134/72	1.00		56/13	1.00			
	rs2071174	Log-additive	CC	55/17	0.71 (0.38–1.33)	0.2800	48/12	0.50 (0.24–1.03)	0.0508	7/5	2.09 (0.53–8.20)	0.2969	0.0518
			TT, TC, CC	248/103	0.71 (0.50–1.00)	0.0463	182/84	0.55 (0.37–0.82)	0.0024	63/19	1.51 (0.69–3.29)	0.2972	0.0095
		Codominant	TT	138/53	1.00		105/44	1.00		31/9	1.00		
			TC	110/40	0.93 (0.56–1.54)		78/34	0.98 (0.56–1.73)		31/6	0.77 (0.23–2.61)		
Dominant		CC	17/17	2.76 (1.24–6.13)	0.0275	14/14	2.66 (1.09–6.50)	0.0785	3/3	4.36 (0.61–31.34)	0.2304	0.7917	
		TT	138/53	1.00		105/44	1.00		31/9	1.00			
rs7944723	Recessive	TC-CC	127/57	1.17 (0.73–1.86)	0.5186	92/48	1.21 (0.71–2.05)	0.4784	34/9	4.36 (0.34–3.20)	0.9471	0.7525	
		TT-TC	248/93	1.00		183/78	1.00		62/15	1.00			
	Log-additive	CC	17/17	2.85 (1.32–6.14)	0.0077	14/14	2.68 (1.13–6.32)	0.0241	3/3	4.98 (0.76–32.71)	0.0967	0.6144	
		TT, TC, CC	265/110	1.36 (0.95–1.93)	0.0893	197/92	1.37 (0.92–2.04)	0.1168	65/18	1.45 (0.59–3.53)	0.4181	0.9817	
	Codominant	CC	169/69	1.00		125/57	1.00		42/12	1.00			
		CG	83/33	0.91 (0.54–1.51)		61/27	0.90 (0.51–1.60)		21/6	0.91 (0.28–2.95)			
Dominant		GG	12/1	0.15 (0.02–1.22)	0.0803	9/0	NA	NA	0.0188	3/1	0.84 (0.07–9.76)	0.9807	0.2190
		CC	169/69	1.00		125/67	1.00		42/12	1.00			
Recessive		CG-GG	95/34	0.79 (0.48–1.31)	0.3667	70/32	0.77 (0.44–1.35)	0.3554	24/7	0.90 (0.29–2.77)	0.8508	0.7682	
		CC-CG	252/102	1.00		186/84	1.00		63/18	1.00			
Log-additive	GG	12/1	0.16 (0.02–1.25)	0.0268	9/0	NA	NA	0.0052	3/1	0.87 (0.08–9.73)	0.9098	0.0814	
		CC, CG, GG	264/103	0.72 (0.47–1.12)	0.1416	195/84	0.67 (0.41–1.12)	0.1191	66/19	0.91 (0.36–2.30)	0.8437	0.5491	

Significant *P* values are given in bold.^aUnconditional logistic regression analysis, adjusted for sex, age, NSAID consumption and lesion at recruitment. NA, not available.

Table V. MUC2 haplotypes and association with evolution of lesions

Region ^a	Haplotype	Frequency	Progression			Regression			
			OR	(CI 95%) ^b	P value	OR	(CI 95%) ^b	P value	
5'	ACCTCG	0.3447	1.00	Reference		1.00	Reference		
	CTCTCG	0.2698	0.72	(0.46–1.12)	0.1478	1.58	(0.99–2.52)	0.0548	
	CTCCCG	0.1576	0.49	(0.28–0.85)	0.0117	1.67	(1.02–2.73)	0.0407	
	CTGCTG	0.0463	0.81	(0.33–1.99)	0.6502	1.72	(0.71–4.17)	0.2268	
	CTCTTG	0.0393	0.60	(0.22–1.61)	0.3080	1.25	(0.45–3.52)	0.6697	
	CTCTTA	0.0349	0.84	(0.32–2.21)	0.7279	0.69	(0.23–2.08)	0.5114	
	CCGCTA	0.0334	0.57	(0.20–1.63)	0.2963	1.81	(0.76–4.32)	0.1786	
	CCGCTG	0.0203	0.83	(0.21–3.30)	0.7898	0.29	(0.03–2.64)	0.2707	
	ACCCCG	0.0153	0.46	(0.09–2.42)	0.3584	2.38	(0.44–13.01)	0.3156	
	ACCTTG	0.0140	0.20	(0.02–1.88)	0.1607	1.49	(0.19–11.56)	0.7014	
	3'	CGCAAAGT	0.3223	1.00	Reference		1.00	Reference	
		CATAGAAC	0.1400	0.46	(0.25–0.86)	0.0156	1.16	(0.69–1.95)	0.5810
		CGCAGAGT	0.0927	0.77	(0.39–1.54)	0.4583	1.00	(0.51–1.97)	0.9989
CGCAGAAC		0.0888	0.73	(0.36–1.45)	0.3667	0.86	(0.45–1.65)	0.6439	
GATAAAGT		0.0771	0.49	(0.21–1.14)	0.0991	1.19	(0.59–2.39)	0.6293	
GACAAAGT		0.0696	1.51	(0.74–3.06)	0.2558	0.52	(0.23–1.17)	0.1130	
CGCGGGC		0.0671	0.83	(0.39–1.79)	0.6378	0.57	(0.26–1.24)	0.1552	
CATAGAGT		0.0340	0.93	(0.34–2.51)	0.8835	1.34	(0.51–3.57)	0.5523	
GATAAAAC		0.0153	1.26	(0.30–5.27)	0.7532	0.86	(0.19–3.87)	0.8436	
GATAGAAC		0.0178	0.86	(0.17–4.29)	0.8543	0.43	(0.07–2.70)	0.3679	

Statistically significant values are given in bold.

^a5' region includes the following SNPs: rs10902073, rs10794281, rs2856082, rs2071174, rs7396030 and rs11245936; 3' region includes the following SNPs: rs7944723, rs6421972, rs10794293, rs11245954, rs7480563, rs7126405, rs3924453 and rs4077759.

^bUnconditional logistic regression analysis, adjusted for sex, age, NSAID consumption and lesion at recruitment, log-additive model.

candidate SNPs in *MUC2*, as well as in *MUC5B*, *MUC6* and *MUC1*, with the risk of chronic atrophic gastritis, a specific GCPL, but no association was identified (34).

Interestingly, polymorphisms associated with regression of the GCPLs were located from the central to the 5'-upstream genomic region of the gene, whereas polymorphisms associated with progression were located in the 3' and downstream region. When haplotype analyses were carried out, it was observed that the CTCCCG haplotype in the 5' region of the gene was associated with regression of the GCPLs and, in agreement with it, it also had a protective role against progression of the lesions. Furthermore, the CATAGAAC haplotype in the central and 3' region of the gene was also significantly associated with a decreased risk of progression of the lesions, thereby also indicating a protective role of this haplotype against progression. Finally, when the haplotype analysis was done with all *MUC2* SNPs, it was observed that the CTCCCGCATAGAAC haplotype, which combines the 5' and 3' protective haplotypes, also shows a clear although not significant tendency toward protection against progression of the GCPLs. All these results suggest that either the combination of different genetic variants in the gene or the effect of a yet an unknown variant in LD with the analyzed ones in each side of the central gene region modify the capacity of the lesions to progress toward gastric carcinoma.

Our results do not inform on the nature of the functional variant/s responsible for the observed associations. Nevertheless, since we used a tagSNP genotyping strategy, it is likely that the causative variants are in strong LD with the genotyped and significant ones. Although functional studies and association analyses with expression levels should be performed to analyze the causal relationships between the associated variants and gastric carcinogenesis, several lines of evidence indicate that this relationship might exist. Thus, rs10902073 and rs10794281 tag a strong LD block in the genomic region 5'-upstream of *MUC2*, suggesting that the variant responsible for increased probability of regression might be present in the upstream regulatory sequences of the gene expression. Rs2071174, which was associated with an increased probability of regression, tags a common non-synonymous Val116Met SNP (rs11825977) that has been suggested to be responsible for significant downregulation of *MUC2* in Crohn's disease, due to reduced messenger RNA expression (35).

Rs7944723, located at the end of central exon 30, which encodes for the highly polymorphic TR region, with properties essential to the surface epithelial protective function of mucins (8). This SNP, together with rs10794293 and rs6421972, tags a 5 kb haplotype block in 3' region of the gene that corresponds to C-terminus domain containing an autocatalytic proteolytic cleavage site predicted to cause disassembly of *MUC2* polymers and to affect the mucus viscosity (36,37). Finally, rs4077759 and rs3924453 are tags of a 4 kb haplotype block containing the last exons of the gene and the 3' downstream genomic region where there are other potentially functional variants in the region, such as some missense coding SNPs and 3'-UTR variants. It is interesting to note that we have not found association with any of the *MUC2* missense SNPs analyzed (Table II), which could be indicative of association being more probably due to variants that regulate *MUC2* expression.

The effect of all associated SNPs with GCPLs was found to be stronger and only significant in *H.pylori*-infected patients, being the interaction statistically significant for rs10902073, rs10794281 and rs4077759. Despite the relative low number of patients negative for *H.pylori* infection (23%) and the consequent statistical power limitations, the calculated ORs do not follow the same trend in infected and non-infected patients, suggesting a real interaction between gene variability and infection by the bacteria in the evolution of the lesions. Although a physical interaction between *H.pylori* and *MUC2* has not been established, the importance of this secreted mucin in bacterial colonization and infection by pathogens has been broadly demonstrated in the intestine (31,38). In stomach, *H.pylori* infection has been reported to cause an aberrant expression of *MUC2* gene in gastric carcinoma cells (39). Moreover, ectopic expression of *MUC2* has been associated with the induction of *CDX1* by *cagA* strain of *H.pylori* in gastric epithelial cells that have been arrested in G1 (40). Therefore, although we do not know the precise molecular mechanism by which variation in *MUC2* may interact with *H.pylori* infection to influence the evolution of GCPLs, the fact that the infection associates with aberrant expression of *MUC2* and that, as stated in the above paragraph, genetic variants of *MUC2* may influence its expression or protein function, provide biological support to the joint effect of these two factors. In contrast, although our results show that the effect of these *MUC2* variants on the evolution of the

lesions is significant only in non-NSAID users, no interaction was found between both factors. Given the protective effect of NSAID consumption on the progression of the lesions observed in this study, it seems that the protective effect of genetic variation in *MUC2* is more clearly observed in those patients not already protected by NSAID consumption.

Strengths of this study are the long and well-defined follow-up and its relatively large size that allowed us to evaluate the risk for progression and regression of lesions conferred by genetic variants in *MUC* genes. To our knowledge, only one study has analyzed the association between genetic variability (in DNA repair genes in this case) with evolution of GCPLs, also with the identification of some significant associations (41). Among the weaknesses, it is to mention that although the sample size is relative large considering the characteristics of the follow up, it was not large enough to avoid power limitations to detect other potential associations with some of the analyzed SNPs and genes, such as *MUC1* and *MUC6*. Furthermore, failure in the genotyping of three of the selected tagSNPs, among which *MUC1* rs4072037 that has recently been found associated with gastric cancer (21–23), has reduced the coverage of the genetic variability of these genes by 40% in the case of *MUC1*, but only <10% of *MUC2* and *MUC6*. Therefore, additional genotyping and association analyses in a larger sample series are needed to comprehensively elucidate and confirm the role in the gastric carcinogenesis pathway of genetic variability in the mucin genes, as well as their interaction with other risk factors such as infection by *H.pylori*.

In conclusion, this is the first study to address the association of mucin gene variants with progression and regression of precursor lesions of gastric cancer. Our analysis has provided evidence that genetic polymorphisms in *MUC2* gene are associated with evolution of GCPLs, especially in *H.pylori*-infected patients, suggesting an important role of this secreted mucin in gastric carcinogenesis.

Funding

Spanish Ministry of Health Research Funds ISCIII Exp.030077 and PI081420 and ISCIII RETICC DR06/0020-0015.

Acknowledgements

We thank the participating patients and their families and Pilar Lopez and Amalia Solana for their valuable collaboration in the field work, as well as Raúl Garcia for his collaboration as data manager. We also want to acknowledge the collaboration of Dr Maria Torres, from the Spanish National Genotyping Center (CEGEN), Santiago node.

Conflict of Interest Statement: None declared.

References

- Gum, J.R. Jr (1992) Mucin genes and the proteins they encode: structure, diversity, and regulation. *Am. J. Respir. Cell Mol. Biol.*, **7**, 557–564.
- Kim, Y.S. *et al.* (1995) Diversity of mucin genes, structure, function, and expression. *Gastroenterology*, **109**, 999–1001.
- Hollingsworth, M.A. *et al.* (2004) Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer*, **4**, 45–60.
- Senapati, S. *et al.* (2008) The MUC gene family: their role in the diagnosis and prognosis of gastric cancer. *Histol. Histopathol.*, **23**, 1541–1552.
- Ho, S.B. *et al.* (1995) Mucin gene expression in normal, preneoplastic, and neoplastic human gastric epithelium. *Cancer Res.*, **55**, 2681–2690.
- Reis, C.A. *et al.* (1999) Intestinal metaplasia of human stomach displays distinct patterns of mucin (*MUC1*, *MUC2*, *MUC5AC*, and *MUC6*) expression. *Cancer Res.*, **59**, 1003–1007.
- Babu, S.D. *et al.* (2006) Expression profile of mucins (*MUC2*, *MUC5AC* and *MUC6*) in *Helicobacter pylori* infected pre-neoplastic and neoplastic human gastric epithelium. *Mol. Cancer*, **5**, 10.
- Andrianifahanana, M. *et al.* (2006) Regulation of mucin expression: mechanistic aspects and implications for cancer and inflammatory diseases. *Biochim. Biophys. Acta*, **1765**, 189–222.
- Kufe, D.W. (2009) Mucins in cancer: function, prognosis and therapy. *Nat. Rev. Cancer*, **9**, 874–885.
- Correa, P. (1992) Human gastric carcinogenesis: a multistep and multifactorial process—First American cancer Society Award Lecture on cancer Epidemiology and Prevention. *Cancer Res.*, **52**, 6735–6740.
- Yuasa, Y. (2003) Control of gut differentiation and intestinal-type gastric carcinogenesis. *Nat. Rev. Cancer*, **3**, 592–600.
- Gonzalez, C.A. *et al.* (2002) Genetic susceptibility and gastric cancer risk. *Int. J. Cancer*, **100**, 249–260.
- Rousseau, K. *et al.* (2004) The complete genomic organization of the human *MUC6* and *MUC2* mucin genes. *Genomics*, **83**, 936–939.
- Swallow, D.M. *et al.* (1987) The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. *Nature*, **328**, 82–84.
- Gendler, S.J. *et al.* (1990) Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.*, **265**, 15286–15293.
- Carvalho, F. *et al.* (1999) *MUC1* gene polymorphism does not explain the different incidence of gastric cancer in Portugal and Denmark. *Ann. Hum. Genet.*, **63**, 187–191.
- Carvalho, F. *et al.* (1997) *MUC1* gene polymorphism and gastric cancer—an epidemiological study. *Glycoconj. J.*, **14**, 107–111.
- Garcia, E. *et al.* (1997) *MUC6* gene polymorphism in healthy individuals and in gastric cancer patients from northern Portugal. *Cancer Epidemiol. Biomarkers Prev.*, **6**, 1071–1074.
- Silva, F. *et al.* (2001) *MUC1* gene polymorphism in the gastric carcinogenesis pathway. *Eur. J. Hum. Genet.*, **9**, 548–552.
- Silva, F. *et al.* (2003) *MUC1* polymorphism confers increased risk for intestinal metaplasia in a Colombian population with chronic gastritis. *Eur. J. Hum. Genet.*, **11**, 380–384.
- Xu, Q. *et al.* (2009) Risk of gastric cancer is associated with the *MUC1* 568 A/G polymorphism. *Int. J. Oncol.*, **35**, 1313–1320.
- Jia, Y. *et al.* (2010) A comprehensive analysis of common genetic variation in *MUC1*, *MUC5AC*, *MUC6* genes and risk of stomach cancer. *Cancer Causes Control*, **21**, 313–321.
- Saeki, N. *et al.* (2011) A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology*, **140**, 892–902.
- Gonzalez, C.A. *et al.* (2010) Gastric cancer occurrence in preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. *Int. J. Cancer*, **127**, 2654–2660.
- Gabriel, S.B. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
- Barrett, J.C. *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
- Gonzalez, C.A. *et al.* (2011) *Helicobacter pylori* *cagA* and *vacA* genotypes as predictors of progression of gastric preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. *Am. J. Gastroenterol.*, **106**, 867–874.
- Gonzalez, J.R. *et al.* (2007) SNPpass: an R package to perform whole genome association studies. *Bioinformatics*, **23**, 644–645.
- Chang, S.K. *et al.* (1994) Localization of mucin (*MUC2* and *MUC3*) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology*, **107**, 28–36.
- Kim, Y.S. *et al.* (2010) Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Curr. Gastroenterol. Rep.*, **12**, 319–330.
- McGuckin, M.A. *et al.* (2011) Mucin dynamics and enteric pathogens. *Nat. Rev. Microbiol.*, **9**, 265–278.
- Silva, E. *et al.* (2002) Mucins as key molecules for the classification of intestinal metaplasia of the stomach. *Virchows Arch.*, **440**, 311–317.
- Mejias-Luque, R. *et al.* (2010) Inflammation modulates the expression of the intestinal mucins *MUC2* and *MUC4* in gastric tumors. *Oncogene*, **29**, 1753–1762.
- Frank, B. *et al.* (2012) Polymorphisms in *MUC1*, *MUC2*, *MUC5B* and *MUC6* genes are not associated with the risk of chronic atrophic gastritis. *Eur. J. Cancer*, **48**, 114–120.
- Moehle, C. *et al.* (2006) Aberrant intestinal expression and allelic variants of mucin genes associated with inflammatory bowel disease. *J. Mol. Med.*, **84**, 1055–1066.
- Lidell, M.E. *et al.* (2003) An autocatalytic cleavage in the C terminus of the human *MUC2* mucin occurs at the low pH of the late secretory pathway. *J. Biol. Chem.*, **278**, 13944–13951.
- Lidell, M.E. *et al.* (2006) *Entamoeba histolytica* cysteine proteases cleave the *MUC2* mucin in its C-terminal domain and dissolve the protective colonic mucus gel. *Proc. Natl Acad. Sci. USA*, **103**, 9298–9303.
- Moran, A.P. *et al.* (2011) Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut*, **60**, 1412–1425.
- Matsuda, K. *et al.* (2008) Quantitative analysis of the effect of *Helicobacter pylori* on the expressions of *SOX2*, *CDX2*, *MUC2*, *MUC5AC*, *MUC6*,

- TFF1, TFF2, and TFF3 mRNAs in human gastric carcinoma cells. *Scand. J. Gastroenterol.*, **43**, 25–33.
40. Murata-Kamiya, N. *et al.* (2007) *Helicobacter pylori* CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene*, **26**, 4617–4626.
41. Li, W.Q. *et al.* (2009) Association between genetic polymorphisms of DNA base excision repair genes and evolution of precancerous gastric lesions in a Chinese population. *Carcinogenesis*, **30**, 500–505.

Received November 10, 2011; revised February 15, 2012; accepted March 3, 2012

Simultaneous Genotyping of *GSTT1* and *GSTM1* Null Polymorphisms by Melting Curve Analysis in Presence of SYBR Green I

Fátima Marín,^{*†} Nadia García,^{*†} Xavier Muñoz,^{*†}
Gabriel Capellà,^{*} Carlos A. González,[†]
Antonio Agudo,[†] and Núria Sala^{*†}

From the Translational Research Laboratory,^{*} and the Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Program,[†] Institut Català d'Oncologia (IDIBELL-ICO), L'Hospitalet de Llobregat, Spain

Due to their ability to metabolize xenobiotics, glutathione S-transferases (GSTs) play an important role in cellular protection. GST family members μ (*GSTM1*) and θ (*GSTT1*) exhibit a common polymorphism that results in the complete deletion of the gene (null allele). Homozygous deletions, which result in the absence of the enzyme, are considered a risk factor for several diseases, including cancer. We report a simple, low cost, and high throughput assay for the simultaneous analysis of the *GSTM1* and *GSTT1* null polymorphisms in a single step. The assay is based on multiplex real-time PCR in the presence of SYBR Green I and genotype discrimination by melting curve analysis in a LightCycler. We have genotyped 792 samples to compare this new approach with conventional PCR followed by gel electrophoresis. Comparison of the methods gave a good agreement, with κ values of 0.88 for *GSTM1* and 0.64 for *GSTT1*. Reanalysis of discrepant samples indicated that absence of amplification of the larger *GSTT1* fragment by conventional PCR accounted for most of the discrepancies. Moreover, the improved amplification efficiency of the real-time PCR results in a significant reduction of missing values. Due to its simplicity and low cost, this assay is well suited for the rapid analysis of GST-null genotypes in studies that involve large number of samples. (J Mol Diagn 2010, 12:300–304; DOI: 10.2353/jmoldx.2010.090076)

Glutathione S-transferases (GSTs) are an important family of enzymes that catalyze the detoxification of a wide variety of xenobiotics, including environmental carcinogens, chemotherapeutic agents, and reactive oxygen species.¹ They are involved in phase II detoxification, mediating the conjugation of reduced glutathione to electrophilic species that leads to the elimination of toxic compounds. GSTs have also been shown to play a role in kinase signaling and glutathionylation.² Therefore, GSTs

play an important role in cellular protection from environmental exposures and oxidative stress and they are also implicated in cellular resistance to drugs.³

The members μ (*GSTM1*) and θ (*GSTT1*) of the multigenic family of GSTs exhibit a common polymorphism that is characterized by the complete deletion of the gene (null allele).² Homozygous deletions (null genotype) thus result in the absence of enzyme and its catalytic activity, which has been shown to decrease the cells' ability to detoxify certain genotoxic agents.⁴ The null genotypes of GSTs have been associated with enhanced genotoxicity and are believed to be key factors in determining susceptibility to diseases associated with exposure to xenobiotics, such as cancer.^{2,3,5} The frequency of the null genotype varies in different populations and occurs in about 0.53 for *GSTM1**0/0 and 0.20 for *GSTT1**0/0 in Caucasians.⁶

Up to now, the most widely used method for genotyping null polymorphisms has been conventional multiplex PCR followed by gel electrophoresis analysis.⁷ This methodology is time consuming and, therefore, may not be the most suitable method for molecular epidemiology and clinical studies that involve large sample numbers and require simple and low cost methodologies with mid-high throughput. In recent years, other techniques based on real-time PCR have been developed. These techniques use specific hybridization probes to distinguish between GSTs homozygous deletions and the presence of at least one GST allele,⁸ or quantitative strategies to identify the three possible genotypes for each polymorphism.^{9–12} Unfortunately, none of these techniques al-

Supported by Health Research Fund (FIS) of the Spanish Ministry of Health (exp. PI020652, PI051932, and PI081420); European Commission FP5, project (QLG1-CT-2001-01049); Fundació "LaCaixa" (exp. BM06-130-0); and the AGAUR, Generalitat de Catalunya (exp. 2002-PIR-00333). X.M. is partially supported by *Instituto de Salud Carlos III* of the National Health System (exp. CA06/0200). The authors are members of ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility), a Network of Excellence of the 6th EU Framework Programme (FP6, FOOD-CT-2005-513 943), and of the ISCIII Spanish Ministry of Health network RTICCC (ISCIII DR06/0020).

Accepted for publication October 9, 2009.

Address reprint requests to Núria Sala, Ph.D., Laboratori de Recerca Translacional, Institut Català d'Oncologia ICO-IDIBELL, Avda. Gran Via de l'Hospitalet 199-203, 08907 L'Hospitalet de Llobregat (Barcelona), Spain. E-mail: nsala@iconcologia.net.

lows simultaneous analysis of the two genes in the same display.

The objective of this work was to develop a simple, single-step assay for the simultaneous analysis of the *GSTM1* and *GSTT1* null polymorphisms in studies that need rapid and effective mid to high-throughput genotyping. To this end, we have developed an alternative method based on multiplex real-time PCR and genotype discrimination by melting curve analysis in the presence of SYBR Green I, a fluorescent dye that specifically binds to double stranded DNA (dsDNA). We have investigated the suitability of this method and assessed its performance as compared with conventional multiplex PCR and gel electrophoresis analysis.

Materials and Methods

DNA Samples

DNA samples from 243 gastric cancer cases and 946 matched controls, selected from the European Prospective Investigation into Cancer and nutrition (EPIC) cohort according to a nested case-control design,¹³ were used to analyze the *GSTT1* and *GSTM1* null polymorphisms by real-time PCR and melting curve analysis in a LightCycler 2.0 (LC2.0). DNAs were extracted with the Puregene DNA Purification System (Gentra Systems, Minneapolis, MN) adapted to the Gentra Autopure LS DNA preparation platform (Gentra Systems) and dried for storage. Before use, DNAs were reconstituted with water.¹³ Moreover, samples of 792 subjects (cases and controls) from this study were reanalyzed by conventional multiplex PCR and gel electrophoresis. A group of 300 DNA samples from healthy subjects selected from the EPIC Spain cohort¹⁴ were used to validate the assay for performance in the LightCycler 480 (LC480). In this study, DNA was prepared by phenol-chloroform extraction and ethanol precipitation. DNA was then dissolved in distilled water.¹⁴

Genotyping by Real-Time PCR and Melting Curve Analysis with SYBR Green I

The real-time PCR was initially performed in the LightCycler (LC) 2.0 instrument, using the fluorescent dye SYBR Green I (Roche Diagnostics, Mannheim, Germany).

The PCR reaction was performed in LC capillaries. Fragments of *GSTT1* (257 bp), *GSTM1* (219 bp), and *BCL2* (155 bp) genes, the last one used as an internal amplification control to exclude false negatives, were simultaneously amplified. The primers used were as previously described¹⁵ but the reverse primer for the *GSTT1* fragment was replaced by 5'-GGAAAAGGGTACAGACTGGGGA-3' (nt9586-9607, AB057594) to obtain a shorter amplification product and, therefore, a lower melting temperature.

Different primer and MgCl₂ concentrations were tested. In the final protocol, PCR mixtures contained 4 mmol/L MgCl₂, 0.3 μmol/L each *GSTT1* primers, 0.4 μmol/L each *GSTM1* primers, 0.5 μmol/L each *BCL2* primers, 1× LightCycler Fast Start DNA Master SYBR

Green I Master Mix (Roche Diagnostics, Mannheim, Germany), and 10 ng of DNA in a final volume of 10 μl. General PCR conditions were initial denaturation at 95°C for 10 minutes and 38 cycles of 95°C for 5 s, 62°C for 15 s, and 72°C for 20 s, at transition rates of 20°C/s. The fluorescence of the dsDNA bound SYBR Green was recorded in each sample after product extension, which allowed for monitoring of the amplification reaction.

Since each amplified sequence is characterized by its apparent melting temperature (T_m), which is a function of product length and base composition, a melting curve analysis was performed subsequent to the amplification of the target sequence. The program for the melting curve analysis was: 95°C for 5 s, 65°C for 20 s, with 20°C/s transition rate, and then ramping to 98°C at 0.2°C/s transition rate. After that, each amplified fragment was identified by converting its specific melting curves, measured as fluorescence emission decrease, to melting peaks by the LC software.

Each series of amplifications included a negative control without DNA template and a positive control with samples of known genotype, which had been previously confirmed by PCR and gel electrophoresis. We reanalyzed 9% (*n* = 110) of the total group of samples as quality control of the technique. The error rate was 1% in each polymorphism.

The final real-time PCR protocol was adjusted for use in the LightCycler 480 (LC480) instrument (Roche Diagnostics, Mannheim, Germany), which uses 384-well plates instead of capillaries. The primers were the same as those used for the LC2.0 assay. The PCR mixtures contained 3.5 mmol/L MgCl₂, 0.3 μmol/L each *GSTT1* primers, 0.4 μmol/L each *GSTM1* primers, 0.5 μmol/L each *BCL2* primers, 1× LightCycler 480 SYBR Green I Master Mix (Roche Diagnostics, Mannheim, Germany), and 10 ng of DNA in a final volume of 10 μl. The fragment amplification was performed as follows: initial denaturation at 95°C for 10 minutes and 33 cycles of 95°C for 10 s, 62°C for 30 s, and 72°C for 25 s, at transition rates of 2 to 4.8°C/s. The program for the melting curve analysis was: 95°C for 10 s, 65°C for 1 minute with 2 to 4.8°C/s transition rate, and then ramping to 97°C at 5 acquisitions per °C.

Genotyping by Multiplex PCR and Gel Electrophoresis

The primers used to simultaneously amplify the fragments of the *GSTT1* (480 bp), *GSTM1* (219 bp), and *BCL2* (155 bp) genes had been previously described.¹⁵ The PCR reaction was performed in a GeneAmp PCR System 9700 instrument (Applied Biosystems), in a total volume of 20 μl with 50 ng of genomic DNA, 0.2 μmol/L of each primer pairs, 200 μmol/L dNTPs, 4 mmol/L MgCl₂, 4 μl of 10× PCR Buffer, and 1 U of TaqDNA polymerase (Roche Diagnostics, Mannheim, Germany). Conventional PCR conditions were: initial denaturation at 95°C for 5 minutes, 40 cycles of 94°C for 90 s, 62°C for 90 s, 72°C for 90 s, and a final step of 72°C for 5 minutes.

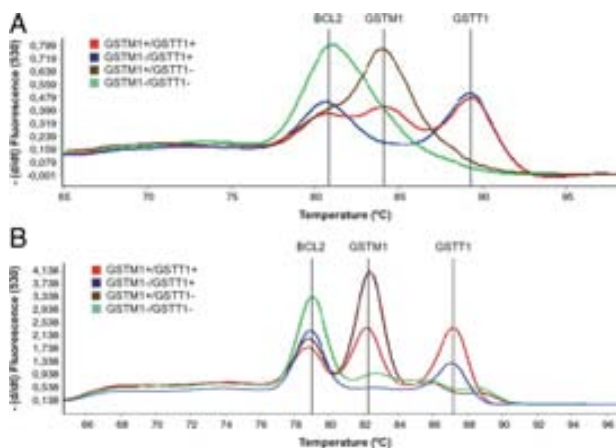


Figure 1. *GSTM1* and *GSTT1* null polymorphism genotyping by real-time PCR and melting curve analysis in the presence of SYBR Green. *T_m* melting peaks of the *GSTT1*, *GSTM1*, and the *BCL2* amplification control gene in: The Light Cycler 2.0 (A), and The Light Cycler 480 (B). Red line corresponds to *GSTM1* and *GSTT1* present genotype, blue to *GSTM1* null and *GSTT1* present genotype, brown to *GSTM1* present and *GSTT1* null genotype and green to *GSTM1* and *GSTT1* null genotype. “+” corresponds to the heterozygous or homozygous genotype for the gene presence. “-” corresponds to the homozygous null genotype.

The PCR products were analyzed by electrophoresis in 2% agarose gels and ethidium bromide staining.

Validation Analysis

Samples with discrepant results between both methods were reanalyzed by singleplex and multiplex long-PCR using combinations of previously described primers.^{16,17}

Statistical Analysis

The genotyping results were classified as *GST* null (homozygous deletion), *GST* present (heterozygous or homozygous genotype for the gene presence) and missing (there was no or very poor internal control fragment amplification). Comparison of proportions was done by means of the McNemar test for paired data.¹⁸ The agreement between results obtained with each method was assessed by means of the kappa index (κ) with its corresponding 95% confidence intervals (CI).¹⁸ This index takes into account the expected proportion of agreement, with a range from -1 to +1; the value 0 corresponds to no agreement beyond what is expected by chance, while +1/-1 correspond to perfect agreement or complete disagreement respectively. A scale has been proposed for the interpretation of κ ,¹⁹ according to which a value below 0.20 indicates poor agreement, for 0.21 to 0.40 the agreement is fair; 0.41 to 0.60, moderate; 0.61 to 0.80, good; and 0.81 to 1, very good. Finally, the proportion of null alleles for both *GST* genes were compared¹⁸ to the theoretical proportion in the population.⁶

Results

Figure 1 shows the melting peaks of the *GSTT1*, *GSTM1*, and *BCL2* fragments amplified in the LC2.0 (Figure 1A)

and in the LC480 (Figure 1B) from four individual DNAs having different combinations of the *GSTM1* and *GSTT1* genotypes. Similar patterns were obtained when the initial DNA amount varied between 3 to 25 ng. Analysis at different primer concentrations indicated that the best melting curve pattern was obtained by increasing the primer concentration of the shorter fragments (data not shown).

Comparison of *GSTM1* null polymorphisms analysis by real-time PCR (LC2.0) and by conventional PCR in 792 samples is shown in Table 1. *GSTM1* genotyping by multiplex PCR and gel electrophoresis produced missing results in 166 samples (21%), while only 10 subjects (1.3%) had missing results by real-time PCR, which is significantly lower (P value <0.0001). Overall, the agreement between both methods is acceptable, with κ (95%CI) 0.59 (0.54 to 0.59). However, it is much better when only the 619 subjects with defined (non-missing) results by both methods were taken into account: the observed proportion of agreement was 94% with a κ value of 0.88 (0.84 to 0.92). Validation analysis revealed that 30 out of 37 discrepant genotypes agreed with those obtained by real-time PCR, while only three agreed with the null genotype obtained by conventional PCR (Table 1). Four samples couldn't be amplified. The proportion of *GSTM1* null genotypes as measured by real-time PCR was 51.4% (402 subjects out of 782 genotyped), while for conventional PCR it was 53.7% (336 out of 626). None of them differed significantly from the 53.1% reported in Caucasian populations.⁶ Since most missing values were due to absence of amplification, all but one samples with a missing value for *GSTM1* were the same as for *GSTT1* (Table 2). Therefore, the pattern of *GSTT1* missing values was similar to that described previously. Overall, the agreement in *GSTT1* results was lower than for *GSTM1*. We found an observed proportion of agreement of 88.7%, with a κ (95% CI) of 0.64 (0.56 to 0.71) among 620 subjects with defined (non-missing) results by both methods, while the overall observed agreement (including missing values) was 70% with κ 0.36 (0.33 to 0.36). Reanalysis of discrepant samples showed that 60 out of 70 genotypes were in agreement with those obtained by

Table 1. Comparison of *GSTM1* Null Polymorphism Genotyping Results Obtained by Real-Time PCR in the Presence of SYBR Green I and by Conventional PCR and Gel Electrophoresis

Conventional	Real-time PCR (LC2.0)			Total
	Missing	Present	Null	
Missing	3 (0.4%)	65 (8.2%)	98 (12.4%)	166 (21.0%)
Present	0 (0.0%)	284 (35.9%)	6* (0.8%)	290 (36.6%)
Null	7 (0.9%)	31† (3.9%)	298 (37.6%)	336 (42.4%)
Total	10 (1.3%)	380 (48.0%)	402 (50.8%)	792 (100.0%)

*Reanalysis of these 6 samples by a third method indicated that 4 had a null genotype. Two did not amplify.

†Reanalysis of these 31 samples by a third method indicated that 26 carried the *GSTM1* gene, 3 were null alleles and 2 did not amplify.

Table 2. Comparison of *GSTT1* Null Polymorphism Genotyping Results Obtained by Real-Time PCR in the Presence of SYBR Green I and by Conventional PCR and Gel Electrophoresis

Conventional	Real-time PCR (LC2.0)			Total
	Missing	Present	Null	
Missing	3 (0.4%)	131 (16.5%)	31 (3.9%)	165 (20.8%)
Present	0 (0.0%)	467 (59.0%)	8* (1.0%)	475 (60.0%)
Null	7 (0.9%)	62† (7.8%)	83 (10.5%)	152 (19.2%)
Total	10 (1.3%)	660 (83.3%)	122 (15.4%)	792 (100.0%)

*Reanalysis of these 8 samples by a third method indicated that 7 had a null genotype and 1 carried the *GSTT1* gene.

†Reanalysis of these 62 samples by a third method indicated that 53 carried the *GSTT1* gene, 3 were null alleles and 6 did not amplify.

real-time, while only four agreed with conventional PCR (Table 2). Six samples couldn't be amplified. The prevalence of *GSTT1* null genotypes obtained by real-time PCR was 15.6% (122 out of 782 genotyped), which was significantly lower (P value = 0.004) than the 19.7% reported in Caucasian populations.⁶ On the contrary, the proportion of *GSTT1* null genotypes obtained by conventional PCR was 24.2%, significantly higher (P value = 0.001) than reported 19.7%.⁶

Furthermore, 300 samples of healthy subjects from the EPIC Spain cohort, previously analyzed for *GSTM1* and *GSTT1* by real-time PCR in the LC2.0,¹⁴ were reanalyzed in the LC480. Among 287 subjects with non-missing values, we found a proportion of observed agreement of 96%, with a κ value of 0.92 for *GSTM1*, and 99.7% (only one discordant result) with κ of 0.99 for *GSTT1*. Reanalysis of the 12 samples with discordant *GSTM1* results (8 *GSTM1* present and 4 null alleles, in the LC480) indicated that 50% of the results agreed with those obtained with the LC480 and 50% agreed with the LC2.0, suggesting that the discrepancy between the instruments was due to chance or sampling errors. The proportion of *GSTM1* null genotypes in this population of Spanish controls was 47.2%, which was similar to the 49.7% reported by Garte⁶ for the Spanish population. By contrast, the proportion of *GSTT1* null genotypes was found to be of 15.0%, which again is lower than the 20.5% reported by Garte.⁶

Discussion

Molecular epidemiology studies, as well as molecular diagnosis, need rapid and low cost methods for the analysis of a great number of samples. Real-time PCR and melting curve analysis in the presence of SYBR Green I offers the opportunity for rapid and specific analysis of the presence/absence of DNA fragments of different melting temperatures and requires only small amounts of DNA.

To our knowledge, the application of this technology to the analysis of *GSTT1* and *GSTM1* null polymorphisms has not been previously described, and the adaptation of the LC2.0 protocol to a platform of higher throughput

such as the LC480 instrument is of great advantage for studies involving large sample numbers. The genotyping method is performed in solution from beginning to end and no separation or purification steps are needed, which minimizes handling errors. The closed system for amplification and genotyping avoids end-product contamination²⁰ and the improved amplification efficiency of the real time PCR versus the conventional PCR²¹ results in a significant reduction of missing values, which is another important point to be considered. In comparison with other real-time methods that use hybridization probes, the cost of genotyping using real-time PCR with SYBR Green I is four times lower when the LC480 is used. Furthermore, genotype discrimination by Tm peaks reduces the time and effort during assay setup and removes the need for downstream relative quantification, allowing for the simultaneous analysis of the two GST genes in the same display.

A disadvantage of this method, compared with those based on quantitative or on long-PCR for the detection of the null GSTs alleles,^{9,11,12,16} is that it does not discriminate between the presence of one or two copies of the gene. Nevertheless, the role of the heterozygote genotype for null GST polymorphisms on disease risk is poorly understood. If association between GST null alleles and disease risk follows a recessive model (risk is only associated with homozygosity for the null allele), then the discrimination between homozygotes and heterozygotes for the gene presence will not be necessary. In this case the application of more economic and rapid methods for the efficient measurement of null allele homozygotes will be preferable.

Although all of the fragments are efficiently amplified with the same primer concentration, the shortest fragments bound less SYBR Green I, which results in low fluorescence emission and a small, hardly detectable, melting peak. Using different primer concentration, we obtained a larger amount of amplification product of the shortest fragments and, therefore, an increment of the fluorescent emission that resulted in a better discrimination of the peaks.

The proportion of agreement in the genotyping results of the two genes, comparing real-time and conventional PCR, was good, although the κ index for *GSTT1* was lower than *GSTM1*. While the frequency of *GSTM1* null genotypes obtained by either method were very similar to the ones previously reported in Caucasian and in Spanish populations, the *GSTT1* null genotype frequencies obtained by the two methods were significantly different from each other and from those previously reported.⁶ The differences obtained when both methods were applied to the same sample group (EPIC samples) could be explained by the fact that the conventional multiplex PCR method used here to amplify a *GSTT1* fragment larger than the one amplified by real-time PCR, gave a higher number of false negatives due to a reduced efficiency in the amplification of large fragments. The results obtained when the discrepant samples were reanalyzed by a third method are in agreement with this hypothesis. We do not know if this could also explain the differences obtained between null *GSTT1* genotype frequencies in our popu-

lations and in those analyzed by Garte et al. The null *GST* genotype frequencies obtained in our study by the melting curve method were similar to those recently obtained by real-time PCR in a large Danish population.¹¹

In summary, the analysis of melting peaks after real-time PCR allows clear discrimination, in one-step and simultaneously for the *GSTM1* and *GSTT1* genes, between homozygotes for the null allele and carriers of any of these genes, with a good proportion of agreement between real-time and conventional PCR methods. Furthermore, the improved amplification efficiency of real-time PCR results in a significant reduction of missing values. Finally, the method described here is relatively inexpensive and labor-saving. Because of its high sensitivity and specificity, rapid performance and robustness, this assay is well suited for the analysis of *GST* null polymorphisms in large scale analysis, and could help overcome the controversy regarding the association between the *GSTT1* and *GSTM1* null polymorphisms and a wide variety of diseases.

Acknowledgments

We acknowledge the personnel from Roche Applied Science in Spain for their support. We also acknowledge Chris Maxwell and Eric J. Duell for their helpful comments and revision of the manuscript.

References

- Hayes JD, Strange RC: Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology* 2000, 61:154–166
- Mcllwain CC, Townsend DM, Tew KD: Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene* 2006, 25:1639–1648
- Daly AK: Pharmacogenetics of the major polymorphic metabolizing enzymes. *Fundamental Clin Pharmacol* 2003, 17:27–41
- Norppa H: Cytogenetic biomarkers and genetic polymorphisms. *Toxicol Lett* 2004, 149:309–334
- Reszka E, Wasowicz W, Gromadzinska J: Genetic polymorphism of xenobiotic metabolising enzymes, diet and cancer susceptibility. *Br J Nutrition* 2006, 96:609–619
- Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, Baranova H, Bathum L, Benhamou S, Boffetta P, Bouchardy C, Breskvar K, Brockmoller J, Cascorbi I, Clapper ML, Coutelle C, Daly A, Dell'Omo M, Dolzan V, Dresler CM, Fryer A, Haugen A, Hein DW, Hildesheim A, Hirvonen A, Hsieh LL, Ingelman-Sundberg M, Kalina I, Kang D, Kihara M, Kiyohara C, Kremers P, Lazarus P, Le Marchand L, Lechner MC, van Lieshout EM, London S, Manni JJ, Maugard CM, Morita S, Nazar-Stewart V, Noda K, Oda Y, Parl FF, Pastorelli R, Persson I, Peters WH, Rannug A, Rebbeck T, Risch A, Roelandt L, Romkes M, Ryberg D, Salagovic J, Schoket B, Seidegard J, Shields PG, Sim E, Sinnet D, Strange RC, Stucker I, Sugimura H, To-Figuerras J, Vineis P, Yu MC, Taioli E: Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 2001, 10:1239–1248
- Abdel-Rahman SZ, el-Zein RA, Anwar WA, Au WW: A multiplex PCR procedure for polymorphic analysis of *GSTM1* and *GSTT1* genes in population studies. *Cancer Lett* 1996, 107:229–233
- Ko Y, Koch B, Harth V, Sachinidis A, Thier R, Vetter H, Bolt HM, Bruning T: Rapid analysis of *GSTM1*, *GSTT1*, and *GSTP1* polymorphisms using real-time polymerase chain reaction. *Pharmacogenetics* 2000, 10:271–274
- Brasch-Andersen C, Christiansen L, Tan Q, Haagerup A, Vestbo J, Kruse TA: Possible gene dosage effect of glutathione-S-transferases on atopic asthma: using real-time PCR for quantification of *GSTM1* and *GSTT1* gene copy numbers. *Human Mutation* 2004, 24:208–214
- Girault I, Lidereau R, Bieche I: Trimodal *GSTT1* and *GSTM1* genotyping assay by real-time PCR. *Int J Biol Markers* 2005, 20:81–86
- Norskov MS, Frikke-Schmidt R, Loft S, Tybjaerg-Hansen A: High-throughput genotyping of copy number variation in glutathione S-transferases M1 and T1 using real-time PCR in 20,687 individuals. *Clin Biochem* 2009, 42:201–209
- Timofeeva M, Jager B, Rosenberger A, Sauter W, Wichmann HE, Bickeboller H, Risch A: A multiplex real-time PCR method for detection of *GSTM1* and *GSTT1* copy numbers. *Clin Biochem* 2009, 42:500–509
- Agudo A, Sala N, Pera G, Capella G, Berenguer A, Garcia N, Palli D, Boeing H, Del Giudice G, Saieva C, Carneiro F, Berrino F, Sacerdote C, Tumino R, Panico S, Berglund G, Siman H, Stenling R, Hallmans G, Martinez C, Bilbao R, Barricarte A, Navarro C, Quiros JR, Allen N, Key T, Bingham S, Khaw KT, Linseisen J, Nagel G, Overvad K, Tjonneland A, Olsen A, Bueno-de-Mesquita HB, Boshuizen HC, Peeters PH, Numans ME, Clavel-Chapelon F, Boutron-Ruault MC, Trichopoulos A, Lund E, Offerhaus J, Jenab M, Ferrari P, Norat T, Riboli E, Gonzalez CA: Polymorphisms in metabolic genes related to tobacco smoke and the risk of gastric cancer in the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev* 2006, 15:2427–2434
- Agudo A, Peluso M, Sala N, Capella G, Munnia A, Piro S, Marin F, Ibanez R, Amiano P, Tormo MJ, Ardanaz E, Barricarte A, Chirlaque MD, Dorronsoro M, Larranaga N, Martinez C, Navarro C, Quiros JR, Sanchez MJ, Gonzalez CA: Aromatic DNA adducts and polymorphisms in metabolic genes in healthy adults: findings from the EPIC-Spain cohort. *Carcinogenesis* 2009, 30:968–976
- Voso MT, D'Alo F, Putzulu R, Mele L, Scardocci A, Chiusolo P, Latagliata R, Lo-Coco F, Rutella S, Pagano L, Hohaus S, Leone G: Negative prognostic value of glutathione S-transferase (*GSTM1* and *GSTT1*) deletions in adult acute myeloid leukemia. *Blood* 2002, 100:2703–2707
- Buchard A, Sanchez JJ, Dalhoff K, Morling N: Multiplex PCR detection of *GSTM1*, *GSTT1*, and *GSTP1* gene variants: simultaneously detecting *GSTM1* and *GSTT1* gene copy number and the allelic status of the *GSTP1* Ile105Val genetic variant. *J Mol Diagn* 2007, 9:612–617
- Sprenger R, Schlagenhauser R, Kerb R, Bruhn C, Brockmoller J, Roots I, Brinkmann U: Characterization of the glutathione S-transferase *GSTT1* deletion: discrimination of all genotypes by polymerase chain reaction indicates a trimodular genotype-phenotype correlation. *Pharmacogenetics* 2000, 10:557–565
- Altman DG: *Practical statistics for medical research*. London, Chapman & Hall, 1991
- Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977, 33:159–174
- Lyon E, Wittwer CT: LightCycler technology in molecular diagnostics. *J Mol Diagn* 2009, 11:93–101
- Bustin SA, Benes V, Nolan T, Pfaffl MW: Quantitative real-time RT-PCR—a perspective. *J Mol Endocrinol* 2005, 34:597–601

Aromatic DNA adducts and polymorphisms in metabolic genes in healthy adults: findings from the EPIC-Spain cohort

Antonio Agudo^{1,*}, Marco Peluso², Núria Sala^{1,3}, Gabriel Capellá³, Armelle Munnia², Sara Piro², Fátima Marín³, Raquel Ibáñez¹, Pilar Amiano^{4,5}, M.José Tormo^{5,6}, Eva Ardanaz^{5,7}, Aurelio Barricarte^{5,7}, M.Dolores Chirlaque^{5,6}, Miren Dorronsoro^{4,5}, Nerea Larrañaga^{4,5}, Carmen Martínez^{5,8}, Carmen Navarro^{5,6}, J.Ramón Quirós⁹, M.José Sánchez^{5,8} and Carlos A.González¹

¹Unit of Nutrition, Environment, and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), IDIBELL, Avenue Gran Via s/n km 2.7, 08907 L'Hospitalet de Llobregat, Spain, ²Cancer Risk Factor Branch, Molecular Biology Laboratory, CSPO-Scientific Institute of Tuscany Region, 50135 Florence, Italy, ³Translational Research Laboratory, Catalan Institute of Oncology (ICO), IDIBELL, Avenue Gran Via, 199-203, 08907 L'Hospitalet de Llobregat, Spain, ⁴Dirección de Salud de Guipúzcoa, 20013 San Sebastian, Spain, ⁵CIBER Epidemiología y Salud Pública (CIBERESP), Spain, ⁶Consejería de Sanidad y Consumo, 3008 Murcia, Spain, ⁷Instituto de Salud Pública de Navarra, 31003 Pamplona, Spain, ⁸Escuela Andaluza de Salud Pública, 18080 Granada, Spain and ⁹Consejería de Sanidad y Servicios Sociales de Asturias, 33001 Oviedo, Spain

*To whom correspondence should be addressed. Tel: +34 93 2607401; Fax: +34 93 2607787; Email: a.agudo@iconcologia.net

Aromatic compounds such as polycyclic aromatic hydrocarbons, arylamines and heterocyclic amines require metabolic activation to form metabolites able to bind to DNA, a process mediated by polymorphic enzymes. We measured aromatic DNA adducts in white blood cells by the ³²P-post-labelling assay in a sample of 296 healthy adults (147 men and 149 women) from five regions of Spain. We also analyzed functional polymorphisms in the metabolic genes *CYP1A1*, *CYP1A2*, *EPHX1*, *GSTM1*, *GSTT1*, *NAT2* and *SULT1A1*. A significant increased level of DNA aromatic adducts was found related to the fast oxidation–hydrolysis phenotype defined by the polymorphism *I462V* in *CYP1A1*, the allele *A* in *IVS1–154C>A* of *CYP1A2* and the combination Tyrosine–Arginine for *Y113H* and *H139R* of *EPHX1*. Geometric means (adducts per 10^{–9} normal nucleotides) were 2.17, 4.04 and 6.30 for slow, normal and fast phenotypes, respectively (*P*-trend = 0.01). Slow acetylation by *NAT2* was associated with a significant decrease in adduct level; subjects with slow alleles **5A* and **7A/B* had in average 1.56 × 10^{–9} adducts, as compared with 5.60 for those with normal *NAT2* activity (*P*-value = 0.01). No association was seen with polymorphisms of other metabolic genes such as *GSTM1*, *GSTT1* or *SULT1A1*. We concluded that the metabolic pathways of oxidation, hydrolysis and acetylation are relevant to the formation of bulky DNA adducts. This could suggest a potential involvement of aromatic compounds in the formation of such adducts; however, given lack of specificity of the post-labeling assay, a firm conclusion cannot be drawn.

Introduction

Chemical carcinogenesis assumes that exposure to a genotoxic compound results in damage to the structural integrity of DNA typically involving the covalent binding to DNA, referred to as carcinogen–

DNA adduct formation (1). Studies in animal models have shown that dose-response curves for chemical exposure, DNA adduct formation, mutagenesis and rates of tumour development are closely correlated. This supports the notion that DNA adducts play a significant role in the carcinogenic process, and therefore represent a measure for exposure as well as cancer risk (2). Several techniques have been used to measure DNA adducts in human tissues. The ³²P-post-labelling assay has been widely applied because of its high sensitivity and the requirement for only microgram amounts of DNA. ³²P-post-labelling typically detects a mixture of adducts often referred to as aromatic or bulky adducts (2). Several groups of aromatic compounds may be relevant to human carcinogenesis, mainly the ubiquitous polycyclic aromatic hydrocarbons (PAHs) formed during the incomplete combustion of organic material, but also the aromatic amines (AA) or heterocyclic amines (HA). Environmental sources of PAH include industrial and urban air pollution, tobacco smoke and diet, the main source of exposure in non-smokers who are not exposed through their occupations (1,3). AAs are occupational contaminants mainly in the dyestuff and rubber industry, whereas HA are formed during cooking foods in the presence of creatinine, mostly meat and fish (1,4).

As many human carcinogens, these compounds are chemically inert and metabolic activation is required to form electrophilic DNA-damaging metabolites. Activation of PAH involves the formation of bay-region diol epoxides; the first step of this process is oxidation by cytochrome P450 (CYP) enzymes, mainly *CYP1A1* and *CYP1A2*, to form epoxides. Then hydrolysis by microsomal epoxide hydrolase (*EPHX*) 1 forms intermediate transdihydrodiols, which are further converted by CYP enzymes to reactive diol epoxides. Intermediate metabolites as well as diol epoxides can be detoxified through conjugation with glutathione by glutathione *S*-transferases (GSTs), mainly *GSTM1* and *GSTT1* (1,3). An alternative pathway of PAH activation is conjugation with sulfate of hydroxylated or tetrol derivatives catalyzed by sulphotransferases (*SULTs*) (5). *N*-hydroxylation mediated by *CYP1A2* is the initial step of activation of AA and HA, followed by *O*-esterification catalyzed by *N*-acetyl-transferases (*NATs*) resulting in a highly reactive metabolite by the generation of arylnitrenium ions. *O*-esterification of HA may be achieved as well by the formation of sulfoxy esters catalyzed by *SULTs* (3,4).

In humans, most metabolic enzymes are highly polymorphic; the main single-nucleotide polymorphisms (SNPs) identified in the genes encoding such enzymes as well as their effects on enzyme function are summarized in Table I. Among the SNPs described for *CYP1A1*, three variants (*m1*, *m2* and *m4*) are often reported in Caucasian populations (3). Initially, 1188 bp 3' of STP C>T (*m1*) was associated with the highly inducible phenotype, although no influence on enzyme activity was found in some studies (3,6). The substitution of valine for isoleucine *I462V* (*m2*) results in an ~2-fold increase in the enzyme activity (3,7), whereas the functional significance of *T461N* (*m4*) is uncertain (7). Two mutations associated with inducibility have been described in *CYP1A2*; the allele C of *IVS1–154C>A*, relatively frequent in several ethnic groups, shows decreased caffeine metabolism only among smokers (8), whereas the variant A of *–3859A>G* expressed a low activity enzyme among Japanese smokers (9) but nothing could be concluded among Caucasians owing to its low population frequency (8). The gene *EPHX1* has two known functional polymorphisms (*Y113H* and *H139R*) in the coding region (10); *in vitro* activity of the protein with histidine at codon 113 is reduced ~40%, whereas the form with arginine at codon 139 increased ~25% (11). Three SNPs in *NAT2* (*L161L*, *R197Q* and *G286E*) are used to assign the alleles *NAT2*5*, *NAT2*6* and *NAT2*7*, respectively (12). The absence of any of these substitutions is considered as the wild-type (*NAT2*4*), although it is not the most common allele in Caucasian populations; subjects with any of the above-mentioned substitutions

Abbreviations: AA, aromatic amine; BaP, benzo(a)pyrene; CI, confidence interval; CYP, cytochrome P450; *EPHX*, epoxide hydrolase; GST, glutathione *S*-transferases; HA, heterocyclic amine; *HWE*, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; *NAT*, *N*-acetyl-transferase; PAH, polycyclic aromatic hydrocarbon; SNP, Single-nucleotide polymorphism; *SULT*, sulphotransferase; WBC, white blood cell.

Table I. Polymorphisms in selected metabolic genes and effect on enzyme function

Gene	Polymorphism ^a	Aminoacid change	refSNP ID ^b	Function of the enzyme	Effect on enzymatic activity
<i>CYP1A1</i>	1188 bp 3' of STP C>T (<i>m1</i>)		rs4646903	Oxidation (phase I, activation)	Higher inducibility Increased oxidation (valine) Not clearly established
	Ex7 + 131A>G (<i>m2</i>)	I462V	rs1048943		
	Ex7 + 129C>A (<i>m4</i>)	T461N	rs1799814		
<i>CYP1A2</i>	IVS1 – 154C>A –3859A>G		rs762551 rs2069514	Oxidation (phase I, activation)	Higher inducibility (allele A) Not clearly established
<i>EPHX1</i>	Ex3 – 28T>C	Y113H	rs1051740	Phase II (detoxification), but hydrolysis of PAH-epoxides together with CYP enzymes acts as phase I (activation)	Decreased activity (histidine)
<i>NAT2</i>	Ex4 + 52A>G Ex2 + 487C>T (*5A)	H139R L161L (I114T) ^c	rs2234922 rs1799929	Phase II, detoxification, but acetylation of AH after CYP-oxidation acts as phase I (activation)	Increased activity (arginine) Slow acetylation (threonine)
	Ex2 – 580G>A (*6A) Ex2 – 313G>A (*7A/B)	R197Q G286E	rs1799930 rs1799931		Slow acetylation (glutamine) Slow acetylation (glutamic Acid)
<i>SULT1A1</i>	Ex9 + 44G>A	R213H	rs9282861	Phase II (detoxification), but sulfation of PAH and AH after CYP-oxidation acts as phase I (activation)	Decreased activity (histidine)
<i>GSTM1</i>	Gene deletion	No protein		Phase II, detoxification	No activity
<i>GSTT1</i>	Gene deletion	No protein		Phase II, detoxification	No activity

^aNucleotide numbering according to the SNP500Cancer Database (<http://snp500cancer.nci.nih.gov>).

^bDatabase of SNPs (<http://www.ncbi.nlm.nih.gov/SNP/>).

^cThe silent mutation in Ex2 + 487C>T is found in the same haplotype (*5A) as the T-to-C substitution at Ex2 + 347 with the aminoacid replacement I114T.

are defined as slow acetylators (12,13). Concerning *SULT1A1*, *R213H* is a SNP relatively common among Caucasians; the variant with histidine in position 213 is related to lower activity and lower thermal stability of the enzyme (14). Finally, deletion of both *GSTM1* and *GSTT1* is relatively common in Caucasians; subjects homozygous for a null allele lack enzymatic activity and are assumed to be detoxification deficient (15).

The role of variation in metabolizing enzymes as potential susceptibility factors has been mainly addressed through the association between SNPs in metabolic genes and several tumour sites, whereas not many studies focused on the dependence of DNA adducts on genotype (1,3). We present the results of a study based upon a sample of healthy adults non-selected on the basis of previous exposure, with the aim of analyzing the relationship between the level of aromatic DNA adducts in white blood cell (WBC) and functional SNPs in *CYP1A1*, *CYP1A2*, *EPHX1*, *GSTM1*, *GSTT1*, *NAT2* and *SULT1A1*. Furthermore, we also addressed the dependence of adducts level on the imputed phenotype and metabolic pathways derived from these genotypes.

Materials and methods

Subjects and study design

The study subjects were selected from the Spanish cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC) study (16). The EPIC-Spain cohort included 41 438 healthy volunteers aged 29–69 years recruited between 1992 and 1996 in five regions in Spain: three from the north (Asturias, Guipúzcoa and Navarra) and two from the south (Granada and Murcia). Information on usual diet and lifestyle factors, anthropometric measurements and a 20 ml blood sample were obtained from each subject at baseline. Blood samples were divided into 0.5 ml aliquots of plasma, serum, concentrated red blood cells and buffy coat and stored in liquid nitrogen tanks at –196°C. For the current study, 300 subjects (60 from each center) aged 35–64 years at recruitment were randomly selected by stratified sampling according to the age–sex structure of the Spanish population (<http://www.inec.es/inecmenu/index.htm>). Biological material was not available for four participants, thus the study population is composed of 296 subjects.

DNA extraction and purification

DNA from WBC was extracted from 100–150 µl buffy coats. It was purified by digestion with ribonuclease A, ribonuclease T1 and proteinase K and extracted with saturated phenol, phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitation (17). DNA was then dissolved in distilled water, and its

concentration and purity was measured by a Beckam DU 800 spectrophotometer. The coded DNA was stored at –80°C until laboratory analysis.

DNA adducts determination: ³²P-DNA post-labelling technique

DNA adducts were analyzed using the nuclease P1 modification of the ³²P-DNA post-labelling technique (18,19). Coded DNA samples (1–5 µg) were initially digested to normal and adducted 3'-mononucleotides with micrococcal nuclease (0.46 U) and spleen phosphodiesterase (0.0174 U). After treatment of samples with 5 µg nuclease P1 for 30 min at 37°C, the hydrolyzate enriched in adducted nucleotides were then labeled by incubation with 25–50 µCi carrier-free [γ -³²P]adenosine triphosphate (3000 Ci/mM) and 10 U T4 polynucleotide kinase. Detection of bulky DNA adducts was carried out by multidirectional polyethyleneimine-cellulose thin-layer chromatography. Plates were first developed overnight using 1 M sodium phosphate (pH 6.8) then the resolution of the bulky DNA adducts was achieved with a two-dimensional chromatography (15) using 4 M lithium formate, 7.5 M urea (pH 3.5), 0.65 M lithium chloride, 0.45 M Tris base and 7.7 M urea (pH 8.0). The plates were finally developed overnight using 1.7 M sodium phosphate (pH 5.0). Detection and quantification of ³²P-labeled DNA adducts and total nucleotides were performed by storage phosphor imaging techniques employing imaging screens from Molecular Dynamics (Sunnyvale, CA). After exposure, the screens were immediately scanned using a Typhoon 9210 (Amersham, Buckinghamshire, UK). Data were processed with ImageQuant (v.5.0) from Molecular Dynamics. The levels of DNA adducts were expressed as relative adduct labeling screen response (volume) in adducted nucleotides/screen response (volume) in total nucleotides (20).

All analyses were carried out blindly, and a standard was routinely included in the analysis; the standard was an adducted DNA from livers of mice intraperitoneal treated with 0.5 mg benzo(a)pyrene (BaP) from a previous inter-laboratory trial (19). The average level of adducts in the BaP–DNA mouse liver samples was 8.5 adducts per 10⁸ nucleotides by mass spectrometry and 11.4 adducts per 10⁸ nucleotides by ³²P-post-labelling. Since the assay efficiency can vary between experiments, we have corrected for run-to-run variation. The detection limit was 0.1 adduct per 10⁹ normal nucleotides, as reported (21). A second aliquot of 1–5 µg of DNA was used to carry out a second independent measurement of adducts by the same assay in 41 of 296 subjects; the estimated intraclass correlation coefficient was 0.98 indicating a high reliability of the measurement (22).

Genotyping

SNPs genotyping was performed in a LightCycler™ 2.0 instrument (Roche Diagnostics, Mannheim, Germany) by melting curve analysis of a fluorescently labeled sensor probe specific for each polymerase chain reaction-amplified variant. The analysis of *CYP1A1*, *CYP1A2*, *EPHX1* and *NAT2* polymorphisms was as described previously (23,24). *SULT1A1 R213H* was genotyped with essentially the same methodology using primers and hybridization probes designed and synthesized by TIB-MOLBIOL (Berlin, Germany), which can be provided by the authors upon request.

Analysis of the *GSTM1* and *GSTT1* null alleles was also performed in the LightCycler 2.0 instrument by melting curve analysis of the fragments amplified in a capillary multiplex polymerase chain reaction that contained specific primers for the amplification of *GSTM1*, *GSTT1* and *BCL2* (used as an amplification control) fragments and the LightCycler™ Fast Start DNA Master SYBR Green I kit (Roche Diagnostics). With the exception of the reverse primer for the *GSTT1* fragment that was modified to obtain a shorter amplification product of 257 bp, the primers used to amplify the *GSTM1* (219 bp), *GSTT1* (257 bp) and *BCL2* (154 bp) fragments had been described previously (25).

Genes have been named according to the Human Genome Organisation Gene Nomenclature Committee (<http://www.genenames.org/guidelines.html>). Metabolic gene allele nomenclature follows published recommendations (26). Polymorphisms are identified according to the SNP500Cancer Database (<http://snp500cancer.nci.nih.gov>) and the accession number from the National Center for Biotechnology Information database of SNPs (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

Imputed phenotypes

Polymorphisms in the same or different genes were merged into one single-variable based upon the known effect on enzymatic function and the genotype analysis from our own study (see below). Three metabolic pathways were considered, namely oxidation, hydrolysis and acetylation; these three pathways were used to define five imputed phenotypes. Increased oxidation was assigned to subjects carrying the allele G for *I462V* of *CYP1A1* and allele A for *IVS1-154C>A* in *CYP1A2*; low oxidation was assigned to homozygous AA for *I462V* of *CYP1A1* and homozygous CC for *IVS1-154C>A*; subjects carrying the homozygous AA for *I462V* and allele A for *IVS1-154C>A* were considered to have an intermediate oxidative activity termed as normal oxidation. The imputed phenotype for hydrolysis was defined as follows: low activity for homozygous CC in *Y113H* (histidine) with allele A in *H139R*; high activity for homozygous GG in *H139R* (arginine) with allele T in *Y133H* and normal activity for those carrying the allele T for *Y113H* and allele A for *H139R*. Three subjects carrying the homozygous genotypes CC and GG for *Y113H* and *H139R*, respectively, were excluded from this analysis because its simultaneous effect on the enzymatic activity is unknown. Those with the *NAT2* homozygous genotype CC for **5A* and GG for **7A/B* were considered as having normal *NAT2* activity; moderately slow activity was assigned to subjects with allele T for **5A* keeping the wild-type CC for **7A/B*, and very slow acetylation was assigned to subjects with allele A for **7A/B* whatever the genotype for **5A*. Since CYP-mediated oxidation and hydrolysis by *EPHX1* are supposed to act in the same pathway, an oxidation–hydrolysis phenotype was created by combining both phenotypes as follows: normal, for normal activity of both oxidation and hydrolysis; low, for low activity of both or low activity of one phenotype and normal for the other and high, for high activity of both or high activity for one phenotype and normal for the other; one subject with high oxidation and low hydrolysis and two subjects with low oxidation and high hydrolysis were excluded. In a similar way, *CYP1A2*-mediated oxidation and acetylation supposed to act in the same pathway were combined to form an oxidation–acetylation phenotype. Concerning *CYP1A2*, those carrying the C allele *IVS1-154C>A* were considered as having slow activity, as compared with homozygous AA, considered normal; thus, the oxidation–acetylation phenotype was defined as follows: normal, for normal activity of both phenotypes; moderately slow, for normal oxidation and moderately slow acetylation or slow oxidation and normal acetylation; slow, for slow oxidation and moderately slow acetylation and very slow, for slow oxidation and very slow acetylation; five subjects with normal oxidation and very slow acetylation were excluded.

Statistical analysis

Genotype frequencies were computed for each polymorphism and departure from Hardy–Weinberg equilibrium (HWE) was assessed by means of an exact test; pairwise linkage disequilibrium (LD) was also tested for SNPs within the same gene (27). The level of DNA adducts was expressed as adducted nucleotides per 10^9 normal nucleotides. Negative samples (those below 0.1 adduct per 10^9 normal nucleotides, the threshold of detection of the technique) were assigned a value of 0.05. The distribution of DNA adducts was right-skewed and the variable was transformed using natural logarithms. All the results for adduct levels are expressed as geometric mean with the corresponding 95% confidence intervals (CIs). After log transformation, the data were normalized; the Shapiro–Wilk *W*-test (28) was non-significant ($P = 0.39$) after exclusion of subjects with values below the detection limit. However, the analyses were repeated by means of non-parametric methods using the Kruskal–Wallis test or the Cuzick's extension of the rank-sum test for trend test across ordered categories (28). Since the results with both approaches gave the same conclusions, only those from the parametric analyses are shown.

Association of each SNP with DNA adduct levels was tested by one-way analysis of variance (28), comparing means in the three genotypes (codominant

model) or grouping homozygotes and heterozygotes (dominant or recessive models). Linear regression (28) was used to assess the additive model, assuming a linear relationship between the DNA adducts level (the dependent variable) and genotype, the independent variable denoted by 0 represents the wild-type (common homozygote), 1 represents the heterozygotes and 2 represents the rare homozygote.

Linear regression was used as well to assess the relationship of multiple SNPs and imputed phenotypes with DNA adducts. Multipoint SNPs association analysis (27) was carried out by including several SNPs in the regression model to take into account their joint distribution. However, when strong LD was present, haplotype-based methods were used to try to capture its correlation structure. Haplotype analysis was achieved by estimating haplotype frequencies and comparing the mean DNA adducts level of each haplotype with the most common haplotype, taken as the reference category (29). Regarding imputed phenotypes, the *t*-test was used to compare the mean of high or low enzymatic activity with the mean of the normal phenotype (reference category). A trend test was carried out by assessing the significance of the regression coefficient of a variable coded as 0 for normal, –1 for low and +1 for high activity for oxidation, hydrolysis and oxidation–hydrolysis or 0 for normal, 1 for moderately slow and 2 for very slow acetylation (28). Finally, interaction between two genes or between a gene and an imputed phenotype was assessed by comparing the likelihood of the model with the corresponding interaction terms and the model containing terms for the main effects only. The likelihood-ratio statistic typically has a Chi-square distribution with degrees of freedom equal to the number of additional parameters (27).

Results

Results were available for 296 subjects (147 men and 149 women), aged 35–64 years with mean age of 49 years at the time of recruitment and blood extraction (in both genders). They were residents in three regions from the North of Spain (60%) and two from the South (40%). Overall 74 (25%) of subjects were current smokers, 48 (16%) were former smokers and 174 (59%) had never smoked; furthermore, 132 (38 subjects) had ever been employed in occupations involving potential exposure to PAH. Aromatic DNA adducts were detected in 255 of 296 samples (86%); the geometric mean (95% CI) was 4.04 (3.26, 4.97) adducts per 10^9 normal nucleotides, whereas the arithmetic mean was 8.30 (7.54, 9.05). Further details of lifestyle features of our population and its relationship with adduct levels have been reported elsewhere (Ibáñez,R., Peluso,M., Munnia,A., Piro,S., González,C.A., Amiano,P., Tormo,M.T., Ardanaz,E., Barricarte,A., Berenguer,A., Chirilaque,M.D., Dorronsoro,M., Jakszyn,P., Larrañaga,N., Martínez,C., Navarro,C., Quirós,J.R., Sánchez,M.J. and Agudo,A submitted for publication).

The genotype distribution of the polymorphisms analyzed with the corresponding geometric mean of adducts level is shown in Table II. The genotype frequencies of all but two SNPs (*I462V*, *T461N* in *CYP1A1*) were conformed to Hardy–Weinberg expectations. Both have a proportion of homozygotes higher than expected according to HWE (1% versus 0.23% for *I462V*, 1% versus 0.18% for *T461N*), with a relative deficit of heterozygotes (7.5% versus 9% *m2*, 6.4% versus 8% *m4*). There was remarkable LD between *m1* and *m2* of *CYP1A1* as well as the two SNPs analyzed in *CYP1A2*; the allele **5A* in *NAT2* was also in strong LD both with **6A* and **7A/B*. A significant LD, although less evident, was observed between *I462V* and *T461N* in *CYP1A1* and between *Y113H* and *H139R* in *EPHX1*. Regarding the differences in adduct levels by genotype, only the presence of allele A in **7A/B* of *NAT2* was significantly associated with lower adduct level (geometric mean 1.56 as compared with 4.24 of GG, P -value = 0.04). Other variants seemed to be associated with increased (*I462V*, *H139R*) or decreased (**5A*, *IVS1-154C>A*) adduct levels, but none of them was statistically significant.

The joint effects of SNPs within *CYP1A1* and *NAT2* were assessed by means of haplotype analysis (Table III). Seven haplotypes were found for *CYP1A1*; compared with the wild-type, no differences in adduct levels were seen for haplotypes showing only the variants corresponding to *m1* (haplotype 2) or *m4* (haplotype 4); however, the haplotype 3, with the variant alleles both for *m1* and *m2* had an increased adduct level, although the difference was non-significant (P -value = 0.11). Haplotypes 6 (variant in the three SNPs) and 7 (variant alleles for *m1* and *m4*) were associated with a strong and

Table II. Frequency of SNP polymorphisms in metabolic genes and DNA adducts per 10⁹ normal nucleotides

Gene	Polymorphism (refSNP ID ^a)	Genotype	n (%)	DNA adducts: Geometric mean (95% CI)	P-value ^b	P-trend ^c
CYP1A1	1188 bp 3' of STP C>T (m1) (rs4646903) dominant	TT	233 (79.0)	3.89 (3.06, 4.94)	0.78	0.53
		TC	60 (20.3)	4.43 (2.77, 7.08)		
		CC	2 (0.7)	7.92 (0.60, 104)		
		TT	233 (79.0)	3.89 (3.06, 4.94)		
		TC/CC	62 (21.0)	4.51 (2.84, 7.16)		
		AA	270 (91.5)	3.89 (3.12, 4.85)		
	I462V (m2) Ex7 + 131A>G (rs1048943) dominant	AG	22 (7.5)	7.12 (3.29, 15.4)	0.16	0.63
		GG	3 (1.0)	1.13 (0.14, 9.19)		
		AA	270 (91.5)	3.89 (3.12, 4.85)		
		AG/GG	25 (8.5)	5.72 (2.76, 11.8)		
		CC	273 (92.5)	4.21 (3.38, 5.24)		
		CA	19 (6.4)	2.58 (1.12, 5.93)		
T461N (m4) Ex7 + 129C>A (rs1799814) dominant	AA	3 (1.0)	0.98 (0.12, 7.95)	0.22	0.089	
	CC	273 (92.5)	4.21 (3.38, 5.24)			
	CA/AA	22 (7.5)	2.26 (1.04, 4.89)			
	AA	122 (41.5)	4.33 (3.12, 6.03)			
	AC	138 (46.9)	4.17 (3.06, 5.69)			
	CC	34 (11.6)	2.59 (1.39, 4.83)			
CYP1A2	IVS1 – 154C>A (rs762551) dominant	AA	122 (41.5)	4.33 (3.12, 6.03)	0.34	0.25
		AC	138 (46.9)	4.17 (3.06, 5.69)		
		CC	34 (11.6)	2.59 (1.39, 4.83)		
		AA	122 (41.5)	4.33 (3.12, 6.03)		
		AC/CC	172 (58.5)	3.80 (2.87, 5.01)		
		AA/AC	260 (88.4)	4.25 (3.39, 5.32)		
	–3859A>G (rs2069514) recessive	CC	34 (11.6)	2.59 (1.39, 4.83)	0.14	0.67
		GG	280 (96.2)	4.01 (3.23, 4.98)		
		GA	11 (3.8)	5.09 (1.71, 15.2)		
		AA	—	—		
		TT	164 (55.8)	4.12 (3.10, 5.48)		
		TC	108 (36.7)	4.03 (2.84, 5.72)		
EPHX1	Y113H Ex3 – 28T>C (rs1051740) dominant	CC	22 (7.5)	3.22 (1.48, 7.01)	0.84	0.65
		TT	164 (55.8)	4.12 (3.10, 5.48)		
		TC/CC	130 (44.2)	3.88 (2.82, 5.34)		
		TT/TC	272 (92.5)	4.08 (3.27, 5.09)		
		CC	22 (7.5)	3.22 (1.48, 7.01)		
		AA	178 (61.0)	3.89 (2.96, 5.12)		
	H139R Ex4 + 52A>G (rs2234922) dominant	AG	96 (32.9)	3.66 (2.52, 5.30)	0.20	0.34
		GG	18 (6.2)	8.53 (3.62, 20.1)		
		AA	178 (61.0)	3.89 (2.96, 5.12)		
		AG/GG	114 (39.0)	4.18 (2.97, 5.89)		
		AA/AG	274 (93.8)	3.81 (3.06, 4.74)		
		GG	18 (6.2)	8.53 (3.62, 20.1)		
NAT2	*5A, L161L Ex2 + 487C>T (rs1799929) dominant	CC	79 (27.3)	4.97 (3.29, 7.51)	0.46	0.29
		CT	156 (54.0)	3.66 (2.73, 4.90)		
		TT	54 (18.7)	3.64 (2.21, 5.99)		
		CC	79 (27.3)	4.97 (3.29, 7.51)		
		CT/TT	210 (72.7)	3.65 (2.84, 4.70)		
		CC/CT	235 (81.3)	4.05 (3.19, 5.15)		
	*6A, R197Q Ex2 – 580G>A (rs1799930) dominant	TT	54 (18.7)	3.64 (2.21, 5.99)	0.70	0.76
		GG	128 (44.3)	3.74 (2.70, 5.17)		
		GA	137 (47.4)	4.26 (3.11, 5.83)		
		AA	24 (8.3)	3.69 (1.74, 7.81)		
		GG	128 (44.3)	3.74 (2.70, 5.17)		
		GA/AA	161 (55.7)	4.17 (3.13, 5.57)		
*7A/B, G286E Ex2 – 313G>A (rs1799931) dominant	GG/GA	265 (91.7)	4.00 (3.19, 5.01)	0.84	0.041	
	AA	24 (8.3)	3.69 (1.74, 7.81)			
	GG	280 (94.9)	4.24 (3.41, 5.26)			
	GA	15 (5.1)	1.56 (0.61, 3.97)			
	AA	—	—			
	GG	159 (53.7)	3.75 (2.81, 5.00)			
SULT1A1	R213H Ex9 + 44G>A (rs9282861) dominant	GA	109 (36.8)	4.26 (3.01, 6.04)	0.74	0.44
		AA	28 (9.5)	4.83 (2.43, 9.60)		
		GG	159 (53.7)	3.75 (2.81, 5.00)		
		GA/AA	137 (46.3)	4.37 (3.21, 5.96)		
		GG/GA	268 (90.5)	3.95 (3.16, 4.93)		
		AA	28 (9.5)	4.83 (2.43, 9.60)		
GSTM1	Ex4 + 10 + > – (gene deletion)	Present	151 (52.8)	3.98 (2.97, 5.34)	0.81	
		Null	135 (47.2)	4.19 (3.07, 5.72)		
GSTT1	Ex5 – 49 + > – (gene deletion)	Present	243 (85.0)	3.94 (3.13, 4.97)	0.46	
		Null	43 (15.0)	4.94 (2.85, 8.57)		

LD analysis: *CYP1A1*: 1188 bp3' of STP C>T and *I462V*, $D' = 0.91$ ($P < 0.001$); *I462V* and *T461N*, $D' = 0.12$ ($P = 0.007$); *CYP1A2* IVS1–154C>A and –3859A>G, $D' = 0.99$ ($P = 0.015$); *EPHX1* *Y113H* and *H139R*, $D' = 0.32$ ($P = 0.013$); *NAT2* *5A and *6A, $D' = 0.98$ ($P < 0.001$); *NAT2* *5A and *7A/B, $D' = 0.76$ ($P = 0.007$). All SNPs in HWE except *I462V* ($P = 0.021$) and *T461N* ($P = 0.011$), both in *CYP1A1*.

^aNational Center for Biotechnology Information database of SNPs (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

^bThe P -values correspond to the analysis of variance (analysis of variance, F -test 2df) for the null hypothesis of equality of means over the three genotypes (codominant model) or t -test (1df) for the dominant or the recessive models.

^cThe P -trend corresponds to the significance of the regression coefficient of a variable coded as 0 for the wild-type, 1 for the heterozygous variant and 2 for the homozygous variant (additive model).

Table III. Frequency and DNA adducts per 10⁹ normal nucleotides in haplotypes of genes *CYP1A1* and *NAT2*

Haplotype	Alleles			Frequency (%)	DNA adducts ^a , difference and 95%-CI	P-value
CYP1A1, alleles of m1, m2, m4						
Haplotype 1	T	A	C	85.4	Reference	
Haplotype 2	C	A	C	6.29	0.09 (-0.53, 0.71)	0.78
Haplotype 3	C	G	C	4.05	0.60 (-0.13, 1.34)	0.11
Haplotype 4	T	A	A	3.40	-0.11 (-0.90, 0.69)	0.79
Haplotype 5	T	G	A	0.34	0.68 (-0.59, 1.95)	0.29
Haplotype 6	C	G	A	0.34	-3.20 (-4.47, -1.94)	<0.0001
Haplotype 7	C	A	A	0.16	-4.31 (-4.32, -4.30)	<0.0001
NAT2 ^b , alleles of *5A, *6A, *7A/B						
Haplotype 1	T	G	G	45.2	Reference	
Haplotype 2	C	A	G	31.0	0.14 (-0.24, 0.51)	0.47
Haplotype 3	C	G	G	21.0	0.26 (-0.14, 0.66)	0.21
Haplotype 4	C	G	A	1.55	-0.37 (-1.70, 0.95)	0.58
Haplotype 5	C	A	A	0.75	-0.94 (-2.71, 0.83)	0.30
Haplotype 6	T	A	G	0.24	-4.25 (-7.38, -1.13)	0.008
Haplotype 7	T	G	A	0.23	0.82 (0.82, 4.12)	0.63

^aDifference between the mean DNA adducts level of each haplotype and the mean of the reference category; DNA adducts level expressed as the natural logarithm of the number of adducts per 10⁹ normal nucleotides.

^bBy default, the most common haplotype is taken as the reference category. However, for *NAT2* the haplotype corresponding to the normal activity of the enzyme is haplotype 3 (CGG). The alleles involving slow acetylation are T for *5A and allele A both for *6A and *7A/B.

significant decrease in adduct level, but they are very rare in the population (frequency <1%) and these results must be taken with caution. Regarding *NAT2*, four of seven haplotypes had frequencies >1%; the estimated frequency for the wild-type (haplotype 3) was 21%. Haplotypes 1, 2 and 4 corresponding, respectively, to slow acetylator alleles *5A, *6A and *7A/B had lower levels of adducts as compared with haplotype 3, but none of the differences was statistically significant. The simultaneous effect of several SNPs was also addressed by means of multipoint association analysis (results not shown). Polymorphisms *m1* and *m2* of *CYP1A1* were strongly correlated; actually all subjects with the variant G of *m2* had the variant C for *m1*, but when they were analyzed together in a model only those with the variant *m2* had a higher level of adducts. On the other hand, *m2* and *m4* had some effect, although in opposite direction; since the functional significance of *m4* is uncertain, only *m2* (*I462V*) was considered in the phenotype analysis. Regarding *NAT2*, when the three SNPs were analyzed together, only the allele *7A/B was statistically significant. The dominant model of *5A was associated with lower level of adducts with *P*-value = 0.11, and no effect was observed for *6A (results not shown); thus, only alleles *7A/B and *5A were used to assess acetylation.

The relationship of adduct levels with imputed phenotypes of oxidation, hydrolysis and acetylation is shown in Table IV. There was a clear trend to a higher level of adducts with oxidative activity as well as with EPHX-mediated hydrolysis, although they did not reach significance. However, when they were considered together as oxidation-hydrolysis pathway, a significant trend appeared (*P*-trend = 0.01). A clear trend was observed between decreasing levels of adduct and slower acetylation activity (*P*-trend = 0.018); furthermore, the very slow acetylators (defined by the allele A in *7A/B) have a significantly lower level of adducts as compared with subjects with normal activity (geometric mean 1.56 and 5.60, respectively, *P*-value = 0.016). There was also a trend between lower levels of adducts and slower oxidation-acetylation activity, although the difference between the extreme categories was not significant. An alternative way to deal with acetylation would be the simple summation of slow genotypes, but this would ignore the fact that individual SNPs may induce slow phenotypes by different mechanisms and different degrees of modification of the enzymatic activity (13). Furthermore, it would not take into account the strong LD between SNPs in our population. Only 11 (4%) subjects had the wild-type genotype, 170 (59%) carried one slow SNP and 108 (37%) had two slow SNPs; none of them had the three slow genotypes.

The corresponding geometric mean (95% CI) were 6.31 (2.09, 19.0), 4.31 (3.25, 5.70) and 3.34 (2.34, 4.75) with a *P*-trend = 0.17; none of the pairwise differences were significant. Thus, combining *5A and *7A/B was considered better to account for the acetylation phenotype than the summation approach in our population.

Since the subject's genotype is determined by an apparently random process, the genotypes distribution may be considered unrelated to environmental and behavioral factors related to adduct level. However, we fit models for adduct levels as a function of each phenotype adjusted by sex, age, center, tobacco smoking, occupation, body mass index, energy intake, alcohol and the consumption of cereals, meat, fish, milk and dairy products. Results from the multivariate model were very similar to those observed using the univariate approach. Furthermore, to prevent possible residual confounding due to effect of relevant exposures (i.e. tobacco smoking) that may account for the observed differences, we carried out a separate analysis among the 174 never smokers. The geometric means (95% CI) of adduct levels for the oxidation-hydrolysis pathway were 2.33 (1.44, 4.74) for low, 4.72 (3.46, 6.46) for normal and 5.80 (2.89, 11.6) for high activity (*P*-trend = 0.24). Regarding the oxidation-acetylation phenotype, the adduct levels were 5.76 (2.86, 11.7) for normal, 5.51 (3.70, 8.22) for moderately slow, 3.40 (2.19, 5.29) for slow and 2.83 (0.70, 11.6) for very slow activity (*P*-trend = 0.054). The pattern of adduct levels among non-smokers for the main imputed phenotypes is quite similar to the one seen in the whole sample, suggesting that the observed differences are not driven by exposure due to smoking.

Finally, we explored the effect of *GSTM1* and *GSTT1* activity on the adducts level, according to the oxidation-hydrolysis phenotype in order to assess whether reduced detoxification owing to lacking of the enzyme because a null genotype vary at different levels of activation (Figure 1). The trend for increased adduct level was evident either when *GSTM1* or *GSTT1* were present or deleted; however, the potential detoxifying effect of GSTs varied at different levels of oxidation-hydrolysis activity. For *GSTM1*, the lack of detoxification due to the absence of the enzyme (null genotype of the gene) translated into an increased level of adducts only among subjects with high oxidative-hydrolytic phenotype, while the lower detoxifying effect of *GSTT1* was associated with increased level of adducts essentially among subjects with low oxidation-hydrolysis activity. However, given the small number of subjects the differences were not significant. There were no significant interactions between the oxidative-hydrolytic pathway and *GSTM1* or *GSTT1* activity. We also explored

Table IV. Imputed phenotypes according to polymorphisms in metabolic genes and DNA adducts per 10⁹ normal nucleotides

Imputed phenotype ^a (genes involved)	Enzymatic activity	<i>n</i>	DNA adducts: geometric mean and 95%–CI	<i>P</i> -value ^b	<i>P</i> -trend ^c	DNA adducts: geometric mean and 95% CI (adjusted) ^d	<i>P</i> -value ^b	<i>P</i> -trend ^c
Oxidation (<i>CYP1A1</i> , <i>CYP1A2</i>)	Low	34	2.59 (1.39, 4.83)	0.18	—	2.59 (1.40, 4.80)	0.042	—
	Normal	234	4.10 (3.23, 5.21)	—	0.095	4.10 (3.24, 5.19)	—	0.045
	High	25	5.72 (2.76, 11.8)	0.40	—	5.72 (2.78, 11.7)	0.56	—
Hydrolysis (<i>EPHX1</i>)	Low	19	2.76 (1.19, 6.40)	0.44	—	2.76 (1.19, 6.38)	0.56	—
	Normal	253	3.88 (3.08, 4.89)	—	0.091	3.88 (3.09, 4.89)	—	0.18
	High	15	8.53 (3.31, 22.0)	0.11	—	8.53 (3.32, 21.9)	0.20	—
Oxidation–hydrolysis (<i>CYP1A1</i> , <i>CYP1A2</i> , <i>EPHX1</i>)	Low	42	2.17 (1.23, 3.84)	0.052	—	2.17 (1.24, 3.83)	0.032	—
	Normal	204	4.04 (3.12, 5.22)	—	0.012	4.04 (3.12, 5.22)	—	0.018
	High	35	6.30 (3.38, 11.7)	0.19	—	6.30 (3.39, 11.7)	0.37	—
Acetylation (<i>NAT2</i>)	Normal	70	5.60 (3.62, 8.64)	—	—	5.60 (3.64, 8.61)	—	—
	Slow, moderate	204	3.78 (2.93, 4.88)	0.13	—	3.78 (2.94, 4.87)	0.14	—
	Very slow	15	1.56 (0.61, 3.99)	0.016	0.018	1.56 (0.62, 3.96)	0.012	0.016
Oxidation–acetylation (<i>CYP1A2</i> , <i>NAT2</i>)	Normal	30	5.92 (3.05, 11.5)	—	—	5.92 (3.07, 11.4)	—	—
	Slow, moderate	125	4.55 (3.29, 6.29)	0.48	—	4.55 (3.30, 6.28)	0.59	—
	Slow	117	3.46 (2.47, 4.84)	0.16	—	3.46 (2.48, 4.83)	0.14	—
	Very slow	10	2.20 (0.70, 6.93)	0.14	0.057	2.20 (0.71, 6.86)	0.090	0.031

^aImputed phenotypes—Oxidation: normal activity, homozygous AA for *I462V* of *CYP1A1* and allele A (AA/AC) for *IVS1–154C>A* of *CYP1A2*; low (decreased oxidative activity), homozygous AA for *I462V* of *CYP1A1* and homozygous CC for *IVS1–154C>A* of *CYP1A2* and high (increased oxidative activity), allele G (AG/GG) for *I462V* of *CYP1A1* and allele A (AA/AC) of *IVS1–154C>A* in *CYP1A2*. Hydrolysis (gene *EPHX1*): normal, allele T (TT/TC) for *Y113H* and allele A (AA/AG) for *H139R*; low, homozygous CC (histidine) for *Y113H* and allele A (AA/AG) for *H139R* and high, allele T (TT/TC) for *Y113H* and homozygous GG (arginine) for *H139R*; 3 subjects homozygous CC for *Y113H* and GG for *H139R* were excluded. Oxidation–hydrolysis: normal, normal activity of both phenotypes (oxidation, hydrolysis); low, low activity of both or low activity for one phenotype and normal for the other and high, high activity of both or high activity for one phenotype and normal for the other; one subject with high oxidation and low hydrolysis, and two subjects with low oxidation and high hydrolysis were excluded. Acetylation (*NAT2*): normal, homozygous CC for *5A and GG for *7A/B; moderately slow, allele T (CT/TT) for *5A and GG for *7A/B and very slow, allele A (GA/AA) for *7A/B whatever the genotype for *5A. Oxidation–Acetylation: normal, normal activity of both phenotypes (*CYP1A2*-oxidation, acetylation); moderately slow, normal oxidation and moderately slow acetylation or slow oxidation and normal acetylation; slow, slow oxidation and moderately slow acetylation and very slow, slow oxidation and very slow acetylation; five subjects with normal oxidation and very slow acetylation were excluded.

^bThe *P*-values correspond to the *t*-test comparing the mean of each phenotype for high or low activity with the mean of the normal phenotype (reference category).

^cThe *P*-trend corresponds to the significance of the regression coefficient of a variable coded as 0 for normal, –1 for low and +1 for high activity phenotypes for oxidation, hydrolysis and oxidation–hydrolysis, as 0 for normal, 1 for slow (moderate) and 2 for very slow acetylation or as 0 for normal, 1 for slow (moderate) and 2 for slow, and 3 for very slow oxidation–acetylation.

^dGeometric means (95% CI) adjusted for sex, age, center, tobacco smoking, occupations with potential PAH exposure, body mass index, total caloric intake and the dietary consumption of cereals, meat, fish, milk and dairy products, vegetables, fruit and alcohol.

the combined effect of *SULT1A1* activity with the *CYP1A2*-mediated oxidation or with the acetylator phenotype since they are thought to act in the metabolism of HA; there was no effect on the adduct levels of any of such combinations (results not shown). Gene–gene interaction between *GSTM1* and *GSTT1* was no significant.

Discussion

We found an increased level of DNA aromatic adducts in WBC from healthy adults in Spain related to higher oxidation mediated by CYP enzymes combined with higher activity of microsomal EPHX. In our population, the high oxidative activity was mainly determined by the presence of the *I462V* in *CYP1A1* and the allele A in *IVS1–154C>A* of *CYP1A2*. The hydrolase activity is related to *Y113H* and *H139R* in *EPHX1*; the isoenzyme with histidine in both codons shows lower activity while the protein carrying the combination tyrosine–arginine is the fast isoform. Furthermore, the slow acetylator phenotype determined by alleles *5A and *7A/B in *NAT2* as well as its combination with the allele C in *IVS1–154C>A* of *CYP1A2* to produce a slow oxidation–acetylation phenotype were associated with a significant decrease in the level of adducts. No association was seen with single SNPs of the above-mentioned genes or with other metabolic genes such as *GSTM1*, *GSTT1* or *SULT1A*.

The oxidation–hydrolysis pathway is thought to be relevant in the activation of PAHs, but the role of *CYP1A2* and *NAT2* suggests the potential involvement of other compounds. We measured DNA adducts by ³²P-post-labelling; studies in settings with high PAH exposure found a positive relationship between adducts measured by such method with levels of specific PAH adduct as measured by specific techniques (30). The nuclease P1 version of the ³²P-post-labelling technique is generally used to analyze the formation of adducts in-

duced from PAHs such as BaP. However, this assay is also effective for the detection of DNA adducts induced from different hydroxyaryl-amines; actually a number of DNA adduct standards formed by *N*-hydroxyderivatives of several HA and AA have been analyzed in our laboratory using the nuclease P1 version of the ³²P-post-labelling assay (18,19). On the basis of our results, we could expect that at least a part of DNA adducts detected in leukocytes is produced from aromatic amines that are resistant to the 3'-dephosphorylating action of nuclease P1, i.e. 2-acetylaminofluorene, 2-acetylaminophenanthrene, methyl-4-azoaminophenyl, 2-aminophenanthrene and methyl-4-azoaminobenzene (31). Most of the epidemiological evidence on these chemical carcinogens arises from groups of subjects with high exposure linked to occupation or tobacco smoking, while our study rather reflects the effect of chronic exposure in general population. On the other hand, we measured adducts in DNA from peripheral WBC; even though WBC are not necessarily the most sensitive cells for DNA adduct formation, some results suggest that they might reflect adduct levels in relevant target organs (32). Furthermore, as a measure of exposure, WBC–DNA adducts have the advantage of accounting for all exposure routes, including inhalation, ingestion and dermal absorption (32).

Several studies have analyzed the effect of SNPs of metabolic genes on adducts measured by ³²P-post-labelling in DNA from peripheral WBC, either leukocytes or lymphocytes (33–48). They were carried out in Caucasian populations with subjects not selected on the basis of specific exposure; some used groups occupationally exposed or heavy smokers, but they always included a control (non-exposed) group. *GSTM1* was the gene most commonly analyzed; the null genotype was associated with significantly higher level of adducts in six studies (33–35,41,43,45) of fifteen (33–48), whereas no association was found with the deletion of *GSTT1* in eight studies (33–35,37–41).

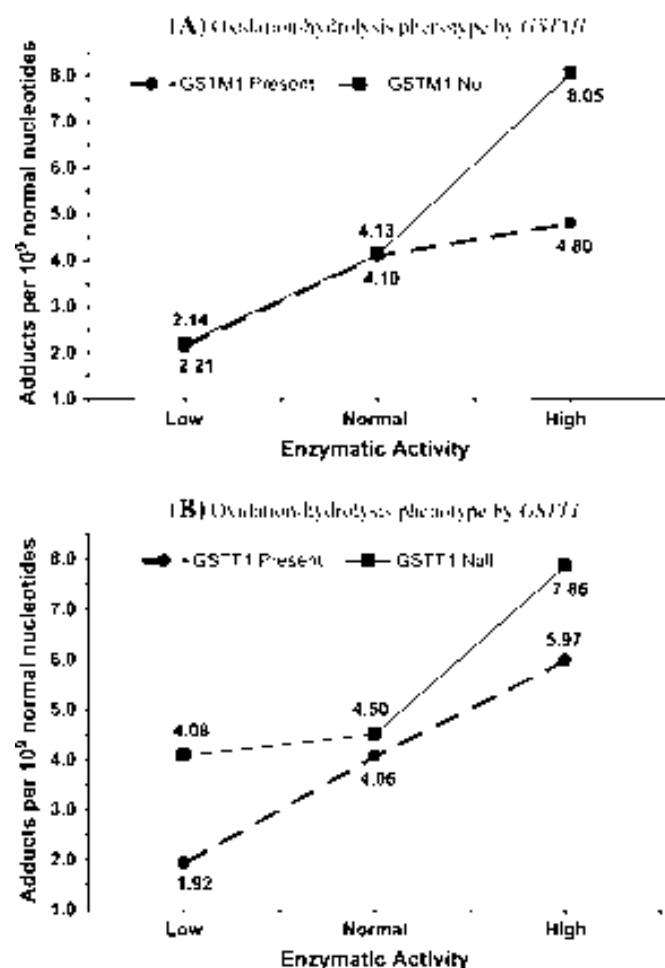


Fig. 1. DNA adducts per 10^9 normal nucleotides (geometric mean) in 273 subjects according to oxidation and hydrolysis phenotype and *GSTM1* (A) or *GSTT1* (B) activity. *GSTM1* and *GSTT1* activity defined, respectively, by the presence or deletion of genes *GSTM1* and *GSTT1*. Oxidative activity defined by allele G for the SNP *I462V* of *CYP1A1* (high activity) or homozygous CC for *IVS1-154C>A* of *CYP1A2* (low activity). Hydrolytic activity defined by two SNPs of *EPHX1*: homozygous CC for *Y113H* (low activity) or homozygous GG for *H139R* (high activity). Oxidation-hydrolysis activity is determined by combination of both phenotypes as follows: normal, normal activity of both; low, low activity of both or low activity for one phenotype and normal for the other; high, high activity of both or high activity for one phenotype and normal for the other. Trend for increasing DNA adducts with oxidation-hydrolysis activity was significant for subjects with *GSTM1* null genotype (P -trend = 0.027) and *GSTT1* present (P -trend < 0.001). Differences in adduct levels by *GSTM1* or *GSTT1* genotype (present versus null) were no significant within each level of oxidation-hydrolysis. No significant interaction (likelihood-ratio test) of oxidation-hydrolysis phenotype with *GSTM1* (P -value = 0.75) or *GSTT1* (P -value = 0.74).

Regarding *CYP1A1*, the variant 1188 bp 3' of STP C>T (*m1*) was associated with increased level of DNA adducts in two studies (34,36) of nine (33,34,36,38–40,43,47,48), while *I462V* was related with increased adduct level in two studies (36,43) and no associated in four (40,42,45,47). No association was found with DNA adducts in the only study looking at the SNP in intron 1 of *CYP1A2* (35), as well as in any of the four studies addressing the two common polymorphisms of *EPHX1* (33–35,37). Slow acetylation assessed by the presence of at least two of the alleles *5A, *6A or *7A was unrelated with adduct level in four studies (33,35,38,46) and positively associated in two (44,45), although the latter did not reach statistical significance (P -value = 0.06). A few studies addressed the combined effect of

SNPs in more than one gene; one study in the Netherlands (35) reported a higher level of adducts for subjects with *GSTM1* null genotype, low activity variant of *Y113H* in *EPHX1* and less efficient *GPX1* activity; the latter is also thought to be involved in PAH detoxification. Another Dutch study (41) found a significant increased adducts level for subjects lacking *GSTT1* and slow acetylation (*NAT2*) as compared with fast acetylators with *GSTT1*. A Greek study reported an increased level of adducts for subjects carrying mutation *m1* in *CYP1A1*, fast genotype for *H139R* in *EPHX1* and null *GSTM1* genotype (37). Finally, one study in Poland (42) addressed the combined effect of *I462V* in *CYP1A1* and deletion of *GSTM1* but no association was found. However, studies that used more specific methods such as high-performance liquid chromatography with fluorescence detection consistently showed that subjects with either *m1* or *m2* in *CYP1A1* and *GSTM1* null have a higher level BaP diol-epoxide adducts (3).

Our results tend to agree with previous works regarding the potential role of 1188 bp 3' of STP C>T and mainly *I462V* as modulators of oxidative activity of *CYP1A1* and subsequent increased adduct formation. Regarding *CYP1A2*, although the SNP in intron 1 has shown to affect inducibility among smokers, most studies have been unable to decipher an effect in the general population (49). Thus, our results concerning intron 1 polymorphism of *CYP1A2* and adduct levels should be taken with caution. On the other hand, previous studies seem to point to a prominent role of *GSTM1* not observed in our population. Although GST enzymes are thought to be involved mainly in detoxification, recent evidence points to the ability of these enzymes to produce reactive glutathione conjugates of some chemicals such as dichloroethane and trichloroethylene (15). Furthermore, the activity of some enzymes does not only depend on the genotype but may change as a consequence of induction by chemicals or by other enzymes. A high inducibility of *CYP1A1* was found associated with *GSTM1* null genotype, whereas the presence of at least one *GSTM1* allele was related to only low levels of induction (50); this could explain the effect on adduct formation at high levels of oxidation-hydrolysis pathway among our *GSTM1* null subjects (Figure 1). In contrast, the null genotype of *GSTT1* was associated with higher adduct level mainly among subjects with low oxidative-hydrolysis phenotype. Regarding *NAT2*, an Italian study (44) reported a strong association between adduct formation and slow acetylation, while our slow acetylators showed decreased adduct levels. The Italian study was mainly based upon subjects with bladder cancer, and thus they might have a relatively high exposure to aromatic amines. It is well known that *NAT* activity may be substrate-specific; while *N*-acetylation has detoxifying effect for AA, *O*-acetylation is seen as an important step in the activation of HA (12,13). It could be that in our population dietary exposure to HA be more relevant than exposure to arylamines.

The distribution of genotype polymorphisms *I462V* and *T461N* in *CYP1A1* were different from the expected according to allelic frequency under HWE. Deviations from HWE can be due to mixture of individuals from heterogenous background; all subjects in our study were recruited in Spain, almost all of them Caucasians, so that our population may be considered homogenous. Furthermore, the observed prevalences of most genotypes were similar to the frequencies reported in other Caucasian populations (51). Hardy-Weinberg disequilibrium could also result from genotyping errors; independent genotyping was carried out in a subsample of 25–30 subjects as quality control. There was 100% agreement in the two SNPs lacking HWE, namely *I462V* and *T461N* in *CYP1A1*, while non-concordant results were observed in one subject for *6A in *NAT2*, one subject for *IVS1-154C>A* in *CYP1A2* and another for *Y113H* of *EPHX1*. When HWE is primarily tested as a data quality check only loci that deviate from HWE at significance level $\alpha < 0.001$ or 0.0001 are usually discarded (29). Therefore, since deviation in our study was rather modest, we decided to include both SNPs in the analysis.

Further limitations in the interpretation of our results refer to the actual composition of bulky adducts and the complex metabolism of chemicals involved in adducts formation. The ^{32}P -post-labelling is unable to determine the structure of the labeled adducts, but even

when a specific compounds is identified there is limited knowledge of their precise mechanisms of activation, and often not all relevant genes involved in the metabolic pathway are included. For instance, other important polymorphic metabolic genes such as nicotinamide quinone oxidoreductase 1 (NQO1) or other genes from the GST family such as GSTP1 and GSTA1 (15,52) could probably lead to the formation of adducts. Finally, the amount of DNA adducts are not only the result of exposure but also of unrepaired DNA damage by chemical carcinogens (2). The normal function of DNA repair enzymes plays a critical role, and SNPs have been found in nearly all human DNA repair genes. A positive relationship has been reported between the level of adducts in WBC and risk alleles in genes involved in the repair of bulky DNA adducts (53).

In conclusion, a pathway defined by the enhanced oxidation related to SNPs *I462V* in *CYP1A1* and *IVS1-154C>A* in *CYP1A2*, and high hydrolase activity of *EPHX1* due to fast variants *Y113H* and *H139R*, is associated with increased DNA adducts in WBC of healthy Spanish adults. On the other hand, decreased adduct level was associated both with slow acetylation by *NAT2* related to alleles **5A* and **7A/B* and its combination with slow *CYP1A2* oxidation related to *IVS1-154C>A*. This suggests a potential role of PAH, but also of other aromatic compounds such as HA, in the formation of these adducts. Identification of specific adducts as well as the inclusion of enzymes involved in other metabolic pathways and DNA repair would be needed to better understand the interplay between susceptibility and resistance genes involved in modulating human responses to carcinogenic compounds.

Funding

European Commission (SO 97 200302 05F02); Spanish Ministry of Health (Exp. 02-0652); Red Temática de Investigación Cooperativa de Centros de Cáncer (C03/10); Regional Governments; International Agency for Research on Cancer (AEP/93/02); Institut d'Investigació Biomèdica de Bellvitge (R.I.); Environmental Cancer Risk, Nutrition and Individual Susceptibility, a Network of Excellence of the 6th EU Framework Program (FP6, FOOD-CT-2005-513 943 to A.A., C.A.G., N.S., G.C.).

Acknowledgements

The authors wish to thank Nadia García (Translational Research Laboratory, ICO) for her valuable contribution to genotyping. We also thank the personnel of TIB-MOBIOL for the design of the LightCycler PCR primers and hybridization probes.

Conflict of Interest Statement: None declared.

References

- Luch,A. (2005) Nature and nurture—lessons from chemical carcinogenesis. *Nat. Rev. Cancer*, **5**, 113–125.
- Poirier,M.C. (2004) Chemical-induced DNA damage and human cancer risk. *Nat. Rev. Cancer*, **4**, 630–637.
- Bartsch,H. *et al.* (2000) Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 3–28.
- Gooderham,N.J. *et al.* (2001) Food-derived heterocyclic amine mutagens: variable metabolism and significance to humans. *Drug Metab. Dispos.*, **29**, 529–534.
- Williams,J.A. (2001) Single nucleotide polymorphisms, metabolic activation and environmental carcinogenesis: why molecular epidemiologists should think about enzyme expression. *Carcinogenesis*, **22**, 209–214.
- Singh,V. *et al.* (2007) A study on the association of cytochrome-P450 1A1 polymorphism and breast cancer risk in north Indian women. *Breast Cancer Res. Treat.*, **101**, 73–81.
- Schwarz,D. *et al.* (2001) Differential metabolism of benzo[a]pyrene and benzo[a]pyrene-7,8-dihydrodiol by human CYP1A1 variants. *Carcinogenesis*, **22**, 453–459.
- Sachse,C. *et al.* (2003) Colorectal Cancer Study Group. Polymorphisms in the cytochrome P450 CYP1A2 gene (CYP1A2) in colorectal cancer patients and controls: allele frequencies, linkage disequilibrium and influence on caffeine metabolism. *Br. J. Clin. Pharmacol.*, **55**, 68–76.
- Nakajima,M. *et al.* (1999) Genetic polymorphism in the 5'-flanking region of human CYP1A2 gene: effect on the CYP1A2 inducibility in humans. *J. Biochem.*, **125**, 803–808.
- Hassett,C. *et al.* (1994) Human microsomal epoxide hydrolase: genetic polymorphism and functional expression *in vitro* of amino acid variants. *Hum. Mol. Genet.*, **3**, 421–428.
- Omicinski,C.J. *et al.* (2000) Epoxide hydrolase—polymorphism and role in toxicology. *Toxicol. Lett.*, **112–113**, 365–370.
- Hein,D.W. *et al.* (2000) Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 29–42.
- Zang,Y. *et al.* (2007) Functional characterization of single-nucleotide polymorphisms and haplotypes of human N-acetyltransferase 2. *Carcinogenesis*, **28**, 1665–1671.
- Raftogianis,R.B. *et al.* (1997) Phenol sulfotransferase pharmacogenetics in humans: association of common SUL1A1 alleles with TS PST phenotype. *Biochem. Biophys. Res. Commun.*, **239**, 298–304.
- Hayes,J.D. (2005) Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.*, **45**, 51–88.
- Bingham,S. *et al.* (2004) Diet and cancer—the European Prospective Investigation into Cancer and Nutrition. *Nat. Rev. Cancer*, **4**, 206–215.
- Reddy,M.V. *et al.* (1990) A comparison of DNA adduct formation in white blood cells and internal organs of mice exposed to benzo[a]pyrene, dibenzo[c,g]carbazole, safrole and cigarette smoke condensate. *Mutat. Res.*, **241**, 37–48.
- Peluso,M. *et al.* (1991) ³²P post-labelling analysis of urinary mutagens from smokers of black tobacco implicates 2- amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhP) as a major DNA-damaging agent. *Carcinogenesis*, **12**, 713–717.
- Phillips,D.H. *et al.* (1999) Standardisation and validation of DNA adduct post-labelling methods: report of interlaboratory trials and production of recommended protocols. *Mutagenesis*, **14**, 301–315.
- Reichert,W.L. *et al.* (1992) Storage phosphor imaging technique for detection and quantitation of DNA adducts measured by the ³²P-post-labelling assay. *Carcinogenesis*, **13**, 1475–1479.
- Beach,A.C. *et al.* (1992) Human biomonitoring and the ³²P-post-labelling assay. *Carcinogenesis*, **13**, 1053–1074.
- Ibañez,R. *et al.* (2005) Reliability of bulky DNA adducts measurement by the nuclease P1 ³²P-post-labelling technique. *Biomarkers*, **10**, 1–9.
- Agudo,A. *et al.* (2006) Polymorphisms in metabolic genes related to tobacco smoke and the risk of gastric cancer in the European prospective investigation into cancer and nutrition. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 2427–2434.
- Agudo,A. *et al.* (2006) No association between polymorphisms in CYP2E1, GSTM1, NAT1, NAT2 and the risk of gastric adenocarcinoma in the European prospective investigation into cancer and nutrition. *Cancer Epidemiol. Biomarkers Prev.*, **5**, 1043–1045.
- Voso,M.T. *et al.* (2002) Negative prognostic value of glutathione S-transferase (GSTM1 and GSTT1) deletions in adult acute myeloid leukemia. *Blood*, **100**, 2703–2707.
- Garte,S. *et al.* (2001) Metabolic gene allele nomenclature. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 1305–1306.
- Thomas,D.C. (2004) *Statistical Methods in Genetic Epidemiology*. Oxford University Press, New York, NY.
- Altman,D.G. (1991) *Practical Statistics for Medical Research*, 1st edn. Chapman and hall, London, UK.
- Solé,X. *et al.* (2006) SNPStats: a web tool for the analysis of association studies. *Bioinformatics*, **22**, 1928–1929.
- Boysen,G. *et al.* (2003) Analysis of DNA and protein adducts of benzo[a]pyrene in human tissues using structure-specific methods. *Mutat. Res.*, **543**, 17–30.
- Gupta,R.C. *et al.* (1988) ³²P-adduct assay: comparative recoveries of structurally diverse DNA adducts in the various enhancement procedures. *Carcinogenesis*, **9**, 1687–1693.
- Godschalk,R.W.L. *et al.* (2003) A critical evaluation of DNA adducts as biological markers for human exposure to polycyclic aromatic compounds. *J. Biochem. Mol. Biol.*, **36**, 1–11.
- Garte,S. *et al.* (2007) Effects of metabolic genotypes on intermediary biomarkers in subjects exposed to PAHS: results from the EXPAH study. *Mutat. Res.*, **620**, 7–15.
- Binkova,B. *et al.* (2007) PAH-DNA adducts in environmentally exposed population in relation to metabolic and DNA repair gene polymorphisms. *Mutat. Res.*, **620**, 49–61.
- Ketelslegers,H.B. *et al.* (2006) Interindividual variations in DNA adduct levels assessed by analysis of multiple genetic polymorphisms in smokers. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 624–629.

36. Georgiadis, P. *et al.* (2005) Interactions between CYP1A1 polymorphisms and exposure to environmental tobacco smoke in the modulation of lymphocyte bulky DNA adducts and chromosomal aberrations. *Carcinogenesis*, **26**, 93–101.
37. Georgiadis, P. *et al.* (2004) Impact of phase I or phase II enzyme polymorphisms on lymphocyte DNA adducts in subjects exposed to urban air pollution and environmental tobacco smoke. *Toxicol. Lett.*, **149**, 269–280.
38. Palli, D. *et al.* (2004) The effects of diet on DNA bulky adduct levels are strongly modified by GSTM1 genotype: a study on 634 subjects. *Carcinogenesis*, **25**, 577–584.
39. Peluso, M. *et al.* (2004) Comparison of DNA adduct levels in nasal mucosa, lymphocytes and bronchial mucosa of cigarette smokers and interaction with metabolic gene polymorphisms. *Carcinogenesis*, **25**, 2459–2465.
40. Lee, J. *et al.* (2002) Influence of GSTM1 genotype on association between aromatic DNA adducts and urinary PAH metabolites in incineration workers. *Mutat. Res.*, **514**, 213–221.
41. Godschalk, R.W. *et al.* (2001) Modulation of DNA and protein adducts in smokers by genetic polymorphisms in GSTM1, GSTT1, NAT1 and NAT2. *Pharmacogenetics*, **11**, 389–398.
42. Grzybowska, E. *et al.* (2000) The effect of the genetic polymorphisms of CYP1A1, CYP2D6, GSTM1 and GSTP1 on aromatic DNA adduct levels in the population of healthy women. *Mutat. Res.*, **469**, 271–277.
43. Brescia, G. *et al.* (1999) The influence of cytochrome P450 1A1 and glutathione S-transferase M1 genotypes on biomarker levels in coke-oven workers. *Arch. Toxicol.*, **73**, 431–439.
44. Peluso, M. *et al.* (1998) White blood cell DNA adducts, smoking, and NAT2 and GSTM1 genotypes in bladder cancer: a case-control study. *Cancer Epidemiol. Biomarkers Prev.*, **7**, 341–346.
45. Butkiewicz, D. *et al.* (1998) Modulation of DNA adduct levels in human mononuclear white blood cells and granulocytes by CYP1A1 CYP2D6 and GSTM1 genetic polymorphisms. *Mutat. Res.*, **415**, 97–108.
46. Binkova, B. *et al.* (1998) Coke oven workers study: the effect of exposure and GSTM1 and NAT2 genotypes on DNA adduct levels in white blood cells and lymphocytes as determined by 32P-postlabelling. *Mutat. Res.*, **416**, 67–84.
47. Hemminki, K. *et al.* (1997) Aromatic DNA adducts in foundry workers in relation to exposure, life style and CYP1A1 and glutathione transferase M1 genotype. *Carcinogenesis*, **18**, 345–350.
48. Ichiba, M. *et al.* (1994) Aromatic DNA adducts, micronuclei and genetic polymorphism for CYP1A1 and GST1 in chimney sweeps. *Carcinogenesis*, **15**, 1347–1352.
49. Jiang, Z. *et al.* (2006) Search for an association between the human CYP1A2 genotype and CYP1A2 metabolic phenotype. *Pharmacogenet. Genomics*, **16**, 359–367.
50. Vaury, C. *et al.* (1995) Human glutathione S-transferase M1 null genotype is associated with a high inducibility of cytochrome P450 1A1 gene transcription. *Cancer Res.*, **55**, 5520–5523.
51. Garte, S. *et al.* (2000) Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 1239–1248.
52. Autrup, H. (2000) Genetic polymorphisms in human xenobiotica metabolizing enzymes as susceptibility factors in toxic response. *Mutat. Res.*, **464**, 65–76.
53. Matullo, G. *et al.* (2003) Combination of DNA repair gene single nucleotide polymorphisms and increased levels of DNA adducts in a population-based study. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 674–677.

Received December 28, 2007; revised May 27, 2008; accepted May 29, 2008