



UNIVERSITAT DE  
BARCELONA

## Susceptibilidad genética al cáncer gástrico y lesiones precursoras, y mecanismos moleculares que intervienen en la progresión de la metaplasia intestinal

Osmel Companioni Nápoles



Aquesta tesi doctoral està subjecta a la llicència Reconeixement- NoComercial – SenseObraDerivada 3.0. Espanya de Creative Commons.

Esta tesis doctoral está sujeta a la licencia Reconocimiento - NoComercial – SinObraDerivada 3.0. España de Creative Commons.

This doctoral thesis is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0. Spain License.



**UNIVERSIDAD DE BARCELONA**  
**Facultad de Medicina**  
**Programa de Doctorado en Biomedicina**

**SUSCEPTIBILIDAD GENÉTICA AL CÁNCER GÁSTRICO Y  
LESIONES PRECURSORAS, Y MECANISMOS  
MOLECULARES QUE INTERVIENEN EN LA PROGRESIÓN  
DE LA METAPLASIA INTESTINAL**

Memoria presentada por Osmel Companioni Nápoles para optar al grado de Doctor por la Universidad de Barcelona.

Este trabajo se ha realizado bajo la dirección de los Doctores Carlos Alberto González y Núria Sala Serra, en el marco del Programa de Investigación en Epidemiología del Cáncer de la Unidad de Nutrición, Ambiente y Cáncer (UNAC) y el Laboratorio de Investigación Traslacional 1 (LRT1) del Instituto Catalán de Oncología (ICO) y el Instituto de Investigación Biomédica de Bellvitge (IDIBELL).

**Autor:** Osmel Companioni Nápoles

**Directores:** Dr. Carlos Alberto González

Dra. Núria Sala Serra



*A mis padres, hija y amigos*

*A Mayelín*



# AGRADECIMIENTOS

A Carlos Alberto González y Nuria Sala Serra por la dirección y el apoyo en la realización del presente trabajo. A Nadia García por su intensa implicación, amistad y estrecha interacción en el trabajo experimental, organización y búsquedas de muestras. A Dori y Nuria Ros con las que compartí muchos intensos momentos de ocio e interacción personal. A otros compañeros de la UNAC y del LRT1 como Ville, Noemie, Omar, Ángela, Silvia San José, Eric, Mireia, Raúl, Xavier Bosch, Ión, Natalia, Gemma, Gorka, Miriam, Gabi, Mar y Elisa.

A Miguel Sanz Anquela y Marisa Pardo por la búsqueda y caracterización histológica de biopsias gástricas de metaplasia intestinal en los hospitales de La Princesa, Gregorio Marañón y Hospital General Virgen del Mirón de Soria. A Victoria Andreu por su implicación en la búsqueda de pacientes del proyecto FIS.

A Juanita Merchant (*Molecular and Integrative Physiology Department* de la Universidad de Michigan) por los trabajos de Schlafen5 en líneas celulares, lesiones precursoras y cáncer gástrico.

A Catalina Bonet por los análisis y asesoramiento estadístico de los estudios de asociación de cáncer gástrico y lesiones precursoras gástricas, Schlafen5 y la elaboración de base de datos del estudio de seguimiento de lesiones precursoras gástricas; así como a Leila Lujan-Barroso que también colaboró en la misma.

A Lara Nonell y Eulalia Puigdecanet por el procesamiento de las muestras y análisis estadísticos del microarray de expresión.

A mi madre que tanto contribuyó a mi educación y alentó el espíritu de conocimiento en mí. A mi padre, hija, hermanos, resto de la familia y amigos que siempre tengo presentes por razones esenciales. A Mayelín que ha compartido conmigo un tiempo muy bonito y ha sido muy paciente y comprensiva.



# ÍNDICE

<b>ÍNDICE DE FIGURAS</b> .....	<b>1</b>
<b>ÍNDICE DE TABLAS</b> .....	<b>3</b>
<b>ABREVIATURAS Y SÍMBOLOS</b> .....	<b>5</b>
<b>PRESENTACIÓN</b> .....	<b>9</b>
<b>INTRODUCCIÓN</b> .....	<b>13</b>
1. CÁNCER GÁSTRICO. DESCRIPCIÓN GENERAL Y EPIDEMIOLOGÍA.....	15
2. ANATOMÍA E HISTOLOGÍA DEL ESTÓMAGO.....	16
3. TIPOS ANATÓMICOS E HISTOLÓGICOS DEL CÁNCER GÁSTRICO.....	19
4. <i>HELICOBACTER PYLORI</i> COMO FACTOR DE RIESGO DEL CÁNCER GÁSTRICO.....	20
4.1. <i>Infección por Helicobacter pylori. Aspectos epidemiológicos</i> .....	20
4.2. <i>Infección por Helicobacter pylori. Factores de virulencia</i> .....	21
4.3. <i>Infección por Helicobacter pylori. Mecanismos y consecuencias en el huésped</i> .....	23
4.4. <i>Infección por Helicobacter pylori. Respuesta inmunitaria</i> .....	26
5. OTROS FACTORES DE RIESGO DEL CÁNCER GÁSTRICO.....	26
5.1. <i>Dieta</i> .....	26
5.2. <i>Tabaquismo</i> .....	27
5.3. <i>Anti-inflamatorios no esteroideos</i> .....	27
6. MECANISMO DE CARCINOGENÉISIS GÁSTRICA.....	27
6.1. <i>Alteraciones moleculares en el cáncer gástrico</i> .....	29
7. LESIONES PRECURSORAS GÁSTRICAS.....	31
7.1. <i>Gastritis crónica no atrófica (superficial o activa)</i> .....	31
7.2. <i>Gastritis atrófica multifocal sin metaplasia intestinal</i> .....	32
7.3. <i>Metaplasia intestinal</i> .....	33
7.4. <i>Displasia</i> .....	35
7.5. <i>Epidemiología de las lesiones precursoras gástricas</i> .....	36
7.6. <i>Estudios de seguimiento de lesiones precursoras gástricas</i> .....	36
8. METAPLASIA INTESTINAL.....	39
8.1. <i>Riesgo de cáncer gástrico según subtipo histológico, localización anatómica y extensión de la metaplasia intestinal</i> .....	39
8.2. <i>Alteraciones moleculares en la metaplasia intestinal</i> .....	40
8.2.1. Factores de transcripción tipo caudal homeobox <i>CDX1</i> y <i>CDX2</i> .....	40
8.2.2. Daño oxidativo.....	41
8.2.3. Supresores tumorales <i>RUNX3</i> y <i>TP53</i> .....	41
8.2.4. Vía de señalización celular Sonic hedgehog.....	41
8.2.5. Mecanismos epigenéticos.....	42
8.2.6. Mecanismos de reparación del ADN. Pérdida de heterocigosidad e inestabilidad de microsatélites.....	42
8.2.7. Aumento de actividad telomerasa.....	42
8.3. <i>Estudios de expresión génica por microarrays en metaplasia intestinal</i> .....	42
8.4. <i>La familia multigénica Schlafen</i> .....	44
8.5. <i>Rol de los genes Schlafen en la metaplasia intestinal</i> .....	45
9. SUSCEPTIBILIDAD GENÉTICA.....	46
9.1. <i>Proyecto de mapeo de haplotipos humanos HapMap y estudios de asociación genética</i> ....	47
9.2. <i>Estudios de asociación de genes candidatos en cáncer gástrico</i> .....	48
9.2.1. Citoquinas pro-inflamatorias.....	49
9.2.2. Reconocimiento de <i>Helicobacter pylori</i> y respuesta inmune innata.....	49



9.3. Estudios de asociación de genoma completo en cáncer gástrico .....	50
9.4 Estudios de asociación en lesiones precursoras gástricas .....	53
9.4.1. Citoquinas pro-inflamatorias .....	53
9.4.2. Reconocimiento de <i>Helicobacter pylori</i> y respuesta inmune innata .....	53
9.4.3. Daño y reparación del ADN.....	54
9.4.5. Metabolismo de xenobióticos .....	54
9.4.6. Funcionalidad y protección de la mucosa gástrica .....	54
<b>HIPÓTESIS .....</b>	<b>61</b>
<b>OBJETIVOS .....</b>	<b>65</b>
<b>MATERIALES Y MÉTODOS .....</b>	<b>69</b>
1. ESTUDIO DE ASOCIACIÓN DE GENES DE SEÑALIZACIÓN DE <i>HELICOBACTER PYLORI</i> Y RIESGO DE CÁNCER GÁSTRICO .....	71
1.1. Pacientes y muestras .....	71
1.2. Selección de tagSNPs en genes de la vía de señalización de <i>Helicobacter pylori</i> .....	71
1.3. Extracción de ADN y genotipado .....	72
1.4. Infección por <i>Helicobacter pylori</i> .....	73
1.5. Análisis estadísticos y bioinformáticos .....	73
2. ESTUDIO DE ASOCIACIÓN GENÉTICA EN EL SEGUIMIENTO DE LESIONES PRECURSORAS GÁSTRICAS .....	74
2.1. Pacientes y muestras .....	74
2.2. Histopatología .....	75
2.3. Infección por <i>Helicobacter pylori</i> .....	75
2.4. Prueba piloto de colecta, conservación y extracción de ADN .....	76
2.4.1. Medición de cantidad y calidad del ADN .....	77
2.5. Diseño de solución estabilizadora de saliva y prueba de conservación .....	78
2.5.1. Protocolo de extracción de ADN a partir de saliva .....	78
2.5.2. Prueba piloto de genotipado de ADN de saliva .....	79
2.5.3. Colecta de saliva en hospitales .....	79
2.6. Extracción de ADN de tejido gástrico parafinado .....	80
2.7. Extracción de ADN de sangre.....	81
2.8. Selección de genes candidatos.....	81
2.9. Selección de tagSNPs en los genes candidatos y análisis bioinformáticos.....	86
2.10. Genotipado de SNPs.....	86
2.11. Análisis estadísticos de los resultados de genotipado de SNPs.....	86
3. ESTUDIO DE EXPRESIÓN GÉNICA POR MICROARRAY EN METAPLASIA INTESTINAL QUE PROGRESA A CÁNCER GÁSTRICO .	94
3.1. Pacientes y muestras .....	94
3.2. Corte de bloques de parafina y evaluación histológica.....	98
3.3. Extracción de ARN total .....	98
3.4. Medición de cantidad y calidad del ARN.....	99
3.5. Prueba piloto de comparación de microarrays Human Gene 2.0 ST vs Almac Xcel .....	100
3.6. Análisis de calidad y estadísticos .....	100
3.7. Análisis de enriquecimiento funcional por IPA y GSEA.....	101
3.8. Validación de expresión génica por qRT-PCR.....	103
3.8.1. Protocolo de qRT-PCR por sondas UPL y cebadores en plataforma Biomark .....	109
3.8.2. Procesamiento de los resultados de qRT-PCR por sondas UPL y cebadores en plataforma Biomark .....	110
4. ESTUDIO DE EXPRESIÓN GÉNICA E INMUNO-HISTOQUÍMICA DE SCHLAFEN5 EN LESIONES PRECURSORAS Y CÁNCER GÁSTRICO.....	112
4.1. Tissue arrays comerciales de tejido gástrico.....	112
4.2. Pacientes y muestras analizadas .....	112
4.3. Histopatología .....	113

4.4. Infección por <i>Helicobacter pylori</i> .....	113
4.5. Inducción de la expresión de <i>Schlafen5</i> por <i>IFN<math>\alpha</math></i> .....	113
4.6. Silenciamiento de <i>Schlafen5</i> y <i>cytospin</i> en células Jurkat .....	114
4.7. Detección inmunohistoquímica de <i>Schlafen5</i> .....	114
4.8. Interpretación de la tinción inmunohistoquímica .....	115
4.9. Análisis estadísticos.....	116
4.10. Doble tinción inmunohistoquímica para <i>Schlafen5</i> y <i>CD2</i> , <i>CD20</i> o <i>Mac2</i> .....	116
<b>RESULTADOS .....</b>	<b>119</b>
1. ESTUDIO DE ASOCIACIÓN DE GENES DE SEÑALIZACIÓN DE <i>HELICOBACTER PYLORI</i> Y RIESGO DE CÁNCER GÁSTRICO .....	121
1.1. Análisis estadísticos de asociación genética .....	121
1.2. Análisis bioinformáticos de efectos funcionales.....	123
2. ESTUDIO DE ASOCIACIÓN GENÉTICA EN EL SEGUIMIENTO DE LESIONES PRECURSORAS GÁSTRICAS .....	128
2.1. Prueba piloto de colecta, conservación y extracción de ADN.....	128
2.2. Diseño de solución estabilizadora de saliva y prueba de conservación .....	130
2.3. Genotipado de ADN de saliva. Prueba piloto .....	131
2.4. Genotipado de ADN en plataforma <i>Sequenom</i> .....	133
2.5. Análisis estadísticos de asociación genética .....	134
2.6 Análisis bioinformáticos de efectos funcionales.....	143
3. ESTUDIO DE EXPRESIÓN GÉNICA POR MICROARRAY EN METAPLASIA INTESTINAL QUE PROGRESA A CÁNCER GÁSTRICO	144
3.1. Medición de cantidad y calidad del ARN .....	144
3.2. Prueba piloto de comparación de <i>microarrays Human Gene 2.0 ST versus Almac Xcel</i> .....	144
3.3. Análisis estadísticos de los resultados del <i>microarray</i> .....	148
3.4. Genes diferencialmente expresados.....	150
3.5. Análisis de enriquecimiento funcional.....	159
3.5.1. <i>Gene Set Enrichment Analysis</i> .....	159
3.5.2. <i>Ingenuity Pathway Analysis</i> .....	171
3.5.3. Análisis de ontología génica .....	176
3.6. Validación de la expresión génica por <i>qRT-PCR</i> .....	180
4. ESTUDIO DE EXPRESIÓN GÉNICA E INMUNO-HISTOQUÍMICA DE <i>SCHLAFEN5</i> EN LESIONES PRECURSORAS Y CÁNCER GÁSTRICO .....	183
4.1. Inducción de la expresión de <i>Schlafen5</i> por <i>IFN<math>\alpha</math></i> y <i>Helicobacter pylori</i> .....	183
4.2. Expresión de <i>Schlafen5</i> en <i>tissue arrays</i> comerciales de tejido gástrico .....	185
4.3. Controles de la tinción inmuno-histoquímica de <i>Schlafen5</i> .....	186
4.4. Resultados cualitativos de la tinción inmuno-histoquímica de <i>Schlafen5</i> .....	186
4.5. Resultados estadísticos de la tinción inmuno-histoquímica de <i>Schlafen5</i> .....	186
4.6. Doble marcaje de <i>Schlafen 5</i> y <i>CD2</i> , <i>CD20</i> o <i>Mac2</i> .....	193
<b>DISCUSIÓN.....</b>	<b>195</b>
1. ESTUDIO DE ASOCIACIÓN DE GENES DE SEÑALIZACIÓN DE <i>HELICOBACTER PYLORI</i> Y RIESGO DE CÁNCER GÁSTRICO .....	197
1.1. <i>NOD2</i> . Asociación genética y plausibilidad biológica.....	197
1.2. <i>CD14</i> . Asociación genética y plausibilidad biológica.....	200
1.3. <i>TLR4</i> y <i>NFKB1</i> . Asociación genética y plausibilidad biológica .....	202
2. ESTUDIO DE ASOCIACIÓN GENÉTICA EN EL SEGUIMIENTO DE LESIONES PRECURSORAS GÁSTRICAS.....	204
2.1. Asociación genética con la evolución de las lesiones precursoras gástricas.....	205
3. ESTUDIO DE EXPRESIÓN GÉNICA EN METAPLASIA INTESTINAL QUE PROGRESA A CÁNCER GÁSTRICO .....	212
3.1 Genes diferencialmente expresados.....	212
3.1.1 Subtipos histológicos de metaplasia intestinal que progresan a cáncer gástrico.....	212
3.1.2 Metaplasia intestinal que no progresa a cáncer gástrico en comparación con mucosa gástrica sana .....	215

---

3.2 Vías de señalización celular y procesos biológicos implicados en la progresión de los subtipos histológicos de metaplasia intestinal a cáncer gástrico .....	218
3.2.1 Gene Set Enrichment Analysis .....	219
3.2.2 Ingenuity Pathway Analysis .....	229
3.3. Validación de la expresión génica por qRT-PCR.....	232
4. ESTUDIO DE EXPRESIÓN GÉNICA E INMUNO-HISTOQUÍMICA DE SCHLAFEN5 EN LESIONES PRECURSORAS Y CÁNCER GÁSTRICO.....	232
<b>CONCLUSIONES .....</b>	<b>239</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>243</b>
<b>TABLAS ANEXAS .....</b>	<b>271</b>
<b>PUBLICACIÓN .....</b>	<b>295</b>

## ÍNDICE DE FIGURAS

<b>FIGURA 1.</b> DISTRIBUCIÓN MUNDIAL DE LA INCIDENCIA DE CÁNCER GÁSTRICO ESTANDARIZADA POR EDAD .....	16
<b>FIGURA 2.</b> ESTRUCTURA ANATÓMICA DEL ESTÓMAGO. ....	16
<b>FIGURA 3.</b> DISPOSICIÓN DE LAS CAPAS CONSTITUYENTES DEL TRACTO DIGESTIVO. ....	17
<b>FIGURA 4.</b> ESTRUCTURA HISTOLÓGICA DE LAS GLÁNDULAS OXÍNTICAS O FÚNDICAS Y PILÓRICAS .....	18
<b>FIGURA 5.</b> ADENOCARCINOMA GÁSTRICO DE TIPO INTESTINAL .....	20
<b>FIGURA 6.</b> ADENOCARCINOMA GÁSTRICO DE TIPO DIFUSO .....	20
<b>FIGURA 7.</b> MUCOSA GÁSTRICA COLONIZADA POR <i>H. PYLORI</i> .....	21
<b>FIGURA 8.</b> MECANISMO DE ACCIÓN DE VACA EN EPITELIO GÁSTRICO.....	22
<b>FIGURA 9.</b> MECANISMOS DE INFECCIÓN DE <i>HELICOBACTER PYLORI</i> EN EL EPITELIO GÁSTRICO. ....	25
<b>FIGURA 10.</b> MECANISMO DE LA CARCINOGENÉISIS GÁSTRICA .....	29
<b>FIGURA 11.</b> MUCOSA NORMAL TEÑIDA CON HEMATOXILINA-EOSINA .....	32
<b>FIGURA 12.</b> GASTRITIS CRÓNICA NO ATRÓFICA .....	32
<b>FIGURA 13.</b> GASTRITIS CRÓNICA ATRÓFICA .....	33
<b>FIGURA 14.</b> METAPLASIA INTESTINAL COMPLETA .....	34
<b>FIGURA 15.</b> METAPLASIA INTESTINAL INCOMPLETA .....	34
<b>FIGURA 16.</b> DISPLASIA GÁSTRICA .....	35
<b>FIGURA 17.</b> SCHLAFEN4 SE SOBRE-EXPRESA EN LA METAPLASIA GÁSTRICA. ....	45
<b>FIGURA 18.</b> MICROARRAYS DE TEJIDO USADOS PARA INMUNOHISTOQUÍMICA DE SCHLAFEN5. ....	112
<b>FIGURA 19.</b> ELECTROFORESIS DE MUESTRAS DE ADN EXTRAÍDAS DE HISOPOS BUCALES DRI-CAPSULES. ....	128
<b>FIGURA 20.</b> ELECTROFORESIS DE MUESTRAS DE ADN EXTRAÍDAS DE HISOPOS BUCALES BUCCALAMP™. ....	129
<b>FIGURA 21.</b> ELECTROFORESIS DE MUESTRAS DE ADN EXTRAÍDAS DE SALIVA EMPLEANDO EL KIT ORAGENE.....	129
<b>FIGURA 22.</b> ELECTROFORESIS DEL ADN EXTRAÍDO DE SALIVA CONSERVADAS EN SOLUCIÓN ESTABILIZADORA # 4.....	131
<b>FIGURA 23.</b> ELECTROFORESIS DEL ADN EXTRAÍDO DE SALIVA COLECTADA EN LOS HOSPITALES EN SOLUCIÓN ESTABILIZADORA # 4.....	131
<b>FIGURA 24.</b> GENOTIPADO DEL SNP rs7578034 EN EL GEN <i>IL1B</i> EMPLEANDO LIGHTCYCLER II. ....	132
<b>FIGURA 25.</b> CLUSTER DE CALIDAD MEDIA EN EL GENOTIPADO SEQUENOM.....	133
<b>FIGURA 26.</b> CLUSTERS DE CALIDAD ÓPTIMA EN EL GENOTIPADO SEQUENOM. ....	133
<b>FIGURA 27.</b> ELECTROFORESIS DE ARN TOTAL OBTENIDO DE MUESTRAS PARAFINADAS EN BIOANALYZER 2100 .....	145
<b>FIGURA 28.</b> DENDROGRAMA DE TODOS LOS GRUPOS ANALIZADOS. ....	148
<b>FIGURA 29.</b> HEATMAP DE LA COMPARACIÓN MII-CG VS MII-NO CG. ....	149
<b>FIGURA 30.</b> HEATMAP DE LA COMPARACIÓN MIC-CG VS MIC-NO CG .....	149
<b>FIGURA 31.</b> HEATMAP DE LA COMPARACIÓN MI-NO CG VS SANOS. ....	150
<b>FIGURA 32.</b> DIAGRAMA DE VENN DE LOS GENES SIGNIFICATIVOS EN LAS COMPARACIONES RELATIVAS A LA MII. ....	151
<b>FIGURA 33.</b> DIAGRAMA DE VENN DE LOS GENES SIGNIFICATIVOS EN LAS COMPARACIONES RELATIVAS A LA MIC.....	152
<b>FIGURA 34.</b> NÚMERO DE GENES SOBRE Y SUB-EXPRESADOS EN LOS GRUPOS ANALIZADOS. ....	152
<b>FIGURA 35.</b> ANÁLISIS DE GENES <i>LEADING EDGE</i> CON LOS CONJUNTOS GÉNICOS SOBRE-EXPRESADOS EN MII-CG.....	169
<b>FIGURA 36.</b> SUBANÁLISIS DE GENES <i>LEADING EDGE</i> DEL NÚCLEO 1 (DEGRADACIÓN PROTEASOMAL). ....	169
<b>FIGURA 37.</b> <i>HEAT MAP</i> DE LOS GENES <i>LEADING EDGE</i> DEL NÚCLEO 1. ....	170
<b>FIGURA 38.</b> SUBANÁLISIS DE GENES <i>LEADING EDGE</i> DEL NÚCLEO 2 (FOSFORILACIÓN OXIDATIVA).....	170
<b>FIGURA 39.</b> ESQUEMAS DE LAS VÍAS CANÓNICAS SOBRE-EXPRESADAS EN LOS GRUPOS MII-CG Y MIC-CG POR IPA... 173	173
<b>FIGURA 40.</b> REDES DE ALTO SCORE IPA DE LAS COMPARACIONES MII-CG VS MII-NO CG Y MI-NO CG VS SANOS. ... 175	175
<b>FIGURA 41.</b> <i>HEATMAP</i> DE LA QRT-PCR EN LA PLATAFORMA BIOMARK HD 96x96 .....	180
<b>FIGURA 42.</b> GRÁFICO DE CAJAS DE LOS VALORES DE Ct DEL GEN <i>ACTB</i> EN LOS GRUPOS MI-CG, MI-NO CG Y SANOS. 181	181
<b>FIGURA 43.</b> GRÁFICA DEL NIVEL DE EXPRESIÓN DEL MICROARRAY Y LA QRT-PCR DE LOS GENES SIGNIFICATIVOS. ....	182
<b>FIGURA 44.</b> INDUCCIÓN DE SCHLAFEN5 POR EL IFNA EN LAS LÍNEAS CELULARES HL-60 Y JURKAT.....	185
<b>FIGURA 45.</b> ANÁLISIS INMUNOHISTOQUÍMICO DE SCHLAFEN5 TRAS INDUCCIÓN POR IFNA EN LA LÍNEA JURKAT. ....	185
<b>FIGURA 46.</b> INMUNO-HISTOQUÍMICA DE LA PROTEÍNA SCHLAFEN5.....	187

<b>FIGURA 47.</b> INMUNO-HISTOQUÍMICA DE LA PROTEÍNA SCHLAFEN5 EN LAS MUESTRAS ESTUDIADAS .....	189
<b>FIGURA 48.</b> REPRESENTACIÓN GRÁFICA DE LA TINCIÓN ESTROMAL DE SCHLAFEN5.....	191
<b>FIGURA 49.</b> CURVAS DE SOBREVIVENCIA PARA LOS MODELOS MI-CG CON Y SIN ADICIÓN DE SCORE H ESTROMAL .....	192
<b>FIGURA 50.</b> CURVAS DE SOBREVIVENCIA PARA LOS MODELOS MII-CG CON Y SIN ADICIÓN DE SCORE H ESTROMAL .....	193
<b>FIGURA 51.</b> DOBLE MARCAJE DE SCHLAFEN5 CON MARCADORES DE SUPERFICIE. ....	194
<b>FIGURA 52.</b> BLOQUES DE HAPLOTIPOS DE <i>MUC2</i> .....	208
<b>FIGURA 53.</b> DIAGRAMA DE VENN DE PROCESOS FUNCIONALES SOBRE-EXPRESADOS COMUNES Y ESPECÍFICOS ENTRE LOS GRUPOS MII-CG Y MI-NO CG, SEGÚN LOS ANÁLISIS GSEA E IPA.....	220
<b>FIGURA 54.</b> PROCESOS FUNCIONALES COMUNES A CONJUNTOS GÉNICOS CON DIFERENTES CRITERIOS DE SELECCIÓN EN LA MII-CG.....	224
<b>FIGURA 55.</b> PROCESOS FUNCIONALES COMUNES A CONJUNTOS GÉNICOS CON DIFERENTES CRITERIOS DE SELECCIÓN EN LA MI-NO CG. ....	228

## ÍNDICE DE TABLAS

<b>TABLA 1.</b> TASAS DE TRANSICIÓN PARA LA PROGRESIÓN Y REGRESIÓN DE LAS LESIONES PRECURSORAS .....	28
<b>TABLA 2.</b> PRINCIPALES ESTUDIOS EPIDEMIOLÓGICOS DE SEGUIMIENTO DE LESIONES PRECURSORAS GÁSTRICAS. ....	38
<b>TABLA 3.</b> RESUMEN DE ESTUDIOS DE ASOCIACIÓN DE GENOMA COMPLETO EN CÁNCER GÁSTRICO. ....	52
<b>TABLA 4.</b> ESTUDIOS DE ASOCIACIÓN GENÉTICA EN LESIONES PRECURSORAS DE CÁNCER GÁSTRICO. ....	56
<b>TABLA 5.</b> SOLUCIONES DE COLECTA Y ESTABILIZADORAS DE SALIVA.....	78
<b>TABLA 6.</b> GENES CANDIDATOS SELECCIONADOS A PARTIR DE ESTUDIOS DE MODELOS MURINOS DE CARCINOGENESIS GÁSTRICA. ....	83
<b>TABLA 7.</b> GENES CANDIDATOS SELECCIONADOS A PARTIR DEL ANÁLISIS DE LOS MECANISMOS DE INFECCIÓN DE <i>HELICOBACTER PYLORI</i> . ....	84
<b>TABLA 8.</b> GENES CANDIDATOS SELECCIONADOS A PARTIR DE ESTUDIOS DE ASOCIACIÓN GENÉTICA CON CÁNCER GÁSTRICO (GWAS) Y CON LESIONES PRECURSORAS GÁSTRICAS (GENES CANDIDATOS). ....	85
<b>TABLA 9.</b> SNPs GENOTIPADOS CON SU INFORMACIÓN ASOCIADA. ....	88
<b>TABLA 10.</b> MUESTRAS USADAS PARA PRUEBA PILOTO DE COMPARACIÓN DE MICROARRAYS DE EXPRESIÓN. ....	95
<b>TABLA 11.</b> MUESTRAS USADAS EN EL ESTUDIO DE EXPRESIÓN DIFERENCIAL POR MICROARRAYS DE LOS SUBTIPOS HISTOLÓGICOS DE MI QUE PROGRESAN A CG, RESPECTO A LOS QUE NO PROGRESAN Y A LA MUCOSA SANA. ....	96
<b>TABLA 12.</b> GENES SIGNIFICATIVOS DEL MICROARRAY DE LAS COMPARACIONES MII-CG vs MII-No CG, MIC-CG vs MIC-No CG Y MI-No CG vs SANOS VALIDADOS POR qRT-PCR. ....	105
<b>TABLA 13.</b> SERIE INDEPENDIENTE DE MUESTRAS USADAS EN LA VALIDACIÓN DE EXPRESIÓN POR qRT-PCR.....	107
<b>TABLA 14.</b> MUESTRAS DE LESIONES PRECURSORAS Y CÁNCER GÁSTRICO USADAS EN EL ESTUDIO INMUNO-HISTOQUÍMICO DE SCHLAFEN5. ....	113
<b>TABLA 15.</b> PRINCIPALES CARACTERÍSTICAS DE LA POBLACIÓN ANALIZADA EN EL ESTUDIO DE ASOCIACIÓN GENÉTICA CON CÁNCER GÁSTRICO.....	121
<b>TABLA 16.</b> CARACTERÍSTICAS DE LOS SNPs ANALIZADOS Y SU ASOCIACIÓN ESTADÍSTICA CON EL CG GLOBAL (CARDIAS + NO CARDIAS/INTESTINAL + DIFUSO). ....	124
<b>TABLA 17.</b> ASOCIACIÓN ESTADÍSTICA DE SNPs CON RESPECTO A LA LOCALIZACIÓN ANATÓMICA DEL CÁNCER GÁSTRICO. ....	125
<b>TABLA 18.</b> ASOCIACIÓN ESTADÍSTICA DE SNPs PREVIAMENTE ASOCIADOS CON CÁNCER GÁSTRICO NO CARDIAS EN INDIVIDUOS INFECTADOS POR <i>HELICOBACTER PYLORI</i> CON RESPECTO AL ESTATUS DE CAGA. ....	126
<b>TABLA 19.</b> ASOCIACIÓN ESTADÍSTICA DE SNPs CON RESPECTO AL SUBTIPO HISTOLÓGICO DE CÁNCER GÁSTRICO. ....	126
<b>TABLA 20.</b> ASOCIACIÓN ESTADÍSTICA DE HAPLOTIPOS RESPECTO A LOS SUBTIPOS HISTOLÓGICOS Y ANATÓMICOS DE CÁNCER GÁSTRICO.....	127
<b>TABLA 21.</b> BÚSQUEDA DE POTENCIALES EQTLs ENTRE LOS SNPs ASOCIADOS Y NO ASOCIADOS A CG.....	128
<b>TABLA 22.</b> PROMEDIO DE LA MASA DE ADN OBTENIDA Y PRESENCIA DE OXIDACIÓN EN LAS DIFERENTES SOLUCIONES ESTABILIZADORAS DE SALIVA.....	131
<b>TABLA 23.</b> RESULTADOS DEL GENOTIPADO DEL SNP rs7578034 EN GEN <i>IL1B</i> EN MUESTRAS DE ADN EXTRAÍDO DE SALIVA. ....	132
<b>TABLA 24.</b> CARACTERÍSTICAS AL RECLUTAMIENTO DE LA POBLACIÓN ESTUDIADA. ....	135
<b>TABLA 25.</b> OR (IC 95%) DE LOS SNPs ASOCIADOS SIGNIFICATIVAMENTE CON PROGRESIÓN Y REGRESIÓN DE LAS LPGs. ....	136
<b>TABLA 26.</b> ASOCIACIÓN CON LA EVOLUCIÓN DE LPGs EN EL PRESENTE Y EN EL ESTUDIO DE SEGUIMIENTO DE LPGs DE SORIA. ....	138
<b>TABLA 27.</b> ASOCIACIONES SIGNIFICATIVAS DE HAPLOTIPOS RESPECTO A LA EVOLUCIÓN DE LAS LPGs. ....	140
<b>TABLA 28.</b> ANÁLISIS DE ASOCIACIÓN EN PORTADORES DE LA COMBINACIÓN DE LOS FACTORES DE VIRULENCIA DE <i>H. PYLORI</i> CAGA+VACAs1/M1 VS OTRAS COMBINACIONES, EN INDIVIDUOS INFECTADOS POR <i>H. PYLORI</i> . ....	141
<b>TABLA 29.</b> POTENCIALES EQTL ENTRE LOS SNPs ESTUDIADOS. ....	143
<b>TABLA 30.</b> MEDICIÓN DE CANTIDAD Y CALIDAD DE LAS MUESTRAS DE ARN USADAS EN LA PRUEBA PILOTO DE COMPARACIÓN DE MICROARRAYS HUMAN GENE 2.0 ST Y ALMAC XCEL. ....	145
<b>TABLA 31.</b> MEDICIÓN DE CANTIDAD Y CALIDAD DE LAS MUESTRAS USADAS EN EL MICROARRAY DE EXPRESIÓN.....	146
<b>TABLA 32.</b> GENES DIFERENCIALMENTE EXPRESADOS ENTRE LOS GRUPOS MII-CG vs MII-No CG. ....	153

<b>TABLA 33.</b> GENES DIFERENCIALMENTE EXPRESADOS ENTRE LOS GRUPOS MIC-CG VS MIC-NO CG.....	157
<b>TABLA 34.</b> GENES DIFERENCIALMENTE EXPRESADOS ENTRE LOS GRUPOS MI-CG VS MI-NO CG.....	158
<b>TABLA 35.</b> CONJUNTOS GÉNICOS SOBRE-EXPRESADOS EN MII-CG CON LOS VALORES MÁS EXTREMOS DEL PARÁMETRO <i>RANK AT MAX.</i> ....	162
<b>TABLA 36.</b> CONJUNTOS GÉNICOS SOBRE-EXPRESADOS EN MI-NO CG CON LOS VALORES MÁS EXTREMOS DEL PARÁMETRO <i>RANK AT MAX.</i> ....	164
<b>TABLA 37.</b> CONJUNTOS GÉNICOS COMPUESTOS POR AL MENOS 3 GENES <i>LEADING EDGE</i> SOBRE-EXPRESADOS EN EL GRUPO MII-CG.....	166
<b>TABLA 38.</b> CONJUNTOS GÉNICOS COMPUESTOS POR AL MENOS 3 GENES <i>LEADING EDGE</i> SOBRE-EXPRESADOS EN EL GRUPO MI-NO CG.....	167
<b>TABLA 39.</b> VÍAS CANÓNICAS DESREGULADAS E INFORMACIÓN FUNCIONAL OBTENIDA EN EL IPA.....	172
<b>TABLA 40.</b> REDES MOLECULARES OBTENIDAS EN EL IPA PARA LAS COMPARACIONES MII-CG VS MII-NO CG Y MI-NO CG VS SANOS.....	177
<b>TABLA 41.</b> RESULTADOS DEL ANÁLISIS DE ONTOLOGÍA GÉNICA EN LAS COMPARACIONES DE ESTUDIO.....	179
<b>TABLA 42.</b> RESULTADOS DE PRUEBA ANOVA DE LOS GENES DE REFERENCIA EN LOS GRUPOS MI-CG, MI-NO CG Y CONTROLES SANOS.....	181
<b>TABLA 43.</b> RESULTADOS DEL TEST DE PERMUTACIONES DEL PROGRAMA <i>BOOTSRTATIO</i> EMPLEADO PARA ESTIMAR DIFERENCIAS SIGNIFICATIVAS ENTRE GRUPOS MI-CG VS MI-NO CG.....	183
<b>TABLA 44.</b> EXPRESIÓN GÉNICA DE <i>SCHLAFEN5</i> EN LÍNEAS CELULARES EXPUESTAS A DIVERSAS CONDICIONES.....	184
<b>TABLA 45.</b> CUANTIFICACIÓN DE LA TINCIÓN ESTROMAL Y GLANDULAR DE <i>SCHLAFEN5</i> MEDIANTE EL SCORE H EN LOS GRUPOS ANALIZADOS. ....	190

## ABREVIATURAS Y SÍMBOLOS

**ADN:** Ácido desoxirribonucleico.

**AINES:** Antinflamatorios no esteroideos.

**ANOVA:** Análisis de varianza.

**ARN:** Ácido ribonucleico.

**ARNm:** Ácido ribonucleico mensajero.

**AUC:** *Area under curve*.

**BM-MSCs:** *Bone marrow derived mesenchymal stem cells*.

**CEGEN:** Centro Nacional de Genotipado.

**CG:** Cáncer gástrico.

**DAB:** Diaminobenzidina.

**DE:** Desviación estándar.

**DISP:** Displasia.

**E:** Extensión.

**EPIC:** *European Prospective Investigation on Cancer*.

**eQTL:** *Expression quantitative trait loci*.

**ES:** *Enrichment score*.

**ESE:** *Exonic splicing enhancer*.

**ESS:** *Exonic splicing silencer*.

**EUA:** Estados Unidos de América.

**FDR:** *False discovery rate*.

**GCA:** Gastritis crónica atrófica multifocal.

**GEO:** *Gene expression omnibus*.

**GNA:** Gastritis no atrófica.

**GO:** *Gene ontology*.

**GSEA:** *Gene set enrichment analysis*.

**GWAS:** *Genome wide association study*.



**HapMap:** Proyecto de mapeo de bloques de haplotipos humanos.

**IARC:** *International Agency for Research on Cancer.*

**IC95%:** Intervalo de confianza del 95%.

**ICO:** Instituto Catalán de Oncología.

**IDIBELL:** Instituto de Investigación Biomédica de Bellvitge.

**IMIM:** Instituto Hospital del Mar de Invesigaciones Médicas.

**IPA:** *Ingenuity pathway analysis.*

**LD:** *Linkage disequilibrium.*

**Limma:** *Linear models for microarray data.*

**LPGs:** Lesiones precursoras gástricas.

**LPS:** Lipopolisacárido.

**LRT1:** Laboratorio de Investigación Traslacional 1.

**MAF:** *Minor allele frequency.*

**MDSC:** *Myeloid derived suppressor cells.*

**MI:** Metaplasia intestinal.

**MI-No CG:** Metaplasia intestinal que no progresa a cáncer gástrico

**MIC:** Metaplasia intestinal completa.

**MIC-CG:** Metaplasia intestinal completa que progresa a cáncer gástrico.

**MIC-No CG:** Metaplasia intestinal completa que no progresa a cáncer gástrico.

**MII:** Metaplasia intestinal incompleta.

**MII-CG:** Metaplasia intestinal incompleta que progresa a cáncer.

**MII-No CG:** Metaplasia intestinal incompleta que no progresa a cáncer gástrico.

**MSI:** *Microsatellite instability.*

**MSigDB:** *Molecular signatures database.*

**NGS:** *Next generation sequencing.*

**NK:** *Natural killers.*

**NLR:** *Nod like receptors.*

**NMD:** *Nonsense mediated mRNA decay.*

**OR:** *Odds ratio.*

**PAMPs:** *Pathogen associated molecular patterns.*

**PCR:** *Polymerase chain reaction.*

**PGI/PGII:** Razón de concentraciones de pepsinógenos I/II.

**PRRs:** *Pattern recognition receptors.*

**qRT-PCR:** *Quantitative real time polymerase chain reaction.*

**RIN:** *RNA integrity number.*

**RM:** Razón de medias.

**ROC:** *Receiver operating characteristic, curva de sobrevivencia.*

**Shh:** Sonic hedgehog.

**SNP:** *Single nucleotide polymorphism.*

**SPEM:** *Spasmolytic polypeptide expressing metaplasia.*

**STA:** *Specific target amplification.*

**TA:** Temperatura ambiente.

**TE:** Tampón Tris-EDTA.

**TFBS:** *Transcription factor binding site.*

**TLE:** Tampón de lisis de eritrocitos.

**TLR:** *Toll like receptors.*

**UNAC:** Unidad de Nutrición, Ambiente y Cáncer.

**V:** Volumen.

**VNTR:** *Variable number tandem repeat.*

**Vs:** Volúmenes.



# **PRESENTACIÓN**



El cáncer gástrico (CG) es la consecuencia de múltiples cambios histológicos en la mucosa gástrica conocidos como la cascada de Correa y en el mismo intervienen múltiples factores ambientales y genéticos. Entre ellos se encuentran la infección por la bacteria *Helicobacter pylori*, factores de susceptibilidad genética como polimorfismos, y variaciones en la expresión de genes y vías de señalización celular, y procesos moleculares condicionantes de la carcinogénesis. La Unidad de Nutrición, Ambiente y Cáncer (UNAC), donde se ha realizado este trabajo, tiene una larga experiencia en el estudio de los factores de riesgo ambientales y genéticos de dicha enfermedad.

Esta tesis consta de cuatro estudios similares en cuanto a que exploran la susceptibilidad del paciente a desarrollar cáncer gástrico; pero diferentes respecto a las muestras y metodología de estudio usadas en cada uno. El contenido de la tesis se estructura según el esquema de introducción, hipótesis, objetivos, materiales y métodos, resultados, discusión, conclusiones y referencias bibliográficas. Finalmente, el apéndice incluye la publicación obtenida como consecuencia directa de este trabajo.



# **INTRODUCCIÓN**

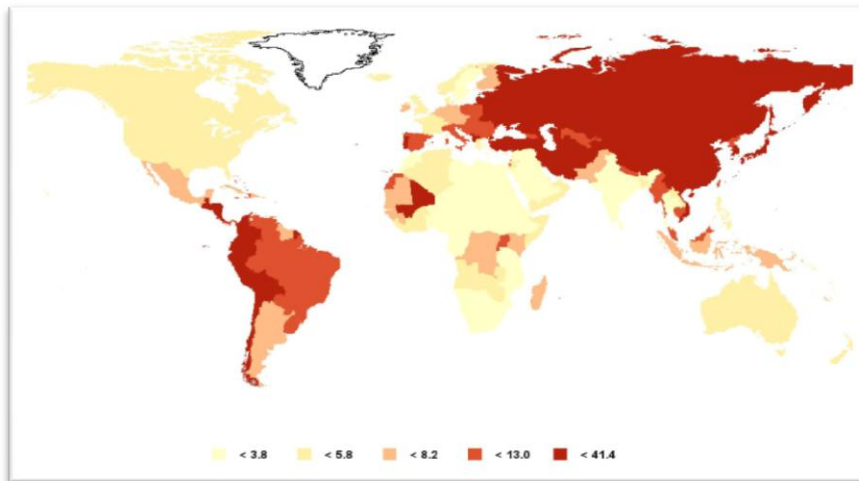




## **1. Cáncer gástrico. Descripción general y epidemiología**

El cáncer gástrico se define como el crecimiento celular tumoral en el estómago. El mismo se divide en CG esporádico, el más frecuente, y en hereditario, el cual presenta una fuerte agregación familiar <sup>1</sup>. A pesar de que la incidencia del CG, especialmente el no cardias, ha disminuido en muchos países en los últimos 50 años, esta enfermedad es la tercera causa de muerte por cáncer y ocupa el tercer lugar a nivel mundial entre los tumores más frecuentes <sup>1</sup>. Existen marcadas diferencias en su incidencia entre regiones geográficas. Las más altas se encuentran en Japón, Corea, China y Mongolia (entre 40 y 60 por cada 100 000 habitantes), Europa del Este (35/100 000) y regiones andinas de América Latina (20-30/100 000). Las más bajas se observan en África (0.6-3/100 000), América del Norte (5-6/100 000) y Holanda (6.9/100 000) (Figura 1). La incidencia en los hombres es el doble respecto a las mujeres <sup>2</sup>, lo cual se ha relacionado con un efecto protector de los estrógenos en las últimas <sup>3</sup>. Su riesgo incrementa con la edad, siendo frecuente en sujetos entre los 55 y 80 años, y raro en los menores de 30 años <sup>1</sup>.

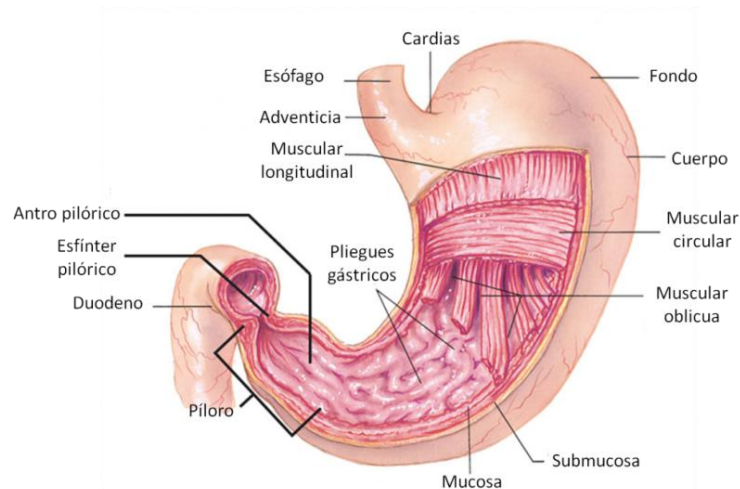
En España la mortalidad por CG en las últimas décadas ha disminuido en casi todas sus provincias. Las mayores tasas de mortalidad corresponden a zonas del centro y norte del país, sobre todo a Castilla-León con 20.14 muertes por cada 100 000 habitantes entre 2000 y 2005, la cual se encuentra entre las más altas de Europa Occidental <sup>4</sup>. La tasa promedio de supervivencia al CG a 5 años es <20%, principalmente debido a un diagnóstico tardío, ya que sus primeras etapas son clínicamente silentes <sup>5</sup>. Si el tumor se detecta y se trata antes de que invada la capa muscular, la supervivencia a los 5 años puede llegar al 90% <sup>6</sup>. A pesar de ello, sólo unos pocos países con alta incidencia del mismo, especialmente Japón, han puesto en marcha amplios programas de detección temprana basados en la identificación radiológica y endoscópica <sup>6</sup>. Sin embargo, dichos programas no son posibles en países de baja y moderada incidencia. Por tanto, la detección y vigilancia de las lesiones preneoplásicas con mayor riesgo de progresión a CG y la erradicación de los principales factores de riesgo es el método más efectivo para su prevención en países de baja y moderada incidencia.



**Figura 1.** Distribución mundial de la incidencia de cáncer gástrico estandarizada por edad. Tomado de Ferlay *et al.*, 1998 <sup>7</sup>.

## 2. Anatomía e histología del estómago

Anatómicamente el estómago se divide en tres regiones: fundus, cuerpo y antro. El mismo posee dos esfínteres: el cardias, que separa el estómago del esófago, y el píloro, que separa el estómago del intestino delgado (Figura 2).



**Figura 2.** Estructura anatómica del estómago. Tomado de Hall *et al.*, 2010 <sup>8</sup>.

Histológicamente, la pared del estómago está formada por capas, nombrándose a partir de la que está en contacto directo con la luz gástrica, las mismas son:

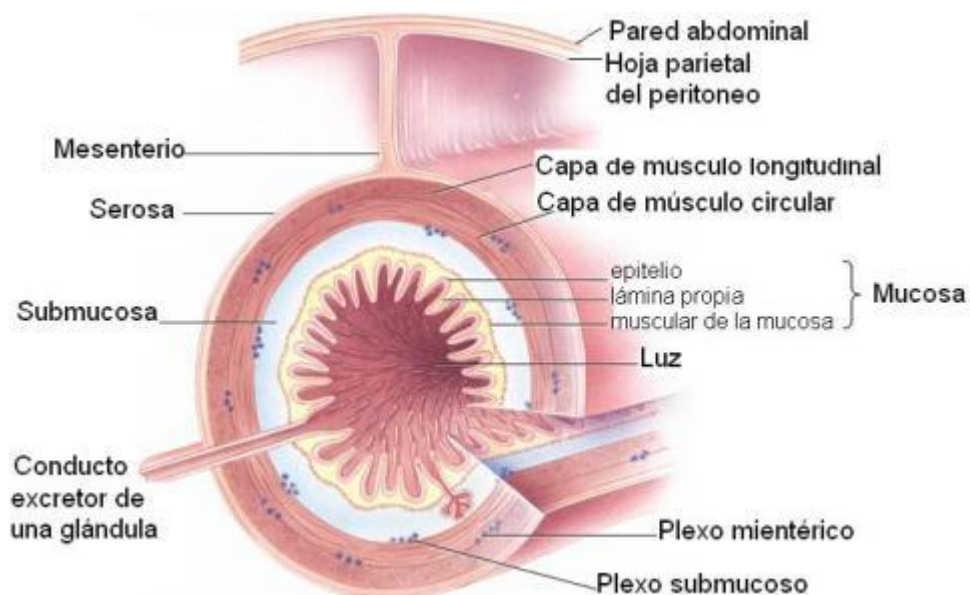
**-Mucosa:** Formada por un epitelio que forma las glándulas gástricas secretoras y que está en contacto con el lumen gástrico. Está constituida, a su vez, por la lámina propia o estroma, formada por vasos sanguíneos, tejido conectivo y células de respuesta inmune como linfocitos, células plasmáticas, eosinófilos, macrófagos y neutrófilos; y una fina capa de músculo liso denominada *mucosa muscularis*, que separa la mucosa de la submucosa, y cuya contracción posibilita la secreción y absorción de sustancias. Su epitelio superficial es formado por las foveolas gástricas

hacia las cuales vierten su contenido las glándulas gástricas (entre 4 y 5 a cada foveola) y estas a su vez vierten hacia el lumen del estómago. Además, presenta células mucosas de superficie que secretan un mucus de pH básico, el cual protege a la mucosa de las secreciones ácidas y sirve como una barrera física entre la mucosa y los contenidos estomacales.

**-Submucosa:** Compuesta por grandes vasos sanguíneos que alimentan al resto de las capas, vasos linfáticos, un plexo nervioso autónomo (sistema nervioso entérico) y glándulas gástricas aisladas.

**-Muscular (*muscularis externa*):** Compuesta por dos capas de tejido muscular liso, una capa interior orientada circularmente y otra exterior dispuesta longitudinalmente. Entre ambas capas existe tejido conectivo que contiene al plexo nervioso mientérico. Las contracciones de la *muscularis externa* bajo el control del sistema nervioso entérico son responsables de la peristalsis, las contracciones que producen acortamientos y alargamientos del tracto gastrointestinal para mezclar y mover el alimento.

**-Serosa:** Estrato más alejado de la luz del estómago, compuesto por epitelio escamoso simple denominado mesotelio y tejido conectivo subyacente. Grandes vasos sanguíneos y linfáticos, así como troncos nerviosos, viajan a través de la serosa hasta la pared del tracto digestivo. En la figura 3 se observa la disposición histológica de las capas antes descritas.



**Figura 3.** Disposición de las capas constituyentes del tracto digestivo. Modificado de Hall *et al.*, 2010 <sup>8</sup>.

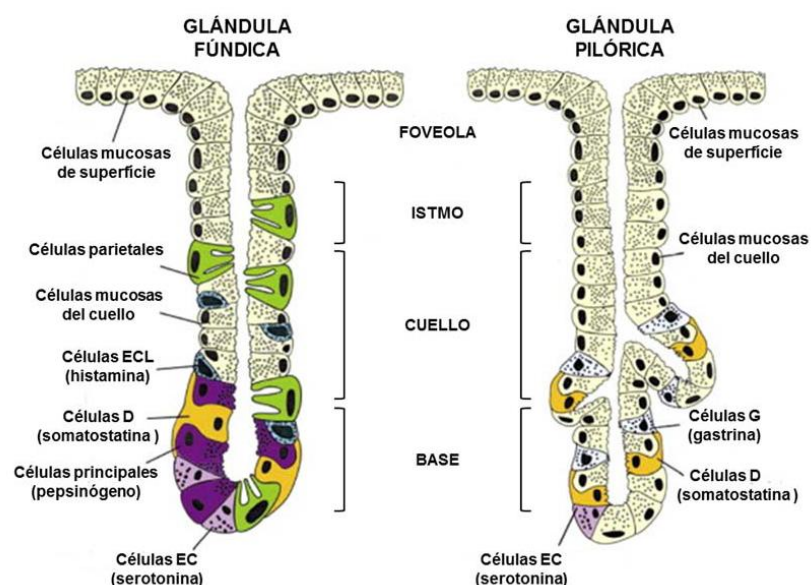
Por otra parte, cada área anatómica del estómago presenta una composición celular característica (Figuras 2 y 4):

**-Cardias:** Compuesto de células mucosas con algunas enteroendocrinas.

**-Fundus y cuerpo:** Constituido por glándulas oxínticas o fúndicas responsables de la digestión, las que a su vez están compuestas por diversos tipos celulares:

- Células mucosas de superficie y del cuello secretoras de mucus.
- Células parietales u oxínticas productoras de ácido clorhídrico y del factor intrínseco que garantiza la absorción de la vitamina B12.
- Células principales localizadas en la base y secretoras de pepsinógeno, el cual por la acción del ácido clorhídrico es convertido en la enzima proteolítica pepsina.
- Células enteroendocrinas o ECL (*enterochromaffin*) productoras de las hormonas gastrointestinales gastrina, ghrelina, secretina, colecistoquinina, péptido gástrico inhibitorio, motilina e histamina, un potente estimulador de la producción de ácido clorhídrico en las células parietales.
- Células madres adultas no diferenciadas localizadas en el istmo, las cuales producen todos los tipos celulares antes descritos.

**-Región antro-pilórica:** Glándulas de estructura tubular ramificada y enrollada en su base, conectadas por foveolas cortas, compuestas por una población celular menos compleja que las de las glándulas oxínticas. La principal función de estas glándulas es la secreción de mucus y 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.



**Figura 4.** Estructura histológica de las glándulas oxínticas o fúndicas y pilóricas. Tomado de Schubert *et al.*, 2008<sup>9</sup>.

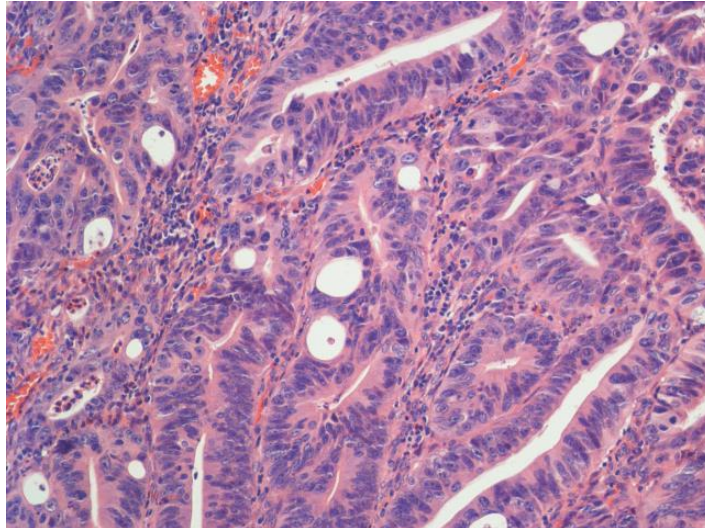
### 3. Tipos anatómicos e histológicos del cáncer gástrico

Alrededor del 95% del CG esporádico está constituido por adenocarcinomas, caracterizados por su surgimiento en el epitelio glandular de la mucosa gástrica, siendo a los mismos a los que nos referiremos en este documento. Existen otros tumores gástricos como linfomas (3-5%) y tumores del estroma gastrointestinal (1-3%) que representan una minoría. Teniendo en cuenta la localización anatómica de los adenocarcinomas gástricos, se clasifican como del cardias y no cardias o distal, este último comprende a los originados en el fundus, cuerpo y la región antro-pilórica <sup>1</sup>. Dichos subtipos anatómicos presentan diferencias en su epidemiología, etiología y patogénesis.

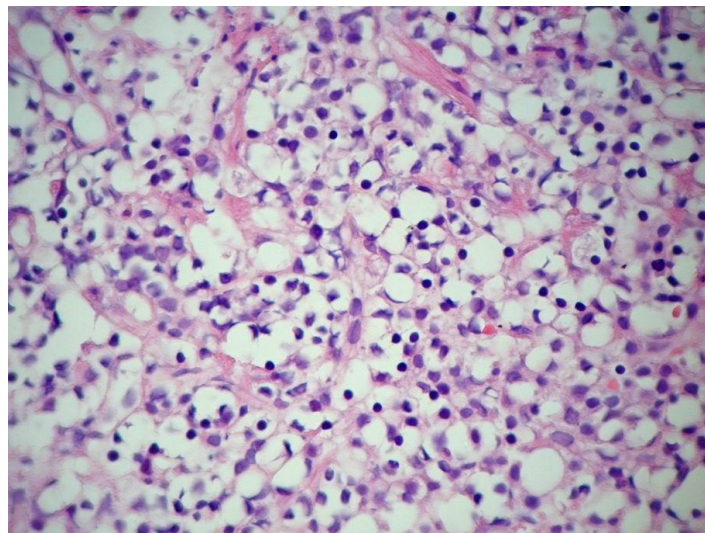
El CG del cardias no se encuentra asociado con la infección por *Helicobacter pylori*. Su incidencia ha aumentado en Europa en las últimas cuatro décadas <sup>10</sup>, ya que alrededor del 30 al 50% de los CGs diagnosticados en países occidentales son del cardias cuando hasta ahora representaban sólo entre el 16 y el 37%. Histológicamente, es de tipo intestinal en su mayoría y se origina probablemente a partir de metaplasia intestinal (MI). Los tumores del cardias se consideran más agresivos y con peor pronóstico que los no cardias, dado que sus sobrevivencias a los 5 años en Estados Unidos de América (EUA) son del 14 y 26%, respectivamente <sup>1</sup>. El factor de riesgo más importante del CG no cardias es la infección por *H. pylori*, siendo responsable de la mayoría de los casos (77-93.2%) <sup>11</sup>. Es el que más ha descendido en su incidencia en el mundo, aunque recientemente se ha observado un aumento en europeos de ambos sexos de EUA en el rango de 25-39 años <sup>12</sup>.

Atendiendo al tipo histológico los CGs se clasifican en intestinal y difuso según la clasificación de Lauren <sup>13</sup>. Los primeros están formados por células tumorales dispuestas cohesivamente en estructuras tubulares irregulares que infiltran el estroma y son semejantes a las del intestino (Figura 5). Representan aproximadamente el 60% de los CGs, su incidencia incrementa con la edad, siendo máxima entre los 55-80 años, su relación de incidencia en hombres:mujeres es 2:1, y predomina en zonas geográficas con alta incidencia de CG <sup>1</sup>.

El CG tipo difuso es un tumor no diferenciado con pérdida de la arquitectura glandular. Las células tumorales son generalmente redondas y pequeñas, con una cohesión intercelular mínima o ausente y con abundante mucinas citoplasmáticas que desplazan el núcleo hacia la periferia <sup>13</sup> (Figura 6). Es más frecuente en poblaciones de bajo riesgo (aunque su localización geográfica es más uniforme), y en los tumores no esporádicos, especialmente en sujetos jóvenes (40-60 años), sin un predominio por sexos y presenta un peor pronóstico que el tipo intestinal <sup>1</sup>.



**Figura 5.** Adenocarcinoma gástrico de tipo intestinal. Tomado de Correa *et al.*, 2012 <sup>14</sup>.



**Figura 6.** Adenocarcinoma gástrico de tipo difuso. Tomado de Correa *et al.*, 2012 <sup>14</sup>.

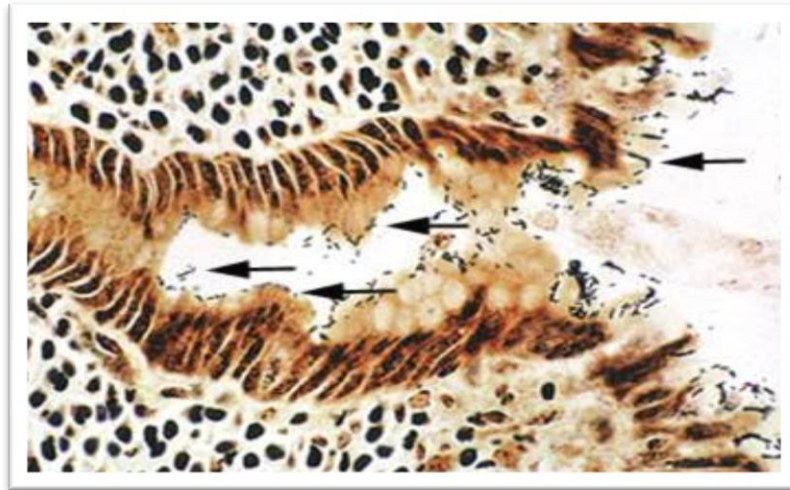
#### **4. *Helicobacter pylori* como factor de riesgo del cáncer gástrico**

##### **4.1. Infección por *Helicobacter pylori*. Aspectos epidemiológicos**

La infección por la bacteria gram-negativa *Helicobacter pylori* es el principal factor de riesgo de gastritis crónica, úlcera péptica, metaplasia intestinal, adenocarcinoma y linfoma gástrico. Se estima que alrededor del 63.4% de todos los casos de CG <sup>15</sup> y hasta el 93.2% del CG no cardias <sup>11</sup> son atribuibles a la misma.

*H. pylori* es una bacteria micro-aerofílica y gram-negativa de localización principalmente extracelular en la mucosa gástrica (Figura 7). Es adquirida fundamentalmente en la niñez temprana, especialmente en los países de más alto

riesgo, hipotéticamente por contacto oral pero se desconoce exactamente su manera de transmisión <sup>1</sup>. Está presente entre el 10-50% de la población adulta de países desarrollados y fue clasificada desde 1994 como un carcinógeno del grupo 1 para humanos <sup>15</sup>. En caso de no ser erradicada la bacteria y/o la inflamación crónica y dependiendo de múltiples factores, el paciente puede evolucionar a lesiones precursoras gástricas (LPGs) de mayor severidad y entre el 0.1-4% de los infectados puede progresar a CG <sup>16</sup>.



**Figura 7.** Mucosa gástrica colonizada por *H. pylori* (indicada por las flechas). Tomado de Correa *et al.*, 2012 <sup>14</sup>.

El hecho de que ciertas áreas geográficas (África, India, Malasia, India, China, Colombia y Costa Rica) presenten elevada prevalencia de infección por *H. pylori* y baja incidencia de CG se conoce como el enigma africano. Las posibles explicaciones a este fenómeno han sido diferencias en la virulencia oncogénica de cepas de *H. pylori*, co-infección con helmintos y desplazamiento de la respuesta inmune desde Th1 a Th2 <sup>16</sup>. Existen evidencias en modelos animales y en humanos de que la respuesta inmune ante la infección de *Helicobacter* puede ser cambiada de un típico patrón Th1 a Th2 debido a la co-infección con helmintos. Este cambio ha sido observado en áreas con una baja incidencia de CG como la costa pacífica colombiana. La prevalencia del parasitismo intestinal en esta población es del 50% con respecto al 25% en las áreas de alto riesgo de las montañas andinas <sup>14</sup>.

#### **4.2. Infección por *Helicobacter pylori*. Factores de virulencia**

Varios componentes del genoma de *H. pylori* como la presencia de *cagA* y las variantes *s1*, *m1* e *i1* de *vacA* son denominados factores de virulencia por estar asociados con un mayor riesgo de lesiones precursoras (apartado 7 de esta sección) y de cáncer gástrico <sup>17</sup>.

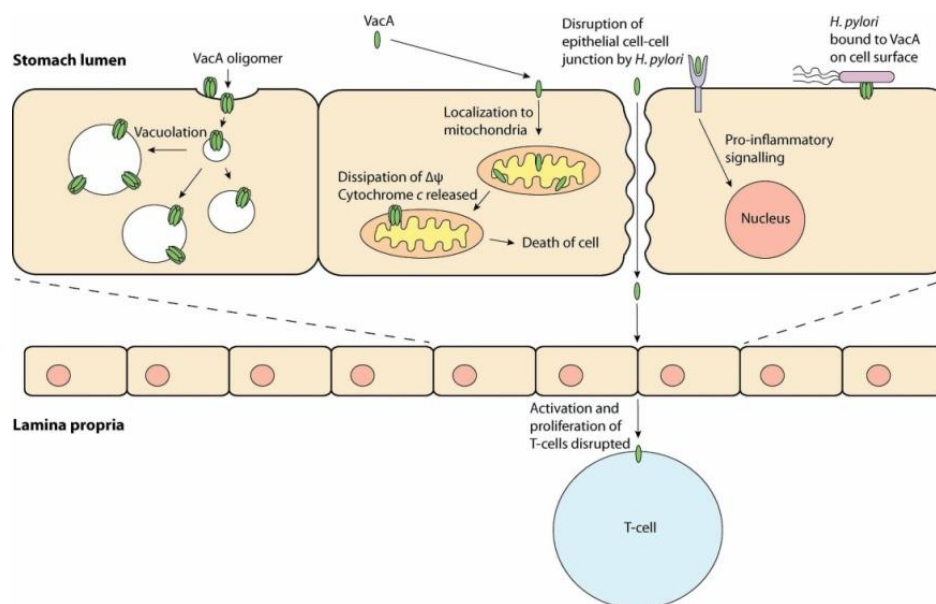
La oncoproteína *cagA* está presente en alrededor del 60% de las cepas de *H. pylori* en los países occidentales, y en casi todas las del este de Asia. *cagA* se encuentra



presente en alrededor del 90% de las cepas de los Andes colombianos (alto riesgo) y en aproximadamente el 70% de las cepas de la costa pacífica colombiana (bajo riesgo) <sup>18</sup>. Un reciente trabajo identificó que en las áreas de bajo riesgo colombianas predominan las cepas de *H. pylori* de origen africano mientras que en las de alto riesgo son más comunes las de origen europeo <sup>19</sup>. Por lo que las cepas africanas serían menos virulentas que las europeas condicionando el riesgo de CG en dependencia del genotipo de las mismas.

En el extremo carboxi-terminal de *cagA* existen unos motivos llamados EPIYA (A-D) los cuales son sitios de fosforilación y se diferencian por la secuencia aminoacídica flanqueante. En los países occidentales predominan los tipos A-C, mientras que en el este de Asia lo hacen A, B y D <sup>17</sup>. La presencia de varios motivos incrementa el riesgo de lesiones precursoras y CG debido a un incremento en la fosforilación y señalización celular inducida por *cagA* <sup>1</sup>.

Otro factor de virulencia de *H. pylori* es la proteína *vacA*, la cual es una citotoxina que produce vacuolas y canales de membrana en las células epiteliales, disrupción de función endosoma/lisosoma y apoptosis (Figura 8). *vacA* está presente en todas las cepas de *H. pylori* y consta de loci polimórficos *s*, *i*, *m*. La progresión de las lesiones premalignas y el CG han sido asociadas con las cepas *s1/i1/m1*, debido a una producción incrementada de toxina con mayor actividad vacuolante respecto a cepas *s2/m2* o *s2/i2/m2* <sup>20</sup>.



**Figura 8.** Mecanismo de acción de *vacA* en epitelio gástrico. Tomado de Palframan *et al.*, 2012 <sup>20</sup>.

#### 4.3. Infección por *Helicobacter pylori*. Mecanismos y consecuencias en el huésped

*H. pylori* posee un complejo mecanismo de infección del epitelio gástrico, en el cual las consecuencias negativas para el huésped son logradas de múltiples maneras (Figura 9) detalladas a continuación:

A) Infección dependiente de cagA fosforilada: cagA penetra en el citoplasma a través del sistema de secreción tipo IV (T4SS) y es fosforilada por SRC, C-ABL o YES1. Posteriormente, interacciona con SHP2 (PTPN11) induciendo la activación de la vía de señalización celular MAPK, específicamente de sus miembros MAPK1 y MAPK3. La consecuencia es el fenotipo colibrí (*hummingbird* en inglés) caracterizado por la elongación, dispersión y migración de las células epiteliales <sup>21</sup>.

B) Infección independiente de cagA fosforilada: Después de penetrar cagA al citoplasma forma un complejo con GRB2, SOS y RAS y es también activada la vía de señalización MAPK como en el mecanismo anterior lo que conlleva a la activación del factor de transcripción AP1 el cual inducirá la expresión de citoquinas pro-inflamatorias IL6, IL8, TNF $\alpha$ , etc., lo que conduce al reclutamiento de neutrófilos y linfocitos, así como a la dispersión y proliferación de las células epiteliales. Las citoquinas inducen inflamación, la cual produce especies reactivas del oxígeno que producen mutaciones al ácido desoxirribonucleico (ADN).

Así mismo, cagA no fosforilada puede localizarse en las uniones intercelulares epiteliales e interactuar con E-cadherina, ZO-1 o JAM-A e inducir la disrupción de la polaridad apico-basal de las células epiteliales y la ruptura de las uniones célula-célula, esenciales para la integridad del epitelio gástrico, debido a la disolución del complejo de adhesión E-cadherina con  $\beta$ -catenina. Esto último provoca la acumulación nuclear y citoplasmática de la  $\beta$ -catenina que, a su vez, induce la expresión de genes intestinales como *CDX1* y *MUC2* en el epitelio gástrico. Además, cagA aumenta la expresión de NFKB1, factor de transcripción que induce la expresión de citoquinas pro-inflamatorias. Adicionalmente, cagA es una oncoproteína que favorece la proliferación descontrolada del epitelio gástrico.

C) Formación de vacuolas por vacA: vacA induce la formación de vacuolas en el epitelio gástrico, las cuales son canales aniónicos que son endocitados por la célula para producir endosomas tardíos y lisosomas tempranos. La actividad de estos canales provoca la acidificación del citosol de las células gástricas y la entrada de agua. Además, vacA induce la liberación del citocromo C lo que causa apoptosis epitelial <sup>22</sup> e inhibe la activación y proliferación de los linfocitos T.

D) Formación de vesículas de membrana externa de *H. pylori*: Dichas vesículas contienen al peptidoglicano de la membrana de *H. pylori* y penetran en las células epiteliales gástricas a través de balsas lipídicas (microdominios de la membrana

plasmática cuya fluidez es mucho menor a la de su entorno). Una vez en el citosol el peptidoglicano es liberado e interacciona con NOD1, aumentando la expresión de NFKB1 e induciendo la expresión de citoquinas pro-inflamatorias <sup>23</sup>.

E) Inhibición de la ATPasa H<sup>+</sup>-K<sup>+</sup>: El receptor del T4SS de *H. pylori*, denominado  $\alpha 5\beta 1$ , al interactuar directamente con *H. pylori* activa a ADAM17, la cual induce la represión transcripcional de la ATPasa H<sup>+</sup>-K<sup>+</sup>, responsable de generar el gradiente de protones H<sup>+</sup> que garantiza el contenido ácido del estómago.

F) Efectos de la inflamación crónica: Las especies reactivas del oxígeno originan mutaciones en el ADN de las células madres mesenquimales procedentes de médula ósea (BM-MSCs, del inglés *bone marrow derived mesenchymal stem cells*) que migran al epitelio gástrico debido a la infección por *H. pylori*, así como en el propio epitelio gástrico. Adicionalmente, el epitelio es susceptible a una apoptosis inducida por citoquinas pro-inflamatorias, la cual induce la muerte de células parietales y principales, provocando la atrofia gástrica. La ausencia de comunicación celular conlleva a la perturbación de los patrones de diferenciación celular, ya que las células parietales son cruciales en la maduración de los precursores epiteliales y su migración dentro de la glándula gástrica <sup>24</sup>. En un modelo murino de CG por infección con *Helicobacter*, la inflamación prolongada induce deposición de BM-MSCs. Estas, al ser cultivadas durante largos periodos se transforman espontáneamente desarrollando un fenotipo tumoral inhibidor de la respuesta inmune, involucrando a las células madres en el origen del CG <sup>25</sup>.

G) Factores transcripcionales homeobox: *H. pylori* es un regulador positivo del factor de transcripción humano CDX2, el cual es un *master gene* responsable de la metaplasia intestinal <sup>26</sup>.

H) Adaptación de *H. pylori* al ambiente gástrico: *H. pylori* posee una potente actividad enzimática ureasa que le permite vivir en el microambiente ácido del lumen gástrico. Esto se logra mediante la hidrólisis de la urea, que produce amonio, el cual protege a la bacteria del pH ácido del lumen gástrico. Otros factores que contribuyen a la persistencia de la bacteria en el estómago son ciertas características del lipopolisacárido (LPS) de la membrana de *H. pylori* que reduce la intensidad de la respuesta inmune del huésped y la expresión de adhesinas que le confieren una estrecha adherencia al epitelio gástrico <sup>27</sup>.

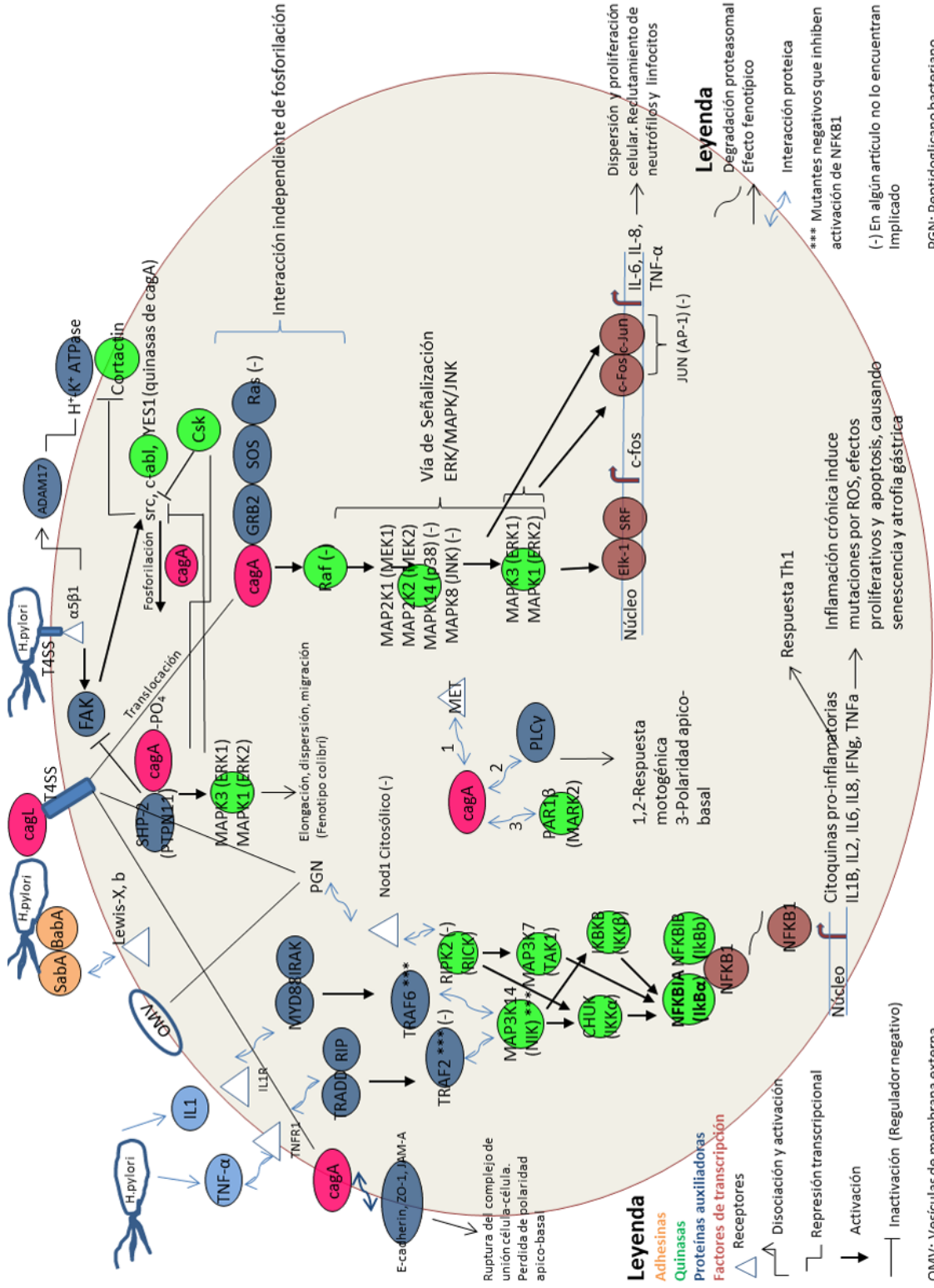


Figura 9. Mecanismos de infección de *Helicobacter pylori* en el epitelio gástrico. Tomado de 17, 21, 23, 27, 28, 29, 30, 31, 32.

#### 4.4. Infección por *Helicobacter pylori*. Respuesta inmunitaria

La respuesta inmune innata es desencadenada por receptores de membrana que se expresan en las células epiteliales gástricas (*Toll Like Receptors* [TLR]: TLR2, TLR4, TLR5, TLR9) y células presentadoras de antígenos (*Nod Like Receptors* [NLR]: NOD1, NOD2). Los TLRs reconocen el lipopolisacárido de la membrana bacteriana mientras que los NLR reconocen el peptidoglicano bacteriano. La activación de los mismos induce la expresión del factor transcripcional NFKB1, el cual induce la expresión de citoquinas pro-inflamatorias como IL1B, IL8, TNFA e IFN $\gamma$ <sup>32</sup>.

La respuesta inmune ante la infección por *Helicobacter* es mayoritariamente de células T en lugar de B, constituyendo la primera una de las más importantes en la eliminación de células tumorales. La respuesta de células T puede ser citotóxica (CD8+) y/o auxiliadora (CD4+). Esta última, basada en la producción de citoquinas, se divide en Th1 (secreción de TNFA e IFN $\gamma$ ) y Th2 (secreción de IL4 e IL5). La respuesta CD8 elimina directamente el agente infeccioso pero es minoritaria en la gastritis atrófica y el CG. La respuesta Th1 induce activación de macrófagos y proliferación de linfocitos T CD8+ mientras que la Th2 induce proliferación de linfocitos B para producir anticuerpos<sup>33</sup>.

En la infección por *H. pylori* existe una respuesta mayoritaria tipo Th1, sin embargo es incapaz de eliminar la infección, contribuyendo a la inflamación crónica<sup>27</sup>. En modelos murinos y pacientes con una respuesta predominante Th2, existe una disminución del riesgo de gastritis crónica atrófica multifocal (GCA) y CG. En ratones con respuesta Th2 la infección con *Helicobacter felis* induce una leve gastritis a pesar de lograr una alta colonización bacteriana. Pacientes infectados con una gastritis leve poseen una respuesta tipo Th0 en la cual secretan tanto citoquinas Th1 como Th2, evidenciando que la respuesta Th2 ejerce protección en comparación con la respuesta Th1<sup>33</sup>.

Como ya hemos comentado anteriormente en poblaciones de bajo riesgo de CG a pesar de una alta incidencia de *H. pylori* como África<sup>16</sup> y la costa pacífica colombiana<sup>14</sup>, la coinfección con helmintos intestinales provoca la respuesta Th2. En ratones coinfectados con *H. felis* y el helminto *Heligmosomoides polygyrus* existe una protección contra la atrofia gástrica y otras LPGs en comparación con infectados solamente con *H. felis*<sup>34</sup>. Por lo que un factor de susceptibilidad que condiciona la severidad de las LPGs es el tipo de respuesta Th1 o Th2 ante la infección por *H. pylori*.

### 5. Otros factores de riesgo del cáncer gástrico

#### 5.1. Dieta

Un reciente meta-análisis muestra que el alto consumo de frutas, pero no de vegetales, disminuye significativamente el riesgo de CG<sup>35</sup>. También, el consumo de fibras provenientes de cereales parece tener un efecto mayor que el de fibras de frutas y

vegetales en la reducción del CG, altos contenidos de los mismos reducen el riesgo de CG en un 31%<sup>36</sup>.

Sin embargo, los consumos elevados de sal, alcohol y carne roja procesada (especialmente en infectados por *H. pylori*)<sup>37</sup> probablemente aumentan el riesgo de CG no cardias. La mayoría de estas asociaciones provienen de estudios caso-control; los pocos estudios de cohortes han sido más inconsistentes y se requieren más estudios de cohorte para arribar a conclusiones definitivas. Las carnes procesadas aportan nitrosaminas y precursores de su formación como los nitratos y nitritos. Las carnes rojas aportan hierro orgánico que es un precursor de la formación endógena de nitrosaminas, las cuales son cancerígenas en modelos de experimentación animal.

El mecanismo para la asociación inversa del riesgo de CG con la ingesta de frutas y posiblemente de algunos vegetales parece estar relacionado con la presencia de las vitaminas C y E, las cuales actúan como antioxidantes y disminuyen la formación endógena de nitrosaminas. Se ha propuesto que la disminución en la incidencia del CG se debe a mejoras en la conservación de los alimentos, especialmente por la refrigeración, que ha permitido reducir el uso de la sal y los ahumados como conservantes y la abundancia de frutas y vegetales frescos en la dieta; así como, a la disminución del consumo de tabaco y de la infección por *H. pylori*<sup>1</sup>.

## 5.2. Tabaquismo

El consumo de tabaco es el factor de riesgo asociado con el mayor número de casos de cáncer en todo el mundo. Se estima que la fracción poblacional atribuible del CG debido al tabaquismo está entre 20-50%<sup>38</sup>.

## 5.3. Anti-inflamatorios no esteroideos

Un meta-análisis encontró que los consumidores regulares de antiinflamatorios no esteroideos (AINES) tuvieron entre un 18-20% de menos riesgo de CG de cardias y entre un 32-36% de reducción del riesgo de CG no cardias<sup>39</sup>. Los AINES inhiben la producción de las enzimas ciclo-oxigenasas COX1 y COX2, disminuyendo la biosíntesis de prostaglandinas y tromboxanos las cuales aumentan la respuesta inflamatoria y por tanto el riesgo de carcinogénesis.

## 6. Mecanismo de carcinogénesis gástrica

El cáncer gástrico se desarrolla como resultado de la sucesión de una serie de lesiones precursoras en la mucosa gástrica. Este proceso fue descrito a partir de observaciones epidemiológicas en poblaciones colombianas con alta (montañas andinas 150/100 000 habitantes) y baja incidencia (2.8/100 000 habitantes) de CG y antes de la identificación de *H. pylori* como un carcinógeno. Mediante un registro de cáncer poblacional en la

ciudad de Cali que poseía un 65.8% de inmigrantes se demostró que la tasa de incidencia de CG era 5 veces mayor en inmigrantes de los Andes respecto a los de la costa del Pacífico. La prevalencia de la metaplasia intestinal siguió un comportamiento análogo a la del CG en dichas poblaciones <sup>14</sup>.

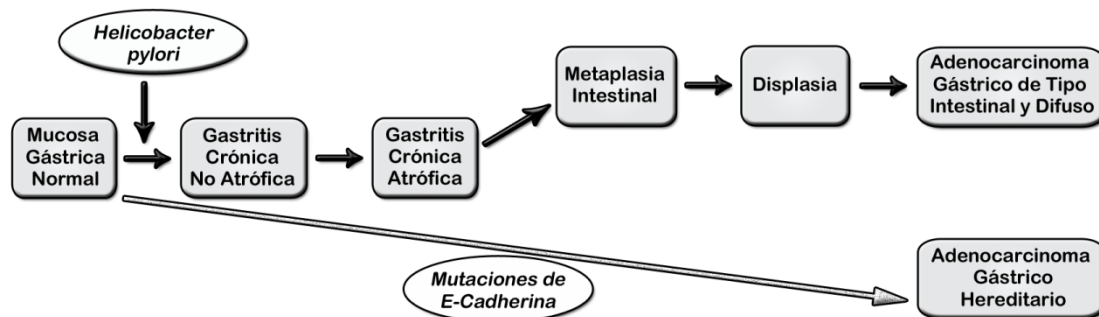
En estudios de seguimiento en dichas poblaciones la severidad de una lesión precursora se correlaciona con la prevalencia de la posterior, indicando una secuencia de eventos. En la tabla 1 se observan las tasas de transición entre la primera y segunda biopsias por 100 personas-año, basado en un seguimiento de 1422 pacientes durante 5 años en un área colombiana de alto riesgo <sup>14</sup>. Se observa que la dinámica de las lesiones sigue un patrón de progresión a lesiones avanzadas y regresión a más leves, siendo más frecuentes las primeras, con excepción de la progresión desde metaplasia a displasia (DISP). Esta dinámica resulta en un lento movimiento no constante hacia la progresión, que difiere entre individuos y está condicionado por la edad (las tasas de progresión aumentan), el muestreo de biopsias gástricas y probablemente por la exposición a otros factores <sup>14</sup>.

<b>Tabla 1. Tasas de transición para la progresión y regresión de las lesiones precursoras. Tomado de Correa <i>et al.</i>, 2012 <sup>14</sup>.</b>	
<b>Tipo de evolución</b>	<b>Tasas de transición por 100 personas-año</b>
Mucosa Normal o Gastritis No Atrófica → Gastritis Atrófica	7.5
Gastritis Atrófica → Mucosa Normal o Gastritis No Atrófica	1.7
Gastritis Atrófica → Metaplasia Intestinal	6.7
Metaplasia Intestinal → Gastritis Atrófica	4.4
Metaplasia Intestinal → Displasia	3.2
Displasia → Metaplasia Intestinal	5.7
Metaplasia Intestinal → Displasia en <40 años	2.1
Metaplasia Intestinal → Displasia en >40 años	4.0

Los resultados anteriores fueron usados para generar un modelo de carcinogénesis gástrica conocido como la cascada de Correa en honor a su descubridor y está compuesto por la siguiente sucesión de LPGs previas a la aparición del CG: gastritis no atrófica (GNA) → gastritis crónica atrófica multifocal → metaplasia intestinal completa (MIC) → metaplasia intestinal incompleta (MII) → displasia de bajo grado → displasia de alto grado → adenocarcinoma gástrico <sup>14</sup>. Numerosos estudios en humanos <sup>1 40</sup> y modelos murinos infectados con *Helicobacter felis* y *Helicobacter pylori* y en otros con expresión transgénica o delección de genes con importantes funciones en la fisiología gástrica han validado dicha teoría <sup>41</sup>.

El proceso se inicia cuando *H. pylori* coloniza la mucosa gástrica de la región antral del estómago, evitando la disminución del pH en el fundus y cuerpo del estómago. Dicha infección por lo general provoca una gastritis crónica no atrófica que puede durar

décadas, a menos que el tratamiento erradique la bacteria; con el tiempo la infección puede extenderse a la mucosa oxíntica. La infección severa y prolongada puede dar lugar a la pérdida de tejido glandular causando una gastritis atrófica multifocal. En algunos pacientes, las glándulas perdidas son reemplazadas por otras con fenotipo intestinal llamada metaplasia intestinal. Una pequeña proporción de sujetos con esta lesión pueden progresar a displasia, así mismo una fracción de los mismos desarrollará CG (Figura 10).



**Figura 10.** Mecanismo de la carcinogénesis gástrica. Modificado de Yuasa, 2003 <sup>42</sup>.

El proceso preneoplásico comienza mucho antes de la aparición de los rasgos histopatológicos neoplásicos, con el crecimiento de tejido epitelial preneoplásico como resultado de la proliferación clonal de células mutadas, las cuales están preconditionadas a un posterior desarrollo tumoral. En este sentido, las glándulas gástricas metaplásicas comparten la mutación c.4680insA en el gen *APC* con glándulas displásicas sugiriendo el origen de la displasia a partir de la metaplasia intestinal <sup>40</sup>.

### 6.1. Alteraciones moleculares en el cáncer gástrico

La adquisición de mutaciones somáticas en oncogenes y genes supresores tumorales es una causa establecida del desarrollo del CG y del cáncer en general. Algunas de las alteraciones genéticas asociadas a la transformación a CG intestinal son la inestabilidad de microsatélites, mutaciones en *TP53*, expresión reducida de *CDKN1B* (p27), incremento en la expresión de *CCNE1* (ciclina E), presencia de transcritos aberrantes de *MET* (c-met), inactivación de genes supresores de tumores como *DCC* y *APC* por pérdida de heterocigosidad o mutación, expresión reducida de *TGFBR1* (receptor tipo I de TGF $\beta$ ) y la amplificación de *ERBB2* (HER2) <sup>43</sup>.

Respecto al CG hereditario de tipo difuso han sido identificadas mutaciones somáticas en el gen de la E-cadherina y expresión reducida de dicha proteína, todo lo cual ocasiona ruptura de la unión de la misma con la  $\beta$ -catenina, y una disminución de la adhesión célula-célula <sup>44</sup>. Otras alteraciones son la pérdida de heterocigosidad en la región cromosómica 17p, donde se localiza *TP53*, la pérdida de p27, inhibidor dependiente de



ciclina que controla la transición G1-S del ciclo celular, y la amplificación de los proto-oncogenes *MET* y *K-sam* contribuyen a la progresión y dispersión metastásica<sup>43</sup>.

Las principales características moleculares (*hallmarks* en inglés) del cáncer<sup>45</sup> se observan también en la carcinogénesis gástrica, las cuales son:

A) Mantenimiento de la señalización proliferativa, lo cual se logra por sobre-expresión autóloga de factores de crecimiento y sus receptores; así como, por activación de vías de señalización activadas por factores de crecimiento, debido a mutaciones en supresores tumorales (*Ras*, *PTEN*, *mTOR*, *TGF $\beta$* ) e inactivación de mecanismos de regulación negativa de dichas vías y evadir la acción de supresores tumorales (*TP53* y *RB1*), los cuales regulan negativamente la proliferación de células con daño genómico.

B) Evitar la muerte celular por apoptosis debido a mutaciones en oncogenes (Ej. *MYC*) o alteraciones en los reguladores de la apoptosis, como el programa intrínseco, extrínseco, *BCL2*, *BAX*, *BAK*, *BIM*, *PUMA*, *IGF1/2* y citocromo C.

C) El mecanismo de autofagia (fusión de fagosomas y lisosomas) permite tanto la sobrevivencia como muerte celular en condiciones de estrés, mediante la ruptura de organelos celulares y el uso de los catabolitos resultantes en el metabolismo energético.

D) Permitir la inmortalidad replicativa debido a la expresión de la enzima telomerasa que impide el acortamiento de los telómeros.

E) Estimulación de angiogénesis por la unión de sus inductores (*VEGFA*, angiopoietina y *FGF*) a receptores del endotelio vascular. Diversas células derivadas de médula ósea como macrófagos, neutrófilos, mastocitos y mieloides contribuyen a este proceso y las células progenitoras vasculares migran al tejido neoplásico.

F) Invasión y metástasis por pérdida de expresión e inactivación mutacional de cadherina E (*CDH1*) y por el mecanismo de transición epitelio-mesenquima (regulado por la familia *Snail*, *Twist* y *ZEB1/2*) y a la que contribuyen macrófagos asociados al tumor que secretan metaloproteinasas de matriz y *EGF* facilitando el proceso.

Otros procesos aún no generalizados son:

G) Reprogramación del metabolismo energético o efecto Warburg.

H) Evasión de la destrucción tumoral por el sistema inmune por pérdida de la expresión de HLA o defectos en la función de células *Natural Killers* (NK), linfocitos T citotóxicos (CD8+) y auxiliares (CD4+). Es importante señalar que ciertos eventos afectan múltiples procesos como es el caso de mutaciones en *RAS* y sobre-expresión de *MYC* que influyen en la proliferación, metabolismo energético, angiogénesis, invasión y sobrevivencia.

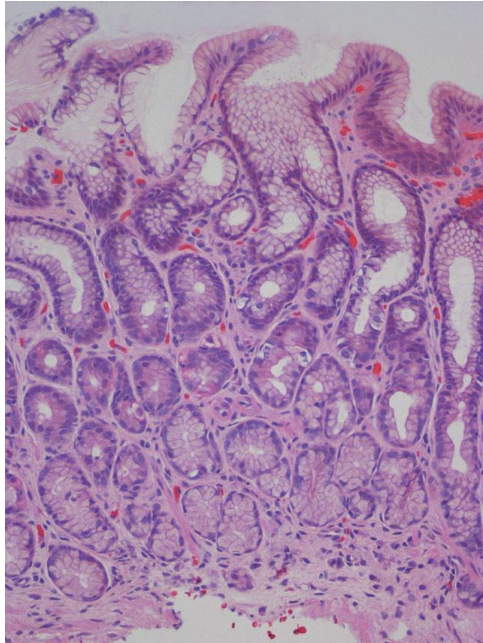
La adquisición de los rasgos antes mencionados es debida a dos características fundamentales: la adquisición de mutaciones endógenas o por carcinógenos exógenos y la inflamación crónica. Esta última es particularmente relevante en la carcinogénesis gástrica a consecuencia de la infección por *Helicobacter pylori* contribuyendo a múltiples *hallmarks* <sup>45</sup>. Adicionalmente a los procesos globales en la carcinogénesis, una serie de vías de señalización y mecanismos moleculares específicos del CG han sido identificados mediante estudios de expresión génica, secuenciación de nueva generación (NGS, del inglés *next generation sequencing*), RNAseq y polimorfismo de nucleótido simple (SNP, del inglés *single nucleotide polymorphism*), los cuales es necesario tener en cuenta en el análisis de las potenciales vías de señalización que condicionan la progresión desde la metaplasia intestinal al CG. En este sentido es relevante la más extensa caracterización genómica del CG realizada por el proyecto internacional Cancer Genome Atlas mediante secuenciación de exoma completo, análisis del número de copias y metilación; determinando patrones moleculares comunes en los CGs según la infección por virus Epstein-Barr, inestabilidad de microsatélites, inestabilidad cromosomal y estabilidad genómica <sup>46</sup>.

## **7. Lesiones precursoras gástricas**

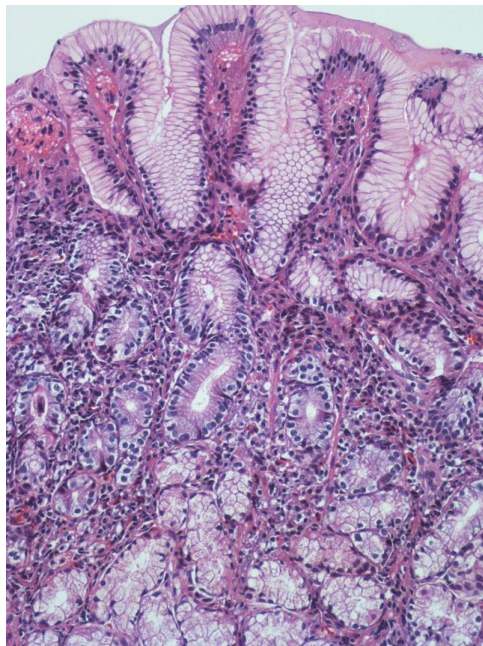
A continuación se describen las LPGs constituyentes de la carcinogénesis gástrica basándose en el sistema de Sydney <sup>47</sup>. Las mismas se localizan inicialmente preferentemente en el antro y zona transicional y con el tiempo se extienden al cuerpo.

### **7.1. Gastritis crónica no atrófica (superficial o activa)**

Gastritis crónica concentrada principalmente en la porción foveolar de la mucosa antral, comenzando justo debajo de la superficie del epitelio. Consiste en la infiltración incrementada en la lámina propia de leucocitos mononucleares representativos de la inflamación crónica (linfocitos T y B, macrófagos, células NK, células plasmáticas, eosinófilos y mastocitos) y de neutrófilos polimorfonucleares característicos de la inflamación aguda. Esta última, así como los agregados linfoides formando centros germinales, son signos distintivos de la infección por *H. pylori* <sup>14</sup>. Tiene una alta tasa de curación si el *H. pylori* es eliminado y de remisiones espontáneas. En las figuras 11 y 12 se observan la mucosa normal y la gastritis crónica no atrófica, respectivamente.



**Figura 11.** Mucosa normal teñida con hematoxilina-eosina. Tomado de Correa *et al.*, 2012<sup>14</sup>.

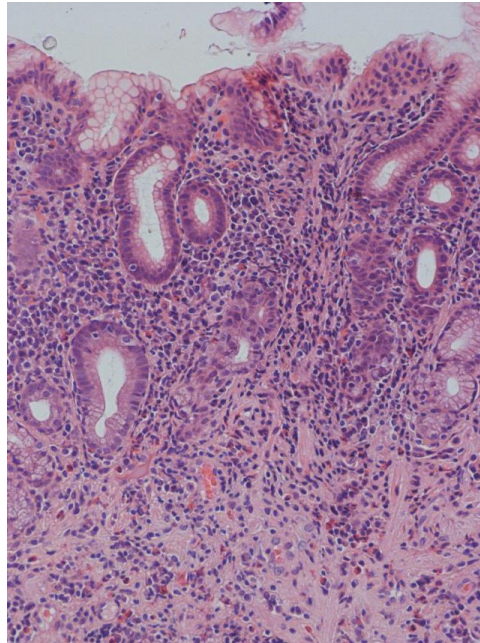


**Figura 12.** Gastritis crónica no atrófica. Tomado de Correa *et al.*, 2012<sup>14</sup>.

## 7.2. Gastritis atrófica multifocal sin metaplasia intestinal

La gastritis atrófica se caracteriza por la pérdida de glándulas antrales y/o oxínticas (Figura 13). En el caso de pérdida de la mayoría de las glándulas gástricas, se denomina gastritis atrófica multifocal o pangastritis, involucrando al antro y al cuerpo gástrico. El diagnóstico de la gastritis atrófica presenta una muy baja concordancia inter-observador ( $k < 0$ )<sup>48</sup>. La pérdida de estas glándulas se acompaña por fibrosis y aparición de glándulas

metaplásicas; sin embargo, la atrofia y la metaplasia intestinal pueden existir independientemente.

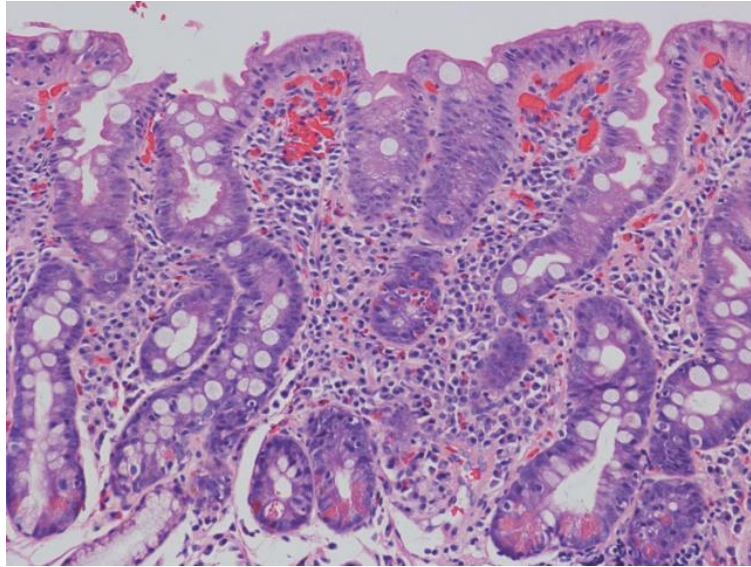


**Figura 13.** Gastritis crónica atrófica. Tomado de Correa *et al.*, 2012 <sup>14</sup>.

Al desaparecer las células parietales disminuye la secreción ácida del estómago (hipoclorhidria), la cual puede provocar la colonización masiva por parte de bacterias y hongos representativos de la flora normal intestinal. Las mismas producen acetaldehído y compuestos N-nitroso, potencialmente carcinogénicos que pueden producir mutaciones en el ADN <sup>49</sup>.

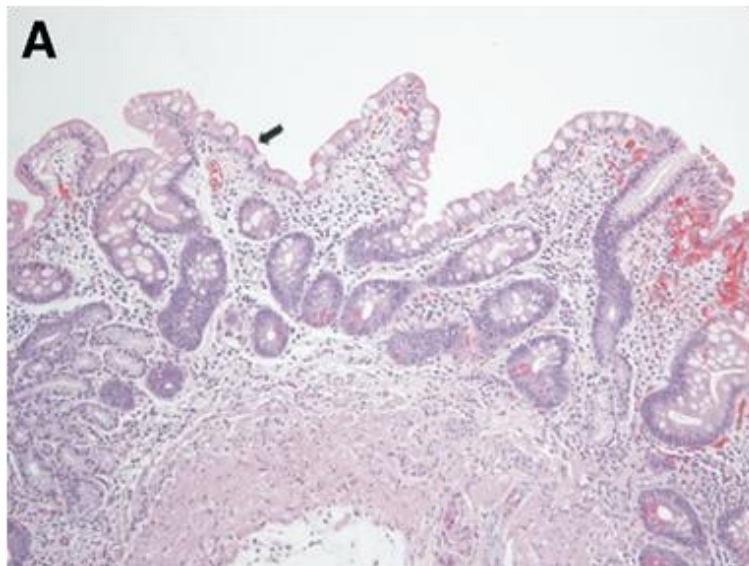
### 7.3. Metaplasia intestinal

La metaplasia intestinal consiste en el reemplazo de la mucosa gástrica normal por otra histológicamente similar a la del intestino, compuesta mayoritariamente por células caliciformes (*goblet cells*) secretoras de mucus. Es la LPG con mayor concordancia inter-observador ( $k=0.81$ ) <sup>48</sup>. Existen 2 tipos, la completa y la incompleta <sup>50</sup>. La MIC está constituida por células epiteliales del intestino delgado, como células caliciformes bien desarrolladas secretoras de sialomucinas, enterocitos absortivos que expresan todas las enzimas digestivas (sucrasa, trehalasa, fosfatasa alcalina) y células de Paneth (Figura 15). En la MIC existe una disminución o ausencia de la expresión de las mucinas gástricas MUC1, MUC5AC y MUC6 y la expresión de la mucina intestinal MUC2. Predomina en sujetos jóvenes <sup>14</sup>.



**Figura 14.** Metaplasia intestinal completa. Tomado de Correa *et al.*, 2012<sup>14</sup>.

La MI está compuesta por células epiteliales del colón, específicamente células columnares y caliciformes de tamaño variable secretoras de sialomucinas y sulfomucinas, ausencia de enterocitos y rara presencia de células de Paneth. Presenta una importante desestructuración glandular y desaparición parcial o total de las enzimas digestivas (Figura 15). Existe expresión de mucinas gástricas o neutrales MUC1, MUC5AC, MUC6 y de la mucina intestinal MUC2. Predomina en pacientes mayores de edad<sup>24</sup>.



**Figura 15.** Metaplasia intestinal incompleta. Tomado de Correa *et al.*, 2007<sup>24</sup>.

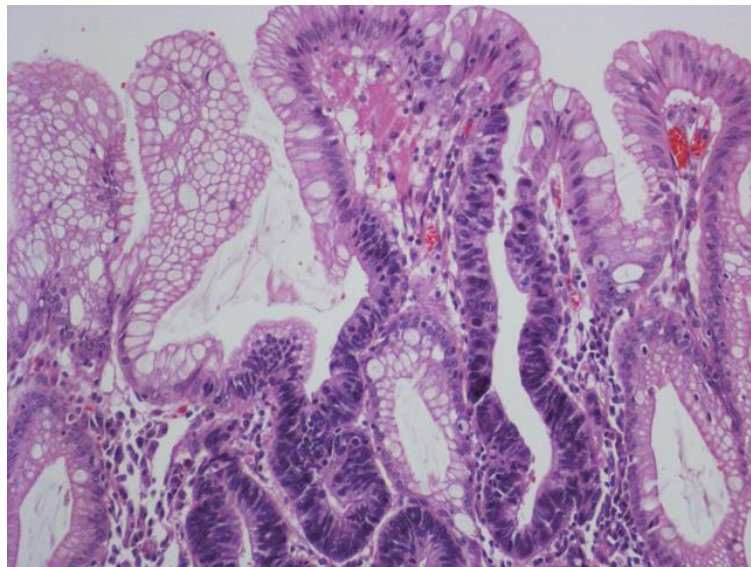
La conexión causal entre la metaplasia intestinal con el CG fue propuesta por vez primera en Java y Sumatra en 1938, comparando pacientes malayos con bajas frecuencias de CG, atrofia y MI, respecto a inmigrantes chinos con altas frecuencias de tales patologías.

Posteriormente, Jarvi y Lauren reportaron que los CGs de tipo intestinal estaban rodeados por metaplasia intestinal más frecuentemente que los CGs de tipo difuso <sup>13</sup>.

El concepto general de metaplasia es la conversión postnatal de células diferenciadas en células madres y su posterior transformación en tipos celulares característicos de otros órganos <sup>51</sup>. La metaplasia ocurre en tejidos expuestos a traumas, inflamación, infección o estimulación hormonal, estando expuestos a continua proliferación. Ejemplos en otros tejidos son la metaplasia escamosa (glandular a escamosa) debido al tabaquismo y el esófago de Barret (escamosa a glandular) causado por reflujo gastroesofágico, caracterizado porque la mucosa esofágica es expuesta a ácidos biliares y gástricos induciendo una inflamación crónica <sup>52</sup>. La metaplasia es considerada una respuesta protectora contra el agente dañino, sin embargo crea una susceptibilidad para la transformación neoplásica, ya que la mayoría de las metaplasias son lesiones precursoras de cáncer <sup>51</sup>.

#### 7.4. Displasia

Lesión neoplásica limitada al epitelio, sin invasión de la lámina propia, caracterizada por atipia celular, esto es, células de formas irregulares, hipertróficas e hiper cromáticas, y en las que se pierde la polaridad apico-basal epitelial; además, hay presencia de frecuentes mitosis y nucléolos prominentes (Figura 16). Esta des-diferenciación del epitelio es progresiva y se convierte en carcinoma invasivo cuando la atipia celular alcanza la membrana basal <sup>14</sup>. Su diagnóstico presenta una baja concordancia inter-observador ( $k=0.42$ ) <sup>48</sup>.



**Figura 16.** Displasia gástrica. Tomado de Correa *et al.*, 2012 <sup>14</sup>.

### 7.5. Epidemiología de las lesiones precursoras gástricas

La prevalencia de las LPGs muestra diferencias dependientes de las zonas geográficas, la incidencia de la infección por *H. pylori*, edad, sexo, exposición a factores de riesgo y si la población de estudio son pacientes que consultan por dispepsia o proceden de la población general<sup>53 54 55</sup>. La prevalencia de las LPGs en dispépsicos oscila en: atrofia (7.8-80%), metaplasia intestinal (5-72%) y displasia (0-20%), siendo menores en no dispépsicos o asintomáticos, aunque también significativamente altos (9.3% en total con LPGs)<sup>55</sup>.

Holanda posee un registro de base poblacional de múltiples enfermedades que ha permitido la realización de múltiples estudios epidemiológicos de las LPGs. En uno de ellos (1991-2005) se demostró que la incidencia de las mismas está disminuyendo quizás debido a la disminución de la infección por *H. pylori*. En el caso de la atrofia gástrica esta disminución es de un 8.2%/año, para la metaplasia 2.9%/año y la displasia 8.1%/año. Basado en ello se puede anticipar una disminución de hasta el 24% en la incidencia de CG en los próximos 10 años sin intervención específica<sup>53</sup>.

En contraposición a lo anterior otro grupo holandés plantea que el CG permanecerá siendo una enfermedad relativamente frecuente en países occidentales, basado en la prevalencia de LPGs en población general sana o asintomática clínicamente<sup>55</sup>. Basado en lo anterior, proponen un programa de vigilancia para poblaciones de bajo riesgo de CG que consistiría en la detección de factores de riesgo epidemiológicos y pesquisaje serológico (anticuerpos anti *H. pylori* y razón de pepsinógenos PGI/PGII). Como resultado de dicho cribado la vigilancia gastroscópica debe ser ofrecida a pacientes con LPGs infectados por *H. pylori*<sup>54 55 56</sup>.

### 7.6. Estudios de seguimiento de lesiones precursoras gástricas

Se han realizado numerosos estudios de seguimiento de lesiones precursoras gástricas con el objetivo de conocer las tasas de progresión a CG desde las mismas. Un artículo de revisión compendia las tasas anuales mundiales para la gastritis atrófica (0-1.8%), metaplasia intestinal (0-10%) y displasia (0-73%). La elevada variación en los rangos es atribuida a diferencias en el estilo de vida, prevalencia de *H. pylori*, sistemas de muestreo de biopsias gástricas y clasificación histológica, criterios de exclusión así como tiempos de seguimiento diferentes<sup>56</sup> (Tabla 2). Las principales conclusiones de dichos estudios es que las tasas de progresión a CG aumentan con la severidad de las lesiones según la cascada de Correa. La MI incompleta (tipos II y III) tiene mayor riesgo que la completa (tipo I)<sup>57</sup> y la displasia de alto grado/severa posee el mayor riesgo entre todas las LPGs (Tabla 2).

En los 2 estudios de seguimiento de LPGs desarrollados por nuestro grupo (provincia de Soria y multicéntrico español) hemos observado que la metaplasia intestinal incompleta confiere el mayor riesgo de progresión a CG<sup>57</sup>. Entre todas las LPGs la metaplasia intestinal es la lesión con mayor concordancia inter-observador ( $k=0.81$ )<sup>48</sup> por poseer

características histológicas fácilmente identificables, y altas tasas de progresión a CG. La otra lesión de interés para el seguimiento es la displasia, sin embargo en nuestra población tiene una prevalencia muy baja (0-4%) y diversos estudios reportan bajas tasas de progresión (0%) y altas de regresión (69.2, 100%)<sup>48</sup>, lo cual concuerda con la mayor tasa de regresión para la transición DISP-MI respecto a MI-DISP previamente reportada<sup>10</sup> (Tabla 1). Basado en lo anterior nos hemos centrado en la metaplasia intestinal como la lesión de mayor riesgo a CG en la población española y por ende su justificación para ser usada en estudios moleculares. Recientemente han sido publicadas guías de manejo clínico de las LPGs, en poblaciones que no son de alto riesgo, las que recomiendan vigilancia endoscópica cada 1-3 años en pacientes infectados y diagnosticados con una metaplasia intestinal extensa, lo cual apoya nuestra hipótesis de estudio<sup>58</sup>. Un programa masivo de pesquizaje mediante gastroscopias en población general de bajo riesgo es inapropiado debido al peligro para los pacientes, objeciones éticas y costes<sup>53</sup>.

La ausencia de una vacuna preventiva contra *H. pylori*, el bajo éxito de la quimioterapia, y la pequeña tasa de sobrevivencia a los 5 años (20%) del CG con diagnóstico tardío, respecto al 60% cuando es detectado precozmente, apuntan a que la temprana detección de las LPGs es decisivo en su pronóstico. En este sentido, la identificación de factores que permitan saber cuáles lesiones progresarán a CG es de gran importancia.



Tabla 2. Principales estudios epidemiológicos de seguimiento de lesiones precursoras gástricas. Tomado y modificado de de Vries <i>et al.</i> , 2007 <sup>56</sup> (con excepción de los estudios de los estudios españoles).							
País	Diseño de estudio	Tamaño de muestra	Tiempo de seguimiento después de endoscopia (años)	Progresión a CG (%)	Progresión a CG (% de todos los estudios)	Incidencia anual de CG (n/100.000 personas años)	Incidencia anual de CG (% de estudios)
<b>Gastritis Atrófica</b>							
Finlandia	Prospectivo	116	15 (máximo)	9 (7.8%)	3.48	≥517	208.6
			23 (máximo)	10 (8.6%)		≥375	
Suecia	Prospectivo	61	2.7 (%)	0 (0%)		0	
China	Prospectivo	1240	5 (máximo)	Ligera: 1 (0.1%) Severa: 0 (0%)		≥ 16	
						0	
Italia	Prospectivo	42	4 (máximo)	0 (0%)		0	
Italia	Prospectivo	106	7 (media)	1 (0.9%)		≈135	
<b>Metaplasia Intestinal</b>							
Italia	Prospectivo	261	9 (%)	12 (4.6%)	11.53	≈ 511	572
Portugal	Prospectivo	124	6 (máximo)	Tipo I-III: 0 (0%)		0	
China	Prospectivo	842	5 (máximo)	Superficial: 2 (0.8%) Severa: 16 (2.7%)		≥ 156	
						≥ 546	
UK	Prospectivo	93	10 (máximo)	10 (11%)		≥ 1075	
España (Soria) <sup>57</sup>	Prospectivo	478	12.8	(MII-21%, MIC 19%)		Global: 3.77 (2.51-5.67)/1000 MII: 16.5 (10.08-6.86)/1000, MII vs MIC (HR 11.3 [3.8-33.9])	
España (multi-céntrico)	Prospectivo	649	10	MIC: 3.23% MII: 6.85%		Global: 3.09 (2.07-4.6)/1000 MIC: 2.76/1000, MII: 5.76/1000 MII vs MIC: HR 2.75 (1.06-6.26)	
<b>Displasia</b>							
Italia	Prospectivo	134	1.6 (%)	Ligera: 5 (6.1%) Moderada: 7 (22.6%) Severa: 6 (46.2%)	64.56	≥ 3811	19 876
						≥ 14.113	
						≥ 28.846	
Italia	Prospectivo	125	10 (máximo)	Bajo grado: 22(27.2%) Alto grado: 36 (81.8%)		≥ 2716	
						≥ 8182	
China	Prospectivo	546	5 (máximo)	Ligera: 14 (2.8%) Moderada/Severa: 3 (7.0%)		≥560	
						≥ 1400	

## **8. Metaplasia intestinal**

### **8.1. Riesgo de cáncer gástrico según subtipo histológico, localización anatómica y extensión de la metaplasia intestinal**

Teniendo en cuenta la relevancia de la MI en la progresión a CG se ha explorado su riesgo diferencial en función del subtipo histológico (completa/incompleta). Algunos de los primeros estudios ya indicaron que la metaplasia intestinal incompleta es la que presenta un riesgo más elevado de cáncer gástrico <sup>50</sup> aunque esta asociación no se confirma en todos los estudios <sup>59</sup>. A pesar de la controversia existente al respecto, una reciente revisión de estudios que evalúan la prevalencia o el efecto del tipo de MI ha demostrado la posible utilidad del subtipaje histológico como marcador de riesgo de CG. La variable resultado investigada fue la prevalencia de los subtipos de MI (completa/incompleta) en el tejido adyacente al tumor (CG) y en el grupo control en los estudios transversales; para los estudios de seguimiento fue la comparación de la incidencia de CG de dichos subtipos al final del seguimiento. Los autores encontraron que en 13 de 14 estudios transversales la prevalencia de la MII fue significativamente mayor en el CG que en otras lesiones gástricas controles <sup>60</sup>. Aunque este tipo de estudios no es aconsejable para encontrar factores pronósticos, si confirman que la MII es la LPG más frecuente hallada en el CG.

Lo más importante es que 6 de 10 estudios de seguimiento (prospectivos, retrospectivos, prospectivos-retrospectivos) encontraron una asociación entre la MII (tipo III) y el riesgo posterior de CG, siendo 4-11 veces mayor para la tipo III respecto a la completa o a la ausencia de tipo III <sup>60</sup>. Estos estudios constituyen los mejores para probar la asociación entre el tipo de MI con el riesgo de CG. Las asociaciones fueron halladas en países con alto y moderado riesgo de CG. Tres de los estudios que demostraron asociación fueron los que poseían un mayor tamaño de muestra y en los que el seguimiento fue más prolongado, factores claves para hallar una asociación significativa.

Adicionalmente en un estudio prospectivo, con una media de seguimiento de 12 años y realizado en 9 hospitales de toda España las tasas de incidencia de CG son mayores en la MII (5.76/1000) respecto a la MIC (2.76/1000), (HR 2.75 [1.06-6.26]), progresando a CG el 6.85% y 3.23% de las MII y MIC, respectivamente (pendiente de aceptación). Previamente un estudio de seguimiento en la provincia de Soria detectó un mayor riesgo de progresar a CG desde la MII respecto a la MIC (HR 11.3 [3.8-33.9]), progresando a CG el 21% y 19% de las MII y MIC, respectivamente <sup>57</sup>. Otro estudio de seguimiento (1989-1990, N=3399) realizado en China encontró que el 13.3% de los pacientes con MII desarrollaron CG en comparación con solo el 1.9% de las MIC <sup>61</sup>.

Las guías de manejo de LPGs no aconsejan el subtipaje de la MI en la práctica clínica, debido a que sus autores consideran que la evidencia de su utilidad es “limitada e inconsistente” y requiere del uso de técnicas inmuno-histoquímicas no comunes<sup>58</sup>. Sin embargo, ambos argumentos son discutibles ya que el estudio anterior muestra la asociación diferencial entre el tipo de MI y el riesgo de CG<sup>60</sup> y el diagnóstico histológico no estaba basado en técnicas de inmunohistoquímica sino en la morfología, utilizando la simple y rutinaria técnica de tinción con hematoxilina-eosina.

Se considera que la localización anatómica de la metaplasia intestinal también proporciona un valor predictivo del riesgo futuro de CG, siendo más elevado en el cuerpo<sup>48</sup>. Debido a que la gastritis atrófica avanza desde el antro hacia el cuerpo la localización en el cuerpo indica un estadio de mayor severidad respecto al antro. Sin embargo, basar el diagnóstico de riesgo de progresión en la extensión de la MI requiere múltiples tomas de biopsias, que no se practican rutinariamente en la clínica. La extensión de la MI también condiciona un mayor riesgo de CG<sup>62</sup>.

## 8.2. Alteraciones moleculares en la metaplasia intestinal

### 8.2.1. Factores de transcripción tipo caudal homeobox *CDX1* y *CDX2*

La transdiferenciación del epitelio gástrico por *CDX1* y *CDX2* es el principal mecanismo causal directo de la metaplasia intestinal. *CDX1* y *CDX2* son los *master genes* encargados de este proceso, factores de transcripción homeobox tipo caudal que regulan el patrón de desarrollo antero-posterior del tracto gastrointestinal<sup>52</sup>. Los mismos no se expresan en el epitelio gástrico normal y son inducidos por *cagA*, ratones transgénicos de *CDX1* y *CDX2* desarrollan MI y progresan a CG<sup>41</sup>. La expresión de *CDX1/2* es responsable de la conversión del epitelio gástrico en metaplásico mediante la inducción de factores de reprogramación de desdiferenciación *SALL4* y *KLF5*<sup>63</sup> y marcadores de diferenciación intestinal como la sacarasa isomaltasa, cadherina LI, *MUC2* y *FURIN*<sup>64</sup>. Además, *CDX2* es un inhibidor de la reparación del ADN no homóloga en cáncer de colon<sup>52</sup>, por lo que es probable que en estómago también demuestre esta función.

Liu Q propone que un nivel muy alto de *CDX2* sería causal de la MIC, la cual es un estadio completamente diferenciado con una baja tasa de progresión a CG, mientras que un nivel menor sería responsable de la MII la cual es un estadio de menor diferenciación y por ende más proclive de avanzar a CG<sup>65</sup>.

*CDX1* y *CDX2* activan la expresión de diversos genes de diferenciación intestinal como sacarasa isomaltasa (*SI*), glucagón (*GCG*), lactasa phlorizin hidrolasa (*LCTL*), calbindin D9k (*S100G*), anhidrasa carbónica I (*CA1*), fosfolipasa A (*PLA1A*), receptor de vitamina D (*VDR*), guanilil ciclasa C, pancreatitis-associated protein I (*PAP I*),

claudina 2 (*CLDN2*), LI-cadherina (*CDH17*), proliferating cell nuclear antigen (*PCNA*), MUC2, UDP-glucuronosyltransferases (*UGT1A8* y *UGT1A10*),  $\beta$ -1,3-galactosyl transferase T5 (*B3GALT5*), glucosa 6 fosfatasa (*G6PC*), resistin-like molecule/found in inflammatory zone (*RELM/FIZZ*) y fosfatasa alcalina intestinal (ALPI). Se ha observado una cooperación entre los factores de transcripción CDX1/2, HNF1A/B (hepatocyte nuclear factor) y la familia GATA para activar la expresión de algunos de estos genes <sup>64</sup>.

### **8.2.2. Daño oxidativo**

En la metaplasia intestinal existen niveles incrementados de marcadores de daño oxidativo al ADN como el 8OHdG <sup>66</sup>, indicativos de la presencia de especies reactivas del oxígeno y el nitrógeno a consecuencia de la infección por *H. pylori*. Dichas especies causan mutaciones en oncogenes, supresores tumorales y genes del mecanismo de reparación del ADN. Adicionalmente, en la MI hay una disponibilidad limitada de secuestradores de dichas especies como el glutatión y la glutatión S-transferasa lo cual contribuye a la persistencia de altos niveles de las mismas en la metaplasia intestinal <sup>67</sup>.

### **8.2.3. Supresores tumorales RUNX3 y TP53**

RUNX3 es un factor transcripcional regulador del crecimiento y la diferenciación del epitelio gástrico que ha sido identificado como supresor tumoral. Se encuentra sub-expresado en la MI y el CG debido a delección, hipermetilación de su promotor, inadecuada localización subcelular proteica y degradación proteasomal inducida por *H. pylori* <sup>68</sup>.

TP53 es un factor transcripcional que controla la reparación del ADN, la apoptosis y el ciclo celular. Se reporta la presencia de mutaciones en el 37.5%, 58.3% y 66.7% de las muestras con MI, DISP y CG, indicando que pueden ser un evento temprano y están correlacionadas con la severidad de las LPGs <sup>69</sup>.

### **8.2.4. Vía de señalización celular Sonic hedgehog**

La vía Sonic hedgehog (Shh) participa en la regulación del desarrollo embrionario del estómago, la proliferación celular y la diferenciación de glándulas gástricas. La inhibición de esta vía conlleva a un programa de diferenciación alterado y pérdida de la secreción ácida gástrica; así como, a un incremento de la proliferación epitelial y una reducción en la expresión de proteínas reguladoras de la expresión específica de tejido y la diferenciación. Shh se expresa en el estómago sano y disminuye en la MI, siendo menor en la MII en comparación con la MIC <sup>70</sup>.

### 8.2.5. Mecanismos epigenéticos

La metilación de islas CpG en la MI disminuye la expresión de los genes *MLH1*, *p14*, *p15*, *p16*, *CDH1*, *RUNX3*, *MGMT*, *THBS1* y *TIMP3*; siendo parcialmente reversible después de la erradicación de *H. pylori*, demostrando que esta bacteria interfiere con mecanismos de regulación epigenética en el epitelio gástrico <sup>71</sup>.

### 8.2.6. Mecanismos de reparación del ADN. Pérdida de heterocigosidad e inestabilidad de microsatélites

Mutaciones en los mecanismos de reparación del ADN permiten la acumulación de mutaciones en el epitelio gástrico infectado por *H. pylori*, como las producidas en regiones de microsatélites con la consiguiente variación en la longitud de los mismos (MSI, del inglés *microsatellite instability*), característica utilizada actualmente en la clasificación molecular del CG. Se reportó MSI en al menos 1 de 9 loci en el 27% de las MI, siendo todas del tipo incompleto y observándose el mismo nivel de MSI en la muestra tumoral circundante <sup>72</sup>. En un trabajo se halló MSI presente en la GNA (13%), MI (20%), DISP (25%) y CG (38%), evidenciando un incremento de su presencia acorde a la severidad de las LPGs en la cascada de Correa <sup>73</sup>.

### 8.2.7. Aumento de actividad telomerasa

El nivel de activación de la enzima telomerasa incrementa según la severidad de las LPGs siendo mayor en la MII que en la MIC <sup>74</sup>.

En resumen, una serie de mecanismos moleculares caracterizan a la MI y algunos de ellos avalan el resultado epidemiológico de un mayor riesgo de progresión a CG de la metaplasia intestinal incompleta respecto a la completa. Estos son, menor expresión de *CDX2* en la MII vs MIC <sup>65</sup>, menor activación de la vía de señalización Sonic hedgehog en la MII <sup>70</sup>, mayor inestabilidad de microsatélites <sup>72</sup>, mayor actividad telomerasa en la MII <sup>74</sup>; así como, la presencia de *H. pylori* de localización intracelular el cual incrementa la severidad inflamatoria, con mayor frecuencia en la MII <sup>75</sup>.

## 8.3. Estudios de expresión génica por microarrays en metaplasia intestinal

Los estudios de expresión génica por microarrays en cáncer han permitido identificar oncogenes, supresores tumorales, marcadores de progresión clínica, y de la respuesta a la quimioterapia en el CG <sup>76</sup>. Esta técnica permite detectar la expresión de transcritos por medio de sondas unidas a un soporte sólido. La misma ha sido aplicada al estudio de la metaplasia intestinal, gastritis y CG con el objetivo de conocer las vías de señalización celular que se encuentran desreguladas.

Meireles *et al.* estudiaron muestras de mucosa gástrica normal, gastritis, MI y CG (intestinal y difuso), hallando un conjunto de genes diferencialmente expresados durante la progresión de las LPGs acorde a la cascada de Correa. Aquellos sobre-expresados en la metaplasia intestinal respecto a mucosa sana son COL4A1, FN1, CTSB, COL1A2, Hs.177781, DAF, y VIM; la mayoría (subrayados) de los cuales poseen funciones relacionadas con la matriz extracelular. Coincidentemente, Zhao *et al.* hallaron que las vías de señalización celular que distinguen las LPGs del CG agrupan genes de la matriz extracelular y adhesión focal integrada principalmente por colágenos<sup>77</sup>.

En otro estudio se identificó que genes del metabolismo lipídico (*AKR1B10*, *ALDH3A2*, *ADH1B*, *CDS1*, *DGKQ*) se encuentran sobre-expresados en MI y que su patrón de expresión es similar al esófago de Barret, la lesión metaplásica esofágica<sup>78</sup>. En concordancia con lo anterior, otro estudio identificó en la metaplasia intestinal una expresión desregulada de genes del metabolismo lipídico (*MTTP*, *FAT*) y, adicionalmente, otros intestinales (*CDX1*) y con funciones gástricas (*CCK*, *TFF1*, *VIL1*)<sup>79</sup>.

De manera reiterada se encontró que en la MI existe una expresión incrementada de genes del metabolismo de lípidos (*MTTP*, *FABP1*, *APOA1*) y la diferenciación intestinal (*CDX1/2*, *HNF4A*), además de transportadores de aminoácidos (*GGT1*) y glucosa (*ALDOB*), proteínas integrales de membrana y del citoesqueleto del epitelio intestinal (*CLDN3*, *CLDN4*, *CDH17* y *VIL1*)<sup>80</sup>. Otro trabajo fue realizado con MI aislada de la periferia de CG intestinal, y comparado el patrón de expresión génica con células principales de sujetos sanos. Demostraron una sobre-expresión génica y proteica de *ACE2*, *MUC13*, *CDH17*, *REG4*, *FABP1*, *LYZ*, *OLFM4*, *MUC5AC*, *KRT20*, *LGALS4*, *AKR1B10* y *DEFA5* respecto a la mucosa normal<sup>81</sup>. Estudios de expresión no microarrays evidencian que *OCT1*, *PDX1*, *TFF3* y *MATH1* se encuentran sobre-expresados mientras que *SOX2*, Sonic hedgehog, *RUNX3*, *TFF1* y *TFF2* disminuyen parcial o totalmente su expresión en la MI<sup>52</sup>.

Las conclusiones de estos estudios es que existe un patrón de expresión génica característico de la metaplasia intestinal caracterizado por la sobre-expresión de un conjunto de genes epiteliales y del estroma gástrico. Los genes sobre-expresados en la MI respecto a la mucosa sana representan a los procesos moleculares de la diferenciación intestinal (*CDX1/2*), moléculas del cito-esqueleto (*CLDN3*, *CLDN4*, *CDH17*, *VIL1*) y mucus intestinal (mucinas, *trefoil factors*), metabolismo lipídico (*MTTP*, *FABP1*, *AKR1B10*, *ALDH3A2*, *ADH1B*, *APOA1*, *CDS1*, *DGKQ*) y adhesión focal mediada por colágenos (*COL1A1*, *COL11A1*, *COL1A2*, *COL4A1*). Es importante destacar como limitantes que en estos estudios se analiza la MI sin diferenciar en sus subtipos histológicos (completa/incompleta) y no se conoce si las mismas evolucionan a CG durante un seguimiento clínico.

#### 8.4. La familia multigénica Schlafen

Los genes Schlafen (dormir en alemán) constituyen una familia multigénica con un alto grado de homología de secuencia, similitud funcional y proximidad cromosomal. Su función es la regulación de la respuesta inmune mediante el control de la diferenciación de timocitos a células T, así como la activación de las mismas <sup>82</sup>. En ratón esta familia está compuesta por 3 grupos basados, entre otras características, en el tamaño de sus proteínas <sup>82</sup>: grupo I (37-42 KDa, Schlafen1 y 2), grupo II (58-68 KDa, Schlafen3 y 4) y grupo III (100-104 KDa, Schlafen5, 8, 9, 10, 14). Todos sus miembros contienen un dominio Slfn de función desconocida, el cual no ha sido encontrado en otras proteínas. Además, contienen un dominio N-terminal AAA con una función hipotética en el metabolismo del ácido ribonucleico (ARN) y la unión a GTP/ATP. Los grupos II y III poseen otro dominio llamado SWADL (Ser-Trp-Ala-Asp-Leu). La localización subcelular de los miembros de los grupos I y II es el citoplasma, mientras que los del grupo III lo hacen en el núcleo debido a una señal de localización nuclear (RKRRR) <sup>82</sup>.

Los genes de los grupos I y II fueron los primeros descubiertos mediante técnicas de hibridación sustractiva para conocer genes reguladores de la diferenciación de timocitos a células T <sup>82</sup>. Posteriormente, usando técnicas similares aplicadas a la infección por *Lysteria monocitogenes* fue descubierto el grupo III. En humanos existen cinco isoformas (Slfn5, 11, 12, 13 y 14), a diferencia de en el ratón solo un miembro tiene localización citoplasmática, Slfn12, el cual carece del dominio helicasa y pertenece al grupo II, los restantes pertenecen al grupo III. La mayoría de los Schlafen murinos no presentan homólogos en humanos y viceversa <sup>82</sup>.

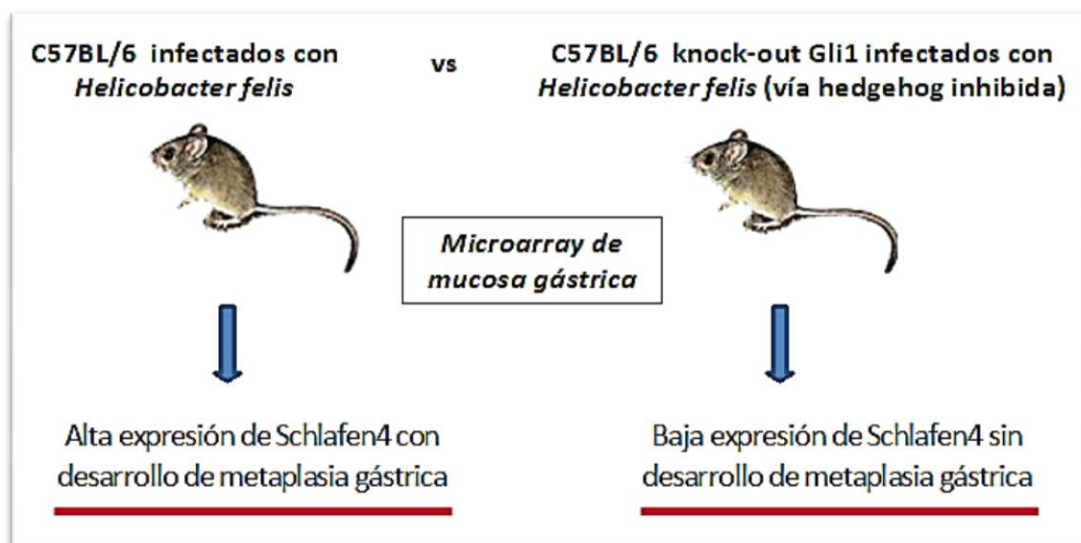
Los genes Schlafen tienen alta expresión durante la diferenciación de timocitos a células T, en células T periféricas inactivas y debido a la infección por bacterias *Brucella* o *Listeria* <sup>82</sup>. Una disminución de su expresión ocurre después de la activación de las células T por medio de una célula presentadora de antígenos <sup>83</sup>. Las funciones de la familia Schlafen incluyen permitir el desarrollo de las células T a partir de timocitos y, además, inducir inhibición del crecimiento de las mismas y otros tipos celulares como fibroblastos por arresto del ciclo celular. Schlafen1 y 2 inducen dicho estado en las células T mediante inhibición de la ciclina D1 <sup>83</sup>.

La similitud funcional de la familia Schlafen se confirma en múltiples estudios. En este sentido, la expresión exógena de Schlafen1 en timocitos y fibroblastos produce un arresto del ciclo celular de las células T <sup>83</sup>. Los knockdowns de Schlafen2 y 3 en melanoma resultan en una proliferación celular incrementada, implicando a estas proteínas como inhibidores de la proliferación celular <sup>84</sup>. La mutación elektra (I135N) en Schlafen2 ocasiona que los ratones sean más susceptibles a infecciones virales y bacterianas debido a que las células T con esta mutación sufren apoptosis en respuesta a un estímulo proliferativo <sup>82</sup>. Schlafen3 es un regulador del desarrollo,

diferenciación y maduración del intestino, además de participar en la diferenciación de las células T y poseer efectos antiproliferativos en cáncer de colon<sup>82</sup>. Schlafen11, miembro del mismo grupo III al que pertenece Schlafen5, es responsable del arresto del ciclo y la muerte celular en líneas celulares tumorales NCI-60 tratadas con quimiofármacos<sup>118</sup>.

### 8.5. Rol de los genes Schlafen en la metaplasia intestinal

Ratones *knock-out* para el ligando GLI1 de la vía de señalización hedgehog (*GLI1*<sup>-/-</sup>) infectados con *Helicobacter felis* no desarrollan metaplasia en contraposición a ratones salvajes infectados con la misma bacteria. Al comparar el nivel de expresión génica en la mucosa gástrica de ambos ratones mediante un microarray de ARN mensajero (ARNm) se observó que el gen Schlafen4 se encontraba sobre-expresado en la metaplasia tipo SPEM (*spasmolytic polypeptide expressing metaplasia* en inglés), y que el mismo era producido por una población de células mieloides que co-expresaban citoquinas pro-inflamatorias<sup>85</sup>. Este estudio evidencia que la activación de la vía hedgehog induce el reclutamiento de células supresoras derivadas de línea mieloides (MDSC, del inglés *myeloid derived suppressor cells*) hacia el estómago murino infectado por *Helicobacter* y que el gen Schlafen4 pudiera ser un biomarcador de la metaplasia tipo SPEM en ratones (Figura 17).



**Figura 17.** Schlafen4 se sobre-expresa en la metaplasia gástrica. Tomado de El-Zaatari *et al.*, 2013<sup>85</sup>.

El ratón transgénico de Schlafen4 presenta una disminución de la densidad de monocitos y macrófagos inflamatorios en la cavidad peritoneal, lo cual es consistente con que Schlafen4 posea un rol regulatorio en el control de la diferenciación de macrófagos. Schlafen4 es inducido por IFN $\beta$  en macrófagos derivados de médula ósea y líneas celulares murinas de fibroblastos NIH3T3<sup>86</sup>.



En otro estudio se comprobó que la expresión murina de Schlafen4 es inducida si la vía hedgehog esta activa y el organismo se encuentra infectado por *H. pylori*. Las células Schlafen4+ migran y adquieren funciones supresoras de células T solamente en el estómago. Las mismas exhiben marcadores de superficie de MDSC que sugieren un cambio metaplásico en la mucosa gástrica (pendiente de aceptación). Las células MDSC pueden desarrollarse bajo condiciones de daño tisular, infección y cáncer; su habilidad para inactivar la respuesta de células T y por ende la respuesta inmune crea un microambiente favorable para la transformación neoplásica <sup>87</sup>.

A pesar de que el homólogo humano de Schlafen4, Schlafen12, no se expresa en mucosa gástrica humana (comunicación personal), otro miembro de esta familia génica, Schlafen5, es inducido por IFN $\alpha$ , citoquina pro-inflamatoria incrementada a consecuencia de la inflamación crónica inducida por *H. pylori* <sup>88</sup>. Schlafen5 es un gen presente solo en humanos y que pertenece al grupo III (masa molecular 100-104 KDa) de la familia de genes Schlafen, los cuales contienen una señal de localización nuclear RKRRR y un dominio C-terminal homólogo al existente en la superfamilia I de ARN helicasas, sugiriendo que posee este tipo de actividad <sup>89</sup>. Schlafen5 es inducido por IFN $\alpha$  mediante el represor traduccional PDCD4 en melanocitos, líneas celulares de melanoma <sup>88</sup>, de fibroblastos (NIH3T3) y fibrosarcoma (L929) <sup>84</sup>. Schlafen 5 disminuye el crecimiento *in vitro* de melanocitos y melanoma humano, por lo que es un inhibidor de la proliferación celular presumiblemente mediante su actividad helicasa <sup>88</sup>.

## **9. Susceptibilidad genética**

Además de los factores de virulencia de *H. pylori* y los componentes ambientales, el largo proceso inflamatorio de respuesta causado por la infección bacteriana es uno de los factores desencadenantes del proceso neoplásico. La susceptibilidad genética del huésped tiene un papel importante en la carcinogénesis gástrica como lo evidencia el hecho de que la heredabilidad del CG se ha calculado en un 28% y que la historia familiar es un factor de riesgo independiente de la infección por *H. pylori* <sup>90</sup>. Aunque el efecto de dicha infección es el principal factor de riesgo de CG, los factores heredados que confieren susceptibilidad genética a la infección y a la severidad de la reacción inflamatoria y otros mecanismos de tumorigénesis son también una parte de este complejo proceso.

El análisis de los polimorfismos genéticos y el estudio de su asociación con enfermedades han sido de gran utilidad para identificar genes asociados a enfermedades complejas como el cáncer. Un polimorfismo genético es una variación en la secuencia de ADN (alelo) presente en más del 1-5% de los individuos de una población. Las enfermedades complejas son aquellas que de manera general están determinadas por la contribución pequeña (*odds ratio* [OR] 1-2) de numerosos genes en combinación con el estilo de vida y la exposición a factores

ambientales. Ejemplos de las mismas son los cánceres no esporádicos, la hipertensión arterial y la diabetes tipo 2<sup>91</sup>.

### **9.1. Proyecto de mapeo de haplotipos humanos HapMap y estudios de asociación genética**

La variación genética más común en el hombre son los polimorfismos de nucleótido simple. Los mismos se distribuyen en bloques de haplotipos, los cuales son regiones en las que existe un elevado desequilibrio de ligamiento (LD, del inglés *linkage disequilibrium*) y poca recombinación genética<sup>92</sup>. El desequilibrio de ligamiento es la asociación no azarosa entre alelos de diferentes loci, debido a su incapacidad de segregación independiente por encontrarse muy próximos en el mismo cromosoma, lo que impide su transferencia a la progenie de manera aleatoria<sup>92</sup>. El análisis de la estructura haplotípica permite la asociación de la variabilidad genética con enfermedades<sup>92</sup>.

Las anteriores conclusiones proporcionaron la base para la construcción del mapa de haplotipos del genoma humano denominado HapMap ([www.hapmap.org/](http://www.hapmap.org/))<sup>92</sup>. El mismo ha permitido reducir el número de SNPs a genotipar manteniendo la potencia para detectar variantes de riesgo. Esta reducción ha sido posible mediante la selección de tagSNPs, los cuales son representativos de la variabilidad genética de una región ya que se encuentran en elevado desequilibrio de ligamiento con otros SNPs. En la actualidad existen programas y algoritmos que analizan el grado de desequilibrio de ligamiento entre SNPs de una determinada región genómica, determinan los bloques de haplotipos que se forman y seleccionan los tagSNPs que los definen<sup>93</sup>.

El proyecto HapMap ha facilitado la realización de una inmensa cantidad de estudios genéticos de asociación, que han permitido identificar nuevos genes asociados a enfermedades complejas. La realización de dichos estudios ha sido motivada por la búsqueda de biomarcadores estables que sean predictores de la predisposición al desarrollo de una enfermedad. A la vez ha sido facilitada por la relativa simplicidad del diseño de estudio caso-control retrospectivo o anidado en una cohorte, la colecta de sangre para extracción de ADN genómico así como el desarrollo acelerado de las técnicas de genotipado automatizado.

En general, los estudios de asociación, y en especial los de genes candidatos, presentan la limitante de la falta de replicabilidad de los resultados al intentar ser replicados en otras poblaciones. Las causas atribuibles son: diferencias entre las muestras en cuanto a su origen étnico y consecuente estructura poblacional, mezcla genética con otras poblaciones o aislamiento genético de las mismas, diferentes patrones de desequilibrio de ligamiento de los genes estudiados, pequeño tamaño muestral, diferencias en el diagnóstico de la enfermedad y en los efectos

ambientales a que están expuestas dichas poblaciones<sup>94</sup>. Con el objetivo de ganar potencia estadística para generar resultados más confiables se han realizado meta-análisis los cuales tienen en cuenta varios estudios incrementando el tamaño muestral y por ende la potencia estadística para detectar una posible asociación<sup>95</sup>.

## 9.2. Estudios de asociación de genes candidatos en cáncer gástrico

Los factores de susceptibilidad genética al CG incluyen genes candidatos involucrados en la respuesta inflamatoria ante la infección, en el reconocimiento de *H. pylori* en el epitelio gástrico, la protección de la mucosa gástrica, el metabolismo y detoxificación de compuestos carcinógenos, y la reparación del ADN, entre otros. Algunas generalidades derivadas del análisis de estos estudios son relevantes a destacar. Una es que se observa un efecto diferencial de asociación en función del origen geográfico de las poblaciones debido a diferentes historias evolutivas de las mismas, lo que condiciona diferencias en las frecuencias alélicas y genotípicas y, por tanto, en el estatus de asociación genética. Como ejemplo de lo anterior, un meta-análisis analizó 225 polimorfismos genéticos de diferentes vías funcionales asociados a CG procedentes de 203 estudios, estando 37 de ellos solamente asociados en asiáticos mientras que otros 12 se asociaron en europeos<sup>96</sup>.

Otra generalidad a destacar es la asociación de determinadas variantes genéticas o el mayor riesgo de las mismas en un subgrupo expuesto a un efecto ambiental (por ejemplo infectados por *H. pylori*, *cagA*+, hábito de fumar, etc.) y no en la población general de estudio o en aquella no expuesta a dicho efecto, en la cual incluso un alelo de riesgo no tendría efecto<sup>97</sup>. Ello es indicativo de una asociación diferencial en función de la exposición al efecto ambiental. Por ejemplo, el genotipo nulo de *GSTT1* se asoció al CG en fumadores mientras que no se asoció en no fumadores<sup>98</sup>. La asociación de *IL1β* con el riesgo de CG es mayor en pacientes infectados con los factores de virulencia *cagA*, *vacA* s1 y m1 que en la población general de dicho estudio<sup>99</sup>.

Una limitante a destacar es que la mayoría de los estudios no analizan el CG teniendo en cuenta sus subtipos histológicos (intestinal/difuso) o localización anatómica (cardias/no cardias). Como se ha explicado previamente estos grupos poseen diferencias en sus etiologías, factores de riesgo, descriptores epidemiológicos como incidencia, prevalencia, sobrevivencia, distribución por sexos; así como, en los procesos moleculares causales. Respecto a los tipos de estudios epidemiológicos usados para los estudios de asociación predominan los caso-control de base hospitalaria, mientras que los prospectivos son los menos frecuentes, y estos son precisamente los que poseen el mayor poder estadístico para detectar asociaciones<sup>97</sup>.

### 9.2.1. Citoquinas pro-inflamatorias

Debido al hecho de que el principal factor de riesgo de CG es la inflamación causada por *H. pylori*, los genes que influyen dicha respuesta son candidatos de interés. Uno de los primeros y más influyentes trabajos realizado al respecto es relativo a polimorfismos en las posiciones -511T (rs16944), -31C (rs1143627) del gen *IL1B* y el antagonista de su receptor (*IL1RN*) como factores de riesgo de CG. Estos polimorfismos se encuentran en las regiones promotoras por lo que deben modular la expresión génica. La *IL1B* es una citoquina pro-inflamatoria sobre-expresada debido a la infección por *H. pylori* y un potente inhibidor de la secreción ácida gástrica. *IL1RN* presenta una repetición en tandem de número variable (VNTR, del inglés *variable number tandem repeat*) de repeticiones de 86 pb (rs2234663)<sup>100</sup>.

Un meta-análisis encontró que el polimorfismo *IL1B* -511\*T e *IL1RN* \*alelo 2 están asociados con el CG no cardias e intestinal en europeos pero no en asiáticos<sup>100 101</sup>. Otro meta-análisis que analiza polimorfismos en los genes *IL1B*, *IL1RN*, *IL8*, *IL10* y *TNFα* estimó un riesgo incrementado de CG intestinal y difuso en los portadores del *IL1RN* \*alelo 2 en poblaciones europeas mientras que en asiáticos existe un riesgo decrementado de CG en los portadores de *IL1B* 31\*C<sup>102</sup>. Las variantes *IL1B* -511\*T (rs16944) y *IL1RN* \*alelo 2 también han sido asociadas con la hipoclorhidria gástrica, la cual permite la diseminación de *H. pylori* al cuerpo, favoreciendo el desarrollo de la atrofia gástrica<sup>103</sup>. Ello indica que la intensidad y el mantenimiento de la respuesta inflamatoria ante la infección son relevantes en los primeros estadios de la carcinogénesis gástrica.

### 9.2.2. Reconocimiento de *Helicobacter pylori* y respuesta inmune innata

Otros genes candidatos son los de reconocimiento del *H. pylori* y mediadores de la respuesta inmune innata, *CD14*, *TLR4*, *NOD2* y *NFKB1*, los cuales activan las vías de señalización que inducen la expresión de citoquinas pro-inflamatorias. Estos polimorfismos pueden modificar la capacidad de reconocimiento de *H. pylori* o sus derivados e incrementar la señalización celular inducida por la infección y por ende los efectos inflamatorios.

*CD14* y *TLR4* reconocen el LPS de bacterias gram-negativas como *H. pylori*, mientras que *NOD2* detecta el peptidoglicano. Estos genes han sido asociados con CG en diferentes poblaciones del mundo. En este sentido, el SNP rs2569190 (-260C/T) de *CD14* ha sido asociado con CG en poblaciones asiáticas; sin embargo, no se asoció en poblaciones europeas<sup>104 105</sup>. Un meta-análisis que analiza las variantes de *CD14* -651C/T y -260C/T sólo encuentra asociación para el *CD14* -260C/T en pacientes infectados por *H. pylori*<sup>106</sup> y no confirma el efecto diferencial por grupo étnico previamente identificado. Mientras que un meta-análisis de la variante -159 C/T no encuentra asociación con CG<sup>107</sup>. Otro meta-análisis encontró asociación entre el

genotipo (insC/insC+insC/-) de *NOD2* rs2066847 (3020insC) con un incremento de riesgo de CG después de analizar la evidencia existente para las variantes rs2066844, rs2066845 y rs2066847 <sup>108</sup>.

Los polimorfismos no sinónimos Asp299Gly (+896 A>G, rs4986790) y Thr399Ile (rs4986791) en el gen *TLR4* se han asociado a CG únicamente en población europea, ya que no están presentes en asiáticos <sup>109</sup>. Los mismos alteran la estructura de la región extracelular de *TLR4* y, por tanto, se ha propuesto que pueden disminuir la capacidad de unión y respuesta al lipopolisacárido bacteriano <sup>110</sup>. Un meta-análisis evaluó el efecto de las variantes de *TLR4* (rs1927914, rs4986790, rs4986791, rs11536889, rs1927911 and rs2149356), concluyendo que rs4986790 y rs4986791 incrementan el riesgo, mientras que rs1927911 lo disminuye <sup>111</sup>. Por último, una variante de inserción/delección (294 ins/del ATTG) de *NFKB1* se asoció con CG en una cohorte china <sup>112</sup>. Otros genes que se han asociado positivamente al CG en meta-análisis en asiáticos y no en europeos son portadores del alelo -160A de *CDH1* <sup>113</sup>, 72Pro (*TP53*) <sup>114</sup>, delección en *GSTT1* <sup>115</sup>, *ERCC2* Lys751Gln y Asp312Asn <sup>116</sup>.

### 9.3. Estudios de asociación de genoma completo en cáncer gástrico

Los estudios de asociación de genoma completo (GWAS, del inglés *genome wide association study*) representan un avance sobre los estudios de genes candidatos debido a que el genoma es inspeccionado en su totalidad mediante el genotipado de cientos de miles de tagSNPs usando chips de ADN y poseen tamaños de muestra en el orden de miles de casos y de controles que permiten tener un poder estadístico adecuado para detectar el pequeño efecto de los SNPs en la causalidad de las enfermedades complejas <sup>117</sup>. En poblaciones asiáticas se han realizado diversos estudios de asociación con marcadores de todo el genoma asociando nuevos genes con el CG cardias y no cardias. En la tabla 3 se indican los SNPs asociados junto con otras características de los estudios.

El primer estudio GWAS fue realizado en población japonesa y reporta una asociación entre dos SNPs (rs2976392) y (rs2294008, Met1Thr) del gen *PSCA* y el CG tipo difuso, la cual ha sido replicada en poblaciones asiáticas y europeas <sup>118</sup>. El SNP rs2294008 es una variante funcional debido a que la sustitución del alelo C>T reduce la actividad transcripcional de este gen. *PSCA* se encuentra sobre-expresado en cánceres de próstata, vejiga y riñón donde actúa promoviendo la progresión tumoral, mientras que está sub-expresado en la MI y el CG difuso, actuando como un gen supresor de tumores. *PSCA* pudiera ejercer un rol en la regulación de la proliferación de la mucosa gástrica y la transducción de señales debido a su localización en la membrana unida a glicosilfosfatidilinositol <sup>118</sup>.

En otro trabajo los SNPs rs2274223 (His>Arg) y rs3765524 de *PLCE1*, y la variante funcional rs4072037 de *MUC1* fueron asociados con CG cardias <sup>119</sup>. El gen *PLCE1*

codifica para una fosfolipasa que hidroliza el fosfatidilinositol bifosfato en los segundos mensajeros diacilglicerol e inositol trifosfato que regulan el crecimiento, la diferenciación y la expresión génica; es un oncogén en cáncer de piel e intestino<sup>120</sup>. El tercer GWAS realizado halló asociación del CG del cardias con la variante rs2274223 de *PLCE1*; así como, con el SNP rs13042395 del transportador de riboflavina expresado en intestino *SLC52A3*<sup>120</sup>.

El último GWAS publicado encontró significación estadística de SNPs intrónicos localizados en región de *PRKAA1-PTGER4* y *ZBTB20* con el CG no cardias<sup>121</sup>. *PRKAA1* codifica para la proteína quinasa activada por AMP 5', un sensor del estado energético celular que garantiza la sobrevivencia en periodos de estrés energético y regulador de la polaridad celular. Por su parte, *PTGER4* participa en la regulación del transporte epitelial intestinal, la expresión de *COX2*, e inhibe el crecimiento de líneas celulares de CG. *ZBTB20* es un factor transcripcional implicado en la hematopoiesis, la oncogénesis y la respuesta inmune.

Tabla 3. Resumen de estudios de asociación de genoma completo en cáncer gástrico.

Autor, año, referencia	Tamaño muestral	Población	Tipo de CG	SNP	Gen	OR (IC 95%)	p-valor	Replicación
Sakamoto <i>et al.</i> , 2008 <sup>118</sup>	926 CG 1397 Controles	Japón	CG difuso	rs2976392	PSCA	1.62 (1.38-1.89)	1.11 x 10 <sup>-9</sup>	1.90 (1.56-2.33)
				rs2294008		1.67 (1.47-1.90)	2.2 x 10 <sup>-15</sup>	1.91 (1.57-2.33)
				rs2070803	MUC1	1.63 (1.33-1.98)	2.2 x 10 <sup>-6</sup>	3.93 x 10 <sup>-5</sup>
Abnet <i>et al.</i> , 2010 <sup>119</sup>	2240 CG 3302 Controles	China	Cardias	rs2274223	PLCE1	1.57 (1.40-1.76)	4.19 x 10 <sup>-15</sup>	1.42 (1.27-1.58)
				rs3765524	MUC1	1.59 (1.40-1.79)	9.94 x 10 <sup>-14</sup>	---
				rs4072037		0.75 (0.65-0.87)	9.45 x 10 <sup>-5</sup>	0.72 (0.63-0.81)
Wang <i>et al.</i> , 2010 <sup>120</sup>	2766 CG 11013 Controles	China Han	Cardias	rs2274223	PLCE1	1.55 (1.45-1.66)	1.74 x 10 <sup>-39</sup>	1.74 x 10 <sup>-39</sup>
				rs13042395	SLC52A3	0.91 (0.85-0.97)	3.02 x 10 <sup>-3</sup>	1.04 (0.94-1.15)
Shi <i>et al.</i> , 2011 <sup>121</sup>	1006 CG 2273 controles	China Han	No Cardias	rs13361707	PRKAA1	1.41 (1.32-1.49)	7.6 x 10 <sup>-29</sup>	1.04 (0.94-1.15)
				rs9841504	ZBTB20	0.76 (0.69-0.83)	1.7 x 10 <sup>-9</sup>	1.29 (1.11-1.51)
				rs2294008	PSCA	---	2.1 x 10 <sup>-7</sup>	---
				rs2976392	PSCA	---	3.7 x 10 <sup>-7</sup>	---
				rs4072037	MUC1	---	1.0 x 10 <sup>-4</sup>	---
				rs13042395	SLC52A3	---	6.9 x 10 <sup>-4</sup>	---

## 9.4 Estudios de asociación en lesiones precursoras gástricas

En la tabla 4 se exponen algunos estudios representativos de la asociación genética con lesiones precursoras gástricas, con el objetivo de ilustrar los resultados de asociación y otras informaciones relevantes de los mismos. Una serie de conclusiones se derivan del análisis de los mismos. Las LPGs más analizadas son la gastritis crónica atrófica y la metaplasia intestinal, mientras que la displasia ha sido analizada en pocos estudios. La mayoría utilizan como controles pacientes afectados de lesiones precursoras como gastritis atrófica para evaluar la asociación con la MI o la displasia. Sin embargo, el análisis de la evolución (progresión y regresión) de las lesiones durante el seguimiento clínico de las mismas es poco frecuente. En diversos estudios el diagnóstico de la GCA no es histológico sino basado en la razón de concentraciones de pepsinógenos I/II séricos (PGI/PGII), la cual es una prueba con baja sensibilidad (50%) y especificidad (80%)<sup>122</sup>. Además, no clasifican a la MI en sus subtipos histológicos completa e incompleta, entidades que muestran un riesgo diferencial de progresar a CG. Adicionalmente, usan diferentes sistemas de clasificación histológica como Sydney y actualizado de Sydney, los cuales utilizan diferentes números de biopsias para el diagnóstico histológico. En algunos estudios un solo patólogo es responsable del diagnóstico histológico mientras que en otros esta tarea es realizada por varios, siendo más confiable el diagnóstico.

### 9.4.1. Citoquinas pro-inflamatorias

Un meta-análisis que evaluó la asociación de polimorfismos *IL1B* -511 T, *IL1RN* VNTR y *TNFA* -308 (rs1800629) halló que el genotipo *IL1RN* LL y portadores del *IL1RN* \*alelo 2 estaban asociados con la MI, siendo los resultados más significativos en europeos e infectados por *H. pylori*<sup>123</sup>. Un estudio de seguimiento desde lesiones precursoras encontró un riesgo incrementado de progresión a CG en los pacientes con GCA o MI portadores del alelo -765C de la ciclo-oxigenasa 2 (*COX2*)<sup>124</sup>. Otros estudios encuentran asociación con la GCA, MI o displasia y los polimorfismos de *IL1B* -511C>T, -31C<sup>125</sup>, +3954C>T (rs143634)<sup>126</sup>, *IL1RN*\*alelo 2, *TNFA* -308 (rs1800629)<sup>123</sup>, *IL2* -330 T>G<sup>127</sup>, *IL8* -251 T>A (rs4073)<sup>128</sup> e *IFNGR1* C-56T<sup>129</sup>.

### 9.4.2. Reconocimiento de *Helicobacter pylori* y respuesta inmune innata

Variantes asociadas con las LPGs en genes de esta vía incluyen a *CD14* 260C>T<sup>130</sup>, *TLR4* +896A/G<sup>131</sup>, *PTPN11* (rs2301756, rs12229892)<sup>132</sup>, entre otros. El mecanismo funcional por el que los SNPs de esta vía incrementan el riesgo de carcinogénesis gástrica está relacionado con el proceso inflamatorio. De este modo *TLR4* +896A/G fue asociado con la hipoclorhidria, un fenotipo que es consecuencia de la GCA<sup>131</sup>.



Un meta-análisis en este mecanismo funcional es el de la variante intrónica de *PTPN11* (rs2301756), que detectó un efecto protector sobre el riesgo de gastritis atrófica en pacientes infectados con *H. pylori*<sup>132</sup>.

Un SNP de cambio de aminoácido, *NOD1* Glu266Lys, se ha asociado con las consecuencias clínicas de la infección por *H. pylori*, como son la úlcera péptica, GCA y MI. Cuatro polimorfismos en el gen *NOD2* también se han asociado con las lesiones precursoras, 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 *NFKB1* en respuesta a *H. pylori*<sup>133</sup>.

#### **9.4.3. Daño y reparación del ADN**

Un estudio de seguimiento de lesiones precursoras al desarrollo de CG en un área de alto riesgo de China halló que los portadores de genotipos combinados de *XRCC1*-194Arg/Trp + Trp/Trp tienen un elevado riesgo de regresión de lesiones (OR 1.44 con intervalo de confianza del 95% [IC95%] 1.06-1.96) mientras que los sujetos que poseían las combinaciones *XRCC1*-399Arg/Gln + Gln/Gln (OR 1.60, IC95% 1.09-2.36) o *OGG1*-326Ser/Cys + Cys/Cys (OR 1.95, IC95% 1.03-3.71) tienen un mayor riesgo de progresión desde la MI o displasia. Así mismo, los portadores de 1 o 2 genotipos de riesgo de *XRCC1*-399 o *OGG1*-326 también poseen un riesgo significativo de progresión<sup>134</sup>. Las mutaciones en el epitelio gástrico inducidas por la inflamación generada por *H. pylori* son reparadas por un mecanismo de escisión de bases entre otros, al que pertenecen X-ray repair cross-complementing group 1 (*XRCC1*) y 8-oxoguanine DNA glycosylase (*OGG1*).

#### **9.4.5. Metabolismo de xenobióticos**

Variantes génicas de delección presentes en glutathione S transferasa *GSTP1*<sup>135</sup>, *GSTM1*<sup>136</sup>, *GSTT1*<sup>136</sup>, y citocromos P450 como *CYP2E1*<sup>135</sup> fueron asociadas con la metaplasia intestinal y la displasia.

#### **9.4.6. Funcionalidad y protección de la mucosa gástrica**

Los alelos S de VNTR de *MUC1* se asociaron con un mayor riesgo de gastritis atrófica y metaplasia intestinal en población portuguesa, con una de las mayores incidencias de CG en Europa<sup>137</sup>. En la población colombiana los genotipos SS y LS de un polimorfismo VNTR de *MUC1* se asocian con la MI<sup>138</sup> mientras que en portugueses los genotipos LL y SS son asociados con la metaplasia intestinal completa e incompleta, respectivamente<sup>139</sup>. Concluyendo que se observa una tendencia a la asociación entre el alelo S de *MUC1* con la metaplasia intestinal.

En un estudio de seguimiento (media de 12.8 años) de lesiones precursoras gástricas en la población española fue evaluada la asociación entre la variabilidad

genética de tagSNPs en *MUC1*, *MUC2* y *MUC6* con la evolución (progresión y/o regresión) de las LPGs, encontrándose asociación con *MUC2*. Los análisis haplotípicos en el extremo 3' de este gen demostraron asociación con un menor riesgo de progresión de lesiones. Mientras que los haplotipos en el extremo 5' fueron asociados con una mayor probabilidad de regresión, indicando un rol de este gen en evitar la progresión de las LPGs a estadios de mayor severidad. Los análisis estratificados evidenciaron que la significación estadística se mantenía solo en pacientes infectados por *H. pylori* y en aquellos que no consumen anti-inflamatorios no esteroideos. Este estudio evidencia que la variabilidad en el gen de la mucina 2 se asocia con la evolución de las lesiones precursoras gástricas, especialmente en infectados por *H. pylori*<sup>140</sup>.

Tabla 4. Estudios de asociación genética en lesiones precursoras de cáncer gástrico.

Referencia, población, año	Gen (Polimorfismo)	Diseño de estudio	Tamaño muestral	OR (IC 95%)	Función del polimorfismo	Observaciones
<b>Citoquinas pro-inflamatorias</b>						
<sup>123</sup> Portugal, Mozambique, 2010	<i>IL1B</i> -511 <i>IL1RN</i> VNTR <i>TNF</i> -308	Asociación con lesiones precursoras: -Casos: MI -Controles: Histológicamente normales y GNA -Sistema actualizado de Sydney  Meta-análisis de previos estudios que evalúan asociación de <i>IL1B</i> , <i>IL1RN</i> , y <i>TNFA</i> con GCA, MI y DISP	Portugueses -Casos, 76 -Controles, 139  Mozambicanos -Casos, 16 -Controles, 88	Asociación con MI en portugueses: - <i>IL1RN</i> VNTR *22 2.83 (1.15-6.96) - <i>IL1B</i> -511 T portadores 1.86 (1.03-3.36)  Meta-análisis, asociación con lesiones precursoras con: - <i>IL1RN</i> VNTR 22 vs LL, 2.27 (1.4-3.7) -Portadores de <i>IL1RN</i> alelo 2, 1.35 (1.12-1.63)	----	Controles afectados por GNA/GCA  Meta-análisis Resultados más significativos en europeos y <i>H. pylori</i> +
<sup>128</sup> China, 2010	<i>IL8</i> -251 <i>MIF</i> -173	Asociación con la evolución de las lesiones:  -Grupo de referencia: No progresión desde Normales o superficial gastritis  -Grupo I: Progresión desde grupo de referencia a GCA  -Grupo II: Progresión desde grupo de referencia a MI  -Sistema actualizado de Sydney	Controles N=137  Casos: N=134  N=101	Asociación con progresión a GCA: - <i>IL8</i> -251 AA vs TT, OR 2.62 (IC 95% 1.23-5.72) -Portadores de alelo A de <i>IL8</i> -251, OR 1.81 (IC 95% 1.06-3.09) - <i>MIF</i> -173 CC vs GG, OR 2.36 (IC 95% 1.38-4.02) -Portadores de alelo C de <i>MIF</i> -173 OR 2.07 (IC 95% 1.21-3.55)  Asociación con progresión a MI: - <i>IL8</i> -251 TA vs TT, 2.27 (1.25-4.14) -Portadores de alelo A de <i>IL8</i> -251, OR 2.07 (IC 95% 1.16-3.69) - <i>MIF</i> -173 CC vs GG, 2.27 (1.16-4.46) -Portadores de alelo C de <i>MIF</i> -173, OR 3.84 (IC 95% 1.58-9.34)	- <i>IL8</i> -251 A produce potencial sitio de unión para C/EBP  - <i>MIF</i> -173 C incrementa la expresión de <i>MIF</i>	Diagnóstico por 3 patólogos en 7 biopsias
<sup>124</sup> Portugal, 2006	<i>COX2</i> -765 G>C	Asociación con la evolución de las lesiones:  -Casos: GCA/ MI -Controles: Sanos	N=37 N=210	Progresión a CG desde GCA/ MI en pacientes portadores del alelo -765C, OR 2.6 (IC 95% 1.03-6.93)	-765 G>C posee baja actividad transcripcional debido a disrupción de TFBS para SP1	Asociación obtenida en grupo compuesto de GCA y MI

Referencia, población, año	Gen (Polimorfismo)	Diseño de estudio	Tamaño muestral	OR (IC 95%)	Función del polimorfismo	Observaciones
<b>Reconocimiento de <i>H. pylori</i> y respuesta inmune innata</b>						
<sup>141</sup> Japón, 2009	<i>PTPN11</i> G/A rs2301756	Asociación con lesiones precursoras: -Casos: GC -Controles: HP+, GCA y no GCA -GCA diagnóstico por PGs	N=583 Controles: -GCA ligera, 327 -GCA severa, 168 -Sin GCA, 442	G/A+A/A vs GG, con GCA severa, 0.62 (0.42-0.90)	Influencia la formación de variantes de empalme alternativo	Análisis de asociación restringido a controles Diagnóstico de GCA basado en la PG I/PGII
<sup>142</sup> Japón, 2009	<i>TLR4</i> +3725 G/C (rs11536889)	Asociación con lesiones precursoras: Casos: CG Controles: GCA y GNA	GCA, 502 GNA, 689	Asociación en <i>H. pylori</i> + : G/C con GCA severa, 1.43 (1.04-2.06)  GC+CC, 1.43 (1.01-2.04)	----	Asociación de GCA restringido a controles Diagnóstico de GCA basado en la PG I/PGII
<sup>130</sup> Venezuela, 2007	<i>CD14</i> C260T (rs2569190)  <i>TLR4</i> Asp299Gly (rs4986790)  <i>NOD2</i> del3020ensC, Gly908Arg (rs2066847)  <i>NOD2</i> Gly908Arg (rs2066845)	Asociación con lesiones precursoras: Casos: GCA, MI y DISP  Controles: -Normales, gastritis superficial y gastritis crónica	-GCA, 289 -MI, 543 -DISP, 118  -Normales y gastritis superficial, 87 -Gastritis crónica, 996	-CD14 con MI, 1.45 (1.11-1.89)	-CD14 260T asociado con expresión incrementada de CD14  -Asp299Gly asociado con respuesta reducida al LPS -Gly908Arg es un SNP no sinónimo	Controles afectados por GCA

Referencia, población, año	Gen (Polimorfismo)	Diseño de estudio	Tamaño muestral	OR (IC 95%)	Función del polimorfismo	Observaciones
<b>Daño y reparación del ADN</b>						
<sup>134</sup> China, 2009	XRCC1 - Arg194Trp XRCC1 - Arg399Gln ADPRT Val762Ala OGG1 - Ser326Cys APE1 - Asp148Glu	Asociación con la evolución de las lesiones	Gastritis superficial y GCA, 400 GCA severa, 132 MI, 458 DISP, 291 Controles sanos, 305	Progresión a CG: -OGG1-326Ser/Cys + Cys/Cys, 1.95 (1.03-3.71) -XRCC1-399Arg/Gln + Gln/Gln, 1.60 (1.09-2.36) Regresión a normalidad XRCC1 194 Arg/Trp + Trp/Trp, 1.38 (1.02-1.89)	SNPs no sinónimos	-Diagnóstico histológico por 3 patólogos
<b>Metabolismo de xenobióticos</b>						
<sup>135</sup> China, 2005	CYP2E1 (RsaI) CYP2E1 (DraI) GSTP1 GSTM1 GSTT1 ALDH2 ODC	Asociación con lesiones precursoras: -Casos: MI severa DISP -Controles: GCA ligera	N=307 N=299 N=302	-Consumo de sal y GSTP1, 1.46 (1.06-2.03) -Consumo de pancake y CYP2E1 RsaI, 2.43 (1.10-5.38) -CYP2E1 DraI y tabaquismo, 1.78 (1.03-3.08)	-CYP2E1 (1053C>T) aumenta transcripción <i>in vitro</i> -Homocigotos de GSTM1 y GSTT1 carecen de actividad GST-μ y GST-θ	-Diagnóstico histológico basado en 7 biopsias realizadas en 2 gastroscopias -Controles afectados de GCA ligera
<sup>136</sup> Taiwan, 2004	GSTM1 GSTT1 CYP2E1	Asociación con lesiones precursoras: -Casos: MI -Controles: enfermedades gastro-intestinales e historia familiar de cáncer gástrico	N= 84 N= 228	Asociación de MI con: -Consumo de pescado salado y vegetales con genotipo GSTM1 nulo. -Consumo de comida salada (carne, vegetales, pescado) con genotipos no nulos de GSTT1 -Consumo de comida salada (carne, vegetales, pescado) con genotipos CYP2E1 c1/c1 vs c1/c2 o c2/c2		Controles incluyen pacientes afectados de enfermedades gastro-intestinales e historia familiar de cáncer gástrico

Referencia, población, año	Gen (Polimorfismo)	Diseño de estudio	Tamaño muestral	OR (IC 95%)	Función del polimorfismo	Observaciones
<sup>136</sup> Taiwan, 2004	GSTM1 GSTT1 CYP2E1	Asociación con lesiones precursoras: -Casos: MI -Controles: enfermedades gastro-intestinales e historia familiar de cáncer gástrico	N= 84 N= 228	Asociación de MI con: -Consumo de pescado salado y vegetales con genotipo GSTM1 nulo. -Consumo de comida salada (carne, vegetales, pescado) con genotipos no nulos de GSTT1 -Consumo de comida salada (carne, vegetales, pescado) con genotipos CYP2E1 c1/c1 vs c1/c2 o c2/c2		Controles incluyen pacientes afectados de enfermedades gastro-intestinales e historia familiar de cáncer gástrico
<b>Protección y funcionalidad de la mucosa gástrica</b>						
<sup>139</sup> Portugal, 2001	MUC1 (VNTR: LL, SS)	Asociación con lesiones precursoras: -Casos: GCA MI -Controles: Sin GCA, ni MI	N=49 N=48 N=125 N=126	-MUC1 LL con GCA, 2.6 (1.0-6.8) y SS 2.5 (1.1-5.7) -MUC1 LL con MI completa, 3.9 (1.1-13.9) -MUC1 SS con MI incompleta MI, 2.9 (1.2-6.0)	----	----
<sup>140</sup> España, 2010	22 tagSNPs en MUC1, MUC2 y MUC6	Asociación con la evolución de las lesiones: Casos -GNA, GCA, MIC, MII, DISP que progresan a LPGs de mayor severidad. Controles -GNA, GCA, MIC, MII, DISP que regresan a lesiones de menor severidad o se mantienen estables	Casos Progresión: 109 Controles Regresión: 111 Estables: 167	-SNPs en extremo 3' de MUC2, (rs10794293, rs3924453 y rs4077759) asociados con menor riesgo de progresión -SNPs en extremo 5' de MUC2, (rs10902073, rs10794281, rs2071174 and rs7944723) asociados con mayor riesgo de regresión		Ausencia de controles sanos



# **HIPÓTESIS**





1. Polimorfismos en genes que participan en el reconocimiento y activación de la respuesta inmune innata ante la infección por *H. pylori* pudieran asociarse con el cáncer gástrico, así como con sus subtipos histológicos y anatómicos.
2. Polimorfismos en genes implicados en el desarrollo de lesiones precursoras y cáncer gástrico en modelos animales y en los mecanismos de infección por *H. pylori*, pudieran asociarse con la evolución (progresión y/o regresión) de las lesiones precursoras gástricas humanas en el tiempo.
3. Si la asociación de polimorfismos genéticos con la evolución de las lesiones precursoras gástricas es cierta, debería replicarse en poblaciones independientes.
4. Teniendo en cuenta que la lesión precursora gástrica con mayor incidencia y prevalencia, menor variabilidad inter-observador y mayor riesgo de progresión a cáncer gástrico en la población española es la metaplasia intestinal; y que dicha progresión es mayor en la metaplasia intestinal incompleta con respecto a la completa en diversas poblaciones, incluyendo a la española. Es posible que existan diferencias en el perfil transcripcional de la metaplasia intestinal incompleta y de la completa que progresan a cáncer gástrico respecto a las mismas lesiones que no progresan, y que el análisis de estos perfiles permita identificar genes y procesos moleculares causales de la progresión. Así mismo, el estudio del transcriptoma de la metaplasia intestinal que no progresa a cáncer gástrico respecto a la mucosa sana permitiría identificar nuevos procesos moleculares de esta lesión y confirmar otros ya identificados.
5. Schlafen4 se encuentra sobre-expresado en la metaplasia gástrica murina por células supresoras derivadas de mieloides, las cuales ejercen un rol supresor de la inmunidad anti-tumoral. Es posible que otro miembro de esta familia, Schlafen5, cuya expresión es inducida por la citoquina pro-inflamatoria IFN $\alpha$  en melanocitos humanos, también se sobre-exprese en células inflamatorias de la MI humana y su alta expresión pudiera ser un marcador de la progresión desde metaplasia intestinal a cáncer gástrico.



# **OBJETIVOS**



1. Determinar si polimorfismos en genes humanos de reconocimiento y activación de la respuesta inmune innata ante la infección por *H. pylori* (*CD14*, *NOD2*, *TLR4* y *NFKB1*) se asocian con el adenocarcinoma gástrico; así como, a sus subtipos histológicos (intestinal/difuso) o por localización anatómica (cardias/no cardias), mediante un estudio de asociación caso-control anidado en la cohorte europea EPIC-Eurgast.
2. Determinar si polimorfismos en genes implicados en la carcinogénesis gástrica y en el desarrollo de lesiones precursoras se encuentran asociados con la evolución (progresión/regresión) de las lesiones precursoras gástricas, mediante un análisis de asociación con muestras de un estudio multicéntrico de seguimiento de lesiones precursoras de cáncer gástrico en España.
3. Determinar si SNPs previamente asociados a la evolución de las lesiones precursoras de cáncer gástrico en un estudio de seguimiento en la provincia de Soria se replican en el estudio multicéntrico de seguimiento de lesiones precursoras en España.
4. Identificar genes y mecanismos moleculares diferencialmente expresados en la progresión de los tipos histológicos de metaplasia intestinal completa e incompleta al cáncer gástrico, y en la metaplasia intestinal que no progresa a cáncer gástrico, mediante un estudio de microarrays de expresión de ARNm.
5. Determinar si el IFN $\alpha$  induce la expresión génica de Schlafen5 en células inflamatorias de la mucosa gástrica y que tipos celulares expresan la proteína Schlafen5 en la metaplasia intestinal. Así como comparar la expresión inmunohistoquímica de Schlafen5 en las diferentes lesiones precursoras de la cascada de Correa, con especial interés en los subtipos histológicos de metaplasia intestinal que progresan y no progresan a cáncer gástrico.



# **MATERIALES Y MÉTODOS**





## 1. Estudio de asociación de genes de señalización de *Helicobacter pylori* y riesgo de cáncer gástrico

### 1.1. Pacientes y muestras

Este estudio se basa en uno previo tipo caso control anidado en la cohorte de la European Prospective Investigation into Cancer and Nutrition (EPIC), denominado Eurgast. La cohorte EPIC se encuentra constituida por 521 457 sujetos sanos de ambos sexos, y edades entre 35-70 años, reclutados entre los años 1992 y 2000 en 10 países europeos (incluyendo España), y que durante el seguimiento (media >15 años) han desarrollado diversas enfermedades <sup>143</sup>. Los casos identificados de CG (N=365) a la fecha del análisis, son adenocarcinomas definidos por el código C16 de la 10<sup>ma</sup> revisión de la Clasificación Internacional de Enfermedades.

Por cada caso se seleccionaron hasta un máximo de 4 controles vivos y no afectados por cáncer en el momento del diagnóstico de los casos, apareados con los casos por centro de reclutamiento, sexo, edad ( $\pm 2.5$  años) y fecha de colecta de sangre ( $\pm 45$  días). El seguimiento se inició en la fecha de reclutamiento y acabó en la fecha de diagnóstico de CG, fecha de fallecimiento o fecha final de seguimiento (Diciembre de 2006). Un panel de patólogos independientes revisaron los cortes histológicos originales así como los informes de cada centro EPIC, con el fin de validar el diagnóstico y clasificar los casos según su localización anatómica (cardias y no cardias) y tipo histológico (intestinal o difuso) <sup>144</sup>. El tamaño de muestra analizada es de 365 casos y 1284 controles.

De todos los individuos, en el momento del reclutamiento se recogió información sobre la dieta, estilo de vida, nivel educacional y social, actividad física, consumo de tabaco y alcohol, historia reproductiva, consumo hormonal, enfermedades y se colectaron muestras de sangre para la obtención de ADN y de suero para evaluar la infección por *H. pylori* <sup>143</sup>. El estudio fue aprobado por los comités éticos de la International Agency for Research on Cancer (IARC) y de cada uno de los centros de reclutamiento EPIC.

### 1.2. Selección de tagSNPs en genes de la vía de señalización de *Helicobacter pylori*

Previamente nuestra unidad había realizado un genotipado de 1287 SNPs en 249 genes candidatos de carcinogénesis gástrica en las muestras del estudio Eurgast descrito en el apartado 1.1 de esta sección. Como resultado de ese trabajo se encontró una asociación entre el gen de *PSCA* con CG de los tipos intestinal y difuso en población europea <sup>145</sup>. Este estudio se centra en el análisis de los resultados de genotipado de 4 genes (*CD14*, *NOD2*, *TLR4*, *NFKB1*) de la vía de señalización de *Helicobacter pylori* constituida por proteínas humanas que interactúan directamente con esta bacteria. Con el objetivo de optimizar el análisis de la

variabilidad genética de los genes seleccionados utilizando el menor número de SNPs, teniendo en cuenta las relaciones de desequilibrio de ligamiento en el genoma humano, se utilizó la base de datos HapMap (<http://www.hapmap.org>) y el programa Haploview versión 4.0 para identificar tagSNPs de los principales bloques de haplotipos.

La selección de tagSNPs se realizó siguiendo una estrategia general previamente descrita <sup>145</sup>. Haciendo uso del navegador del sitio web de HapMap (<http://www.hapmap.org>) en población de origen europeo (CEU), *release* #23a o 24, basado en la versión 126 de dbSNP y el NCBI *genome build* 36, fueron cargados los datos genotípicos de cada gen seleccionado; así como, un mínimo de 10 y 3 kilobases adicionales de sus regiones 5' y 3'. Posteriormente, dichos datos genotípicos fueron cargados en el programa Haploview v.4.0 (<http://www.broad.mit.edu/mpg/haploview>) y se seleccionaron los SNPs con frecuencia del alelo menor (MAF, del inglés *minor allele frequency*) de 0.05. Para la definición de bloques en desequilibrio de ligamiento (haplotipos) se aplicó el método de los intervalos de confianza de Gabriel *et al.* <sup>92</sup>, utilizando los parámetros por defecto en Haploview v.4.0. Posteriormente, fueron seleccionados tagSNPs mediante la opción *Pairwise Tagger* con un umbral de  $r^2 \geq 0.8$  entre cada par de tagSNP y su conjunto de SNPs *tagged*; así como, haplotipos *tagged* con frecuencia  $\geq 0.05$ .

Con el objetivo de que la variabilidad analizada en cada región fuera la mayor posible 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 seleccionados tuvieron que ser sustituidos por otros de las mismas características (preferiblemente en total desequilibrio de ligamiento con los seleccionados) para poder ser correctamente genotipados. Los tagSNPs seleccionados y genotipados fueron 6 en *CD14*, 5 en *NOD2*, 7 en *TLR4* y 12 en *NFKB1*. Los bloques de haplotipos definidos por estos tagSNPs son 2 en *CD14* y *NOD2* mientras que 1 para *TLR4* y *NFKB1* (Tabla 20 en la sección Resultados).

### 1.3. Extracción de ADN y genotipado

El ADN genómico fue extraído a partir de la capa leucocitaria empleando el kit Puregene DNA Purification System (Gentra Systems, EUA), y genotipado en el Centro Nacional de Genotipado (CEGEN) del Instituto Carlos III usando la plataforma Illumina BeadStation Platform y la tecnología GoldenGate (Illumina Inc., EUA). Un total de un 5% de muestras duplicadas fueron incluidas como controles del error de genotipado <sup>145</sup>.

#### 1.4. Infección por *Helicobacter pylori*

La infección por *H. pylori* fue determinada mediante la detección de anticuerpos anti-cagA en plasma<sup>146</sup> o inmunoblot empleando el HELICOBLOT 2.1 (MP Biomedicals, EUA)<sup>11</sup>.

#### 1.5. Análisis estadísticos y bioinformáticos

Las posibles diferencias de la distribución de variables clínicas entre casos y controles fueron evaluadas mediante la prueba de chi cuadrado. Para probar la asociación estadística entre los SNPs genotipados y el CG en general, así como con sus subtipos anatómicos (cardia/no cardias) e histológicos (intestinal/difuso), se calcularon los OR ajustados (IC95%) mediante regresión logística no condicional ajustando por las variables de emparejamiento edad, sexo y país de reclutamiento. Los OR fueron calculados para los siguientes modelos genéticos: log-aditivo, dominante, recesivo y codominante. Debido a la baja MAF (<0.05) en los SNPs rs4986790 y rs4986791 del gen *TLR4* solamente se consideró el modelo codominante. En todos los análisis el alelo o genotipo más frecuente fue considerado como la categoría de referencia. Se realizó un test de heterogeneidad para estimar el efecto diferencial de las variantes significativas sobre los distintos subtipos histológicos y localizaciones anatómicas de CG usando el modelo log-aditivo. Para corregir el efecto de las comparaciones múltiples se realizó un test de permutaciones (N=1000) a nivel de genes<sup>147</sup>. El p-valor de permutación de cada gen fue entonces corregido por el número de genes analizado (N=4) atendiendo a la corrección de Bonferroni. La significación después de dicha corrección fue establecida en  $p < 0.0125$ .

Teniendo en cuenta que la localización anatómica no cardias es la única asociada con la infección por *H. pylori* analizamos si la presencia de anticuerpos anti-cagA modifica el efecto genotípico. Para ello, en casos con CG no cardias y controles infectados por *H. pylori* calculamos un OR comparando aquellos individuos cagA+ respecto a los cagA-.

La asociación de haplotipos con CG en general, o sus localizaciones anatómicas y subtipos histológicos también se analizó mediante regresión logística usando el haplotipo más frecuente como referencia. Las frecuencias haplotípicas fueron inferidas mediante el algoritmo de expectación-maximización implementado en la librería haplo.stats del paquete estadístico R. El resto de los análisis estadísticos fueron realizados usando la librería SNPassoc implementada en R<sup>148</sup>. Un resultado fue considerado significativo si  $p < 0.05$ .

Se predijeron las posibles consecuencias funcionales de los SNPs debido a su localización en regiones intrónicas, intergénicas, UTR 5' y 3', reguladoras, *upstream*

y *downstream*, sitios de *splicing* alternativo, SNPs sinónimos y no sinónimos empleando el programa bioinformático Variant Effect Predictor <sup>149</sup> ([http://www.ensembl.org/Homo\\_sapiens/Tools/VEP](http://www.ensembl.org/Homo_sapiens/Tools/VEP)). Adicionalmente, se utilizó el programa PupaSuite v.3.1 (<http://pupasuite.bioinfo.cipf.es/>) con el objetivo de predecir consecuencias funcionales de los SNPs localizados en sitios de unión a factores de transcripción (TFBS, del inglés *transcription factor binding site*), potenciadores exónicos del *splicing* alternativo (ESE, del inglés *exonic splicing enhancer*), silenciadores exónicos del *splicing* alternativo (ESS, del inglés *exonic splicing silencer*), regiones conservadas entre humanos y ratón, así como SNPs en miRNAs y los genes regulados por los mismos <sup>150</sup>.

Además, los tagSNPs asociados y no asociados a cáncer gástrico o a la evolución de las lesiones precursoras gástricas fueron introducidos en el buscador eQTL (<http://www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi>) para evaluar cuales eran *expression Quantitative Trait Loci* (eQTL), es decir, reguladores del nivel de expresión del ARNm. Este programa busca si el SNP de interés es un eQTL por correlacionarse significativamente sus genotipos con los niveles de expresión de uno o más ARNm, según análisis realizados en tejido humano linfoblastoide <sup>151</sup>, hígado <sup>152</sup> y cerebro <sup>153</sup>. También, se empleó con el mismo objetivo el buscador eQTL del grupo Gilad/Pritchard (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>) que recopila información de eQTLs en hígado <sup>152</sup>, cerebro <sup>154</sup>, tejido linfoblastoide <sup>155</sup>, monocitos <sup>156</sup> y diversos tejidos <sup>157</sup>. Después de ser identificados los posibles efectos funcionales se realizó la prueba exacta de Fisher, para ver si existían diferencias en la distribución de cada localización o efecto entre los SNPs asociados respecto a los no asociados al CG o a la evolución de las LPGs.

## 2. Estudio de asociación genética en el seguimiento de lesiones precursoras gástricas

### 2.1. Pacientes y muestras

Las muestras utilizadas proceden de un estudio multicéntrico longitudinal de lesiones precursoras de CG desarrollado en 9 hospitales públicos de Madrid, Barcelona, Zaragoza, Valladolid y Guipuzcoa con el siguiente diseño de estudio.

Se realizó una búsqueda de pacientes que entre 1995 y 2004 tenían un diagnóstico histológico de GCA, MI o DISP. Del total de pacientes identificados fueron excluidos 12 073 pacientes por presentar algunos de los siguientes criterios de exclusión:

- Edad fuera del rango de 25-69 años (N=3 348).
- Úlcera péptica (N=583).
- Cáncer gástrico u otro cáncer (N=468).
- Esófago de Barret (N=200).

- Gastrectomía previa (N=303).
- Falta de calidad de la biopsia gástrica (N= 6 349).
- Falta de información (N= 591).
- Pacientes que viven fuera del área de influencia hospitalaria (N=231).

Se seleccionaron 1 538 pacientes para ser incluidos en el estudio que cumplían con los siguientes criterios:

- Edad en el rango 25-69 años.
- Ausencia de úlcera péptica, cáncer gástrico u otro cáncer, esófago de Barret y gastrectomía previa.
- Biopsia gástrica con 3 fragmentos donde al menos 1 pertenece al antro y otro al cuerpo.

Los mismos fueron invitados a ser sometidos a una gastroscopia con toma de muestras gástricas e independientemente de saliva, para la extracción del ADN. Así como, a responder un cuestionario personal acerca de historia de tabaquismo, uso de anti-inflamatorios no esteroideos, diagnóstico y tratamiento de infección por *Helicobacter pylori*, historia familiar de CG, grupo sanguíneo, ovariectomía y uso de hormonas anticonceptivas. Dicha información epidemiológica fue obtenida de las historias clínicas o a partir de familiares en el caso de aquellos pacientes con un punto final predefinido al seguimiento y los que no pudieron ser entrevistados. Todos los participantes firmaron un consentimiento informado aprobado por el comité ético de cada hospital participante. Para el presente estudio de asociación genética el tamaño de muestra fue de 559 pacientes afectados con lesiones precursoras gástricas.

## **2.2. Histopatología**

La caracterización histopatológica de las lesiones al reclutamiento y al final del seguimiento se realizó por tinción hematoxilina-eosina en la mayoría de los casos, aunque algunos hospitales usaron azul de Alcian-ácido peryódico de Shiff pH 2.5. El diagnóstico fue realizado según la clasificación de Sydney (apartado 7 de la sección Introducción) por el mismo patólogo en cada hospital. Se clasificó la muestra según la lesión más severa en la cascada de Correa observada en cualquiera de los fragmentos<sup>47</sup>.

## **2.3. Infección por *Helicobacter pylori***

La infección por *Helicobacter pylori* fue determinada a partir de tinción de Giemsa, el test rápido de la ureasa, o como resultado de la entrevista con el paciente. La presencia de un resultado positivo en alguna de estas fuentes fue considerada como un resultado positivo en general para dicho paciente. Adicionalmente, el Servicio Digestivo de la Corporación Sanitaria Parc Taulí de Barcelona genotipó los

factores de virulencia *cagA* y *vacA* en ADN extraído de las biopsias gástricas al reclutamiento.

#### **2.4. Prueba piloto de colecta, conservación y extracción de ADN**

Debido a la dificultad de varios hospitales participantes en el proyecto a la hora de coleccionar, almacenar y enviar sangre para la posterior extracción de ADN genómico con el fin de ser empleado en el estudio de asociación genética, nos planteamos la posibilidad de evaluar diferentes métodos alternativos de colecta de muestras biológicas para la extracción del ADN, como hisopos bucales (Isohelix, Epicentre) y saliva (Oragene, DNA Genotek). Para ello se realizó una prueba piloto con sujetos normales del departamento de trabajo con el objetivo de probar la calidad del ADN extraído de dichos procedimientos.

Así, fueron colectadas muestras de mucosa bucal con hisopos bucales Dri-Capsules (SGC-50, Isohelix, Inglaterra) (N=20) y BuccalAmp™ DNA Extraction Kit (Epicentre, EUA) (N=20). También, muestras de saliva mediante el kit Oragene DNA (OG-500, DNA Genotek, Canadá) (N=10) siguiendo las recomendaciones de los fabricantes. Todas se incubaron a temperatura ambiente (TA) durante 30 días y, posteriormente, se extrajo el ADN según el protocolo de los fabricantes.

##### A) Protocolo Dri-Capsules (SGC-50, Isohelix, Inglaterra)

1. Añadir 500 µl de tampón de lisis LS y 20 µl de proteinasa K al hisopo bucal, mezclar con vórtex.
2. Incubar a 60°C durante 1 hora.
3. Transferir el hisopo en posición invertida a otro tubo y centrifugar a 16 500 g durante 1'.
4. Añadir 500 µl de tampón CT y mezclar con vórtex.
5. Centrifugar a 16 500 g por 10' a 4°C y eliminar el sobrenadante.
6. Añadir 150 µl de tampón Tris-EDTA (TE) 1X (10 mM Tris, 1 mM EDTA, pH 8) y centrifugar a 16 500 g por 2'.
7. Transferir el sobrenadante a un tubo nuevo.

##### B) Protocolo BuccalAmp™ DNA Extraction Kit (BQ0901SSC, Epicentre, EUA)

1. Añadir 380 µl de tampón SET 1X (Tris 10 mM pH 7.4, SDS 10%, EDTA 50 mM pH 8, ajustar pH 7.8) y 20 µl de proteinasa K 20 mg/ml, mezclar con vórtex.
2. Incubar a 60°C durante 105' con agitación manual a intervalos.
3. Transferir el hisopo en posición invertida a otro tubo y dar un golpe de centrifuga.
4. Añadir 200 µl de fenol y 200 µl de cloroformo, mezclar con vórtex 30''.
5. Centrifugar a 16 500 g por 10' a TA.
6. Transferir la fase acuosa a un tubo nuevo.
7. Añadir 800 µl de isopropanol, 20 µl de NaCl 5M y 1 µl de glicógeno 20 µg/µl.
8. Incubar a -20°C por 2 horas.
9. Centrifugar a 16 500 g por 15' a 4°C y desechar el sobrenadante.

10. Añadir 1 ml de EtOH 70% y mezclar con vórtex.
11. Centrifugar a 16 500 g por 5' a TA.
12. Secar a temperatura ambiente.
13. Añadir 50 µl de tampón TE 1X y disolver.

C) Colecta de saliva (kit Oragene DNA, OG-500, DNA Genotek, Canadá)

1. No comer, beber o fumar durante los 30 minutos anteriores a la colecta.
2. Enjuagar la boca con agua.
3. Esperar 2-3 minutos y coleccionar alrededor de 2 ml de saliva.
4. Mezclar con la solución de preservación de saliva.

D) Protocolo de extracción de ADN de saliva (kit Oragene DNA, OG-500, DNA Genotek; modificado añadiendo paso de fenol: cloroformo: isoamilalcohol<sup>158</sup>)

1. Incubar a 50°C durante toda la noche.
2. Transferir el volumen (V) a un tubo de 15 ml.
3. Añadir 1/25 volúmenes (Vs) de Oragene Purifier y mezclar con vórtex.
4. Incubar en hielo 10'.
5. Centrifugar a 4 000 g por 20' a TA.
6. Transferir el sobrenadante a un tubo limpio.
7. Añadir 1 V de EtOH 100% y 1/20 Vs de NaCl 5M, invertir varias veces.
8. Centrifugar a 4 000 g por 20' a TA.
9. Eliminar el sobrenadante.
10. Lavar con 1 ml de EtOH 70%.
11. Secar el precipitado a TA.
12. Añadir 600 µl de TE 1X, mezclar con vórtex e incubar a 50°C durante 1 hora.
13. Enfriar y añadir 12 µl de RNAsa A e incubar 1 hora a TA.
14. Añadir 1 V de fenol: cloroformo: isoamilalcohol, mezclar con vórtex 30''.
15. Centrifugar a 16 500 g-5' a TA.
16. Transferir fase acuosa a tubo nuevo.
17. Añadir 2 Vs de EtOH 100%, 1/20 Vs de NaCl 5M e invertir.
18. Centrifugar a 16 500 g por 5' a TA.
19. Lavar 2 veces con 1 ml de EtOH 70% y desechar el sobrenadante.
20. Secar a temperatura ambiente.
21. Diluir en 50-100 µl de TE 1X, de acuerdo al tamaño del precipitado.

**2.4.1. Medición de cantidad y calidad del ADN**

Para todos los casos la concentración de ADN fue medida mediante espectrofotometría usando el equipo Nanodrop 1000 (Thermo Scientific, USA). También se determinaron las relaciones de absorbancia 260/280 nm y 260/230 nm para determinar la posible contaminación proteica (260/280<1.8) y con fenol o sales (260/230<1.8), respectivamente.



## 2.5. Diseño de solución estabilizadora de saliva y prueba de conservación

Teniendo en cuenta que los mejores resultados de la anterior prueba piloto fueron obtenidos con el ADN proveniente de saliva y que los kits eran de un coste elevado (13.9 euros/muestra, Abyntek distribuidor en Europa de DNA Genotek) para nuestro presupuesto, se planteó la posibilidad de confeccionar en el laboratorio un similar de la solución estabilizadora de saliva que usan los kits Oragene DNA. Se realizó una prueba piloto con las diferentes soluciones estabilizadoras de saliva propuestas en la aplicación de patente asociada a dichos kits: "Compositions and methods for obtaining nucleic acids from sputum. Appl. No 12/338.848"<sup>159</sup>. Esta prueba fue necesaria debido a que el lenguaje ambiguo de las patentes no permite conocer cuál es la solución elegida por los autores como la de mejores resultados (Tabla 5).

Tabla 5. Soluciones de colecta y estabilizadoras de saliva. Tomado de Birnboim, 2004 <sup>159</sup> .									
Soluciones	Tris base <sup>a</sup>	LiCl <sup>b</sup>	Acetato de Sodio <sup>c</sup>	Urea <sup>d</sup>	SDS <sup>e</sup>	CDTA <sup>f</sup>	Vitamina C	Etanol	pH
1	0.3 M	No	0.67 M	0.67 M	0.6 %	3.3	0.1 M	30%	8
2	0.2 M	No	0.5 M	No	1%	10 mM	0.15 M	30%	9.5
3	0.2 M	No	0.5 M	No	1%	10 mM	No	30%	9.5
4	33 mM	0.67 M	No	0.67 M	0.6 %	3.3	No	30%	8

<sup>a</sup> Trisaminometano (HOCH<sub>2</sub>)<sub>3</sub>CNH<sub>2</sub>, <sup>b</sup> Cloruro de litio, <sup>c</sup> C<sub>2</sub>H<sub>3</sub>NaO<sub>2</sub>, <sup>d</sup> CO(NH<sub>2</sub>)<sub>2</sub>, <sup>e</sup> Dodecilsulfato sódico NaC<sub>12</sub>H<sub>25</sub>SO<sub>4</sub>, <sup>f</sup> 1,2-cyclohexanediaminetetraacetic acid low sodium.

La prueba piloto consistió en mezclar (N=20 individuos) los 2 ml de saliva con 2 ml de cada una de las soluciones estabilizadoras siguiendo el protocolo de colecta para el kit Oragene (OG-500, DNA Genotek, Canadá) e incubándolos a temperatura ambiente durante 60 días, a continuación se extrajo el ADN mediante el protocolo descrito.

### 2.5.1. Protocolo de extracción de ADN a partir de saliva

Modificación del método de extracción de ADN basado en salificación<sup>160</sup> (*salting-out*) y fenol: cloroformo: isoamilalcohol<sup>158</sup>.

1. Mezclar la mezcla de saliva y la solución de preservación e incubar a 50 °C toda la noche en baño térmico.
2. Añadir 1/6 Vs de NaCl ≈6M, mezclar con vórtex 20'' y dejar 10' en hielo.
3. Centrifugar a 1 792 g por 10' a TA.
4. Recuperar el sobrenadante.
5. Añadir 1 V de ETOH 100% helado e invertir hasta la formación del precipitado de ADN, incubar 10' en hielo.
6. Centrifugar a 1 792 g por 20' a 4°C.
7. Desechar el sobrenadante y dejar secar el precipitado a TA durante 10-15'.

8. Añadir 500 µl de TE 1X y disolver el precipitado (incubar a 50°C si es necesario).
9. Pasar a tubo Eppendorf.
10. Añadir 2 µl de RNAsa A 10 mg/ml e incubar 1 hora a TA.
11. Añadir 1 V de fenol: cloroformo: isoamilalcohol y mezclar con vórtex 30''.
12. Centrifugar a 16 500 g por 10' y recoger la fase acuosa en un tubo nuevo.
13. Añadir 150µl de TE 1X a la fase orgánica restante y mezclar con vórtex 30''.
14. Centrifugar a 16 500 g por 10' a TA, tomar el sobrenadante y combinar con el recogido en el paso 12.
15. Añadir 1 V de fenol: cloroformo: isoamilalcohol y mezclar con vórtex 30''.
16. Centrifugar a 16 500 g durante 10' y recoger el sobrenadante.
17. Añadir 2 Vs de etanol 100%. Si no se observa el precipitado de ADN incubar toda la noche a -20°C.
18. Centrifugar a 16 500 g por 10' a TA y desechar el sobrenadante.
19. Lavar con 1 ml de EtOH 70%.
20. Dejar secar a temperatura ambiente.
21. Añadir 50-150 µl de TE 1X y disolver.

### 2.5.2. Prueba piloto de genotipado de ADN de saliva

Con el objetivo de probar la calidad del ADN extraído de la saliva para el genotipado de polimorfismos tipo SNP hicimos una prueba usando sondas TaqMan y el método Endpoint Genotyping con el equipo de reacción en cadena de la polimerasa (PCR, del inglés *polymerase chain reaction*) en tiempo real LightCycler® 480 Real-Time PCR (qRT-PCR) System (Roche, Suiza). Evaluamos los SNPs rs7578034 y rs16944 (*IL1B*), rs895819 (miRNA27a), empleando el siguiente protocolo:

SNP	ADN (ng)	TaqMan® Universal PCR Master Mix (µl)	20X Assay Mix (µl)	H <sub>2</sub> O (µl)
rs7578034 y rs16944	10	2.5	0.25	1.25
rs895819	100	5	0.5	2.5

La TaqMan® Universal PCR Master Mix está compuesta de una mezcla de oligonucleótidos no marcados para PCR y sonda TaqMan® MGB (marcadas FAM™ y VIC®). El programa de amplificación fue desnaturalización inicial de 10'a 95°C y 40 ciclos de pasos 15'' a 92°C y 1' a 60°C.

### 2.5.3. Colecta de saliva en hospitales

La solución escogida después del estudio comparativo y buen genotipado (apartados 2.1 y 2.3 de la sección Resultados) fue la #4 (Tabla 5). La misma se preparó en mayor cantidad, se esterilizó con un filtro de 0.2 µm (Millipore, Alemania) y se prepararon alícuotas de 2 ml en tubos Eppendorf que fueron enviadas a los hospitales participantes con las instrucciones precisas para la recogida de las muestras de saliva. Las mismas fueron conservadas a 4°C antes y después de la colecta de saliva. Los hospitales enviaron a nuestro departamento en

el ICO tandas de entre 5 y 20 muestras de saliva de sus pacientes conservadas a 4°C para la extracción del ADN.

## **2.6. Extracción de ADN de tejido gástrico parafinado**

Se extrajo ADN del tejido gástrico de las biopsias del reclutamiento con el objetivo de genotipar *cagA* y *vacA* de acuerdo al siguiente protocolo, basado en el método del fenol: cloroformo: isoamilalcohol<sup>158</sup>.

1. Cortar bloques de parafina de 5 µm de espesor atendiendo a la cantidad de tejido (Suficiente 10-15 cortes, Regular 20-25, Poco 30-35). Colocar los cortes en tubos Eppendorf de 1.5 ml resistentes al fenol y limpiar el micrótopo entre cortes con Histoclear (HS-202, National Diagnostics, EUA). Cortar controles negativos de tejido no gástrico de ratón (pulmón) al inicio, medio y final de cada tanda de corte.
2. Añadir 1 ml de xileno, mezclar con vórtex 10'', 5' a 50°C, mezclar con vórtex 10''.
3. Centrifugar a 16 500 g por 3' y desechar el sobrenadante.
4. Repetir los pasos 2-3.
5. Añadir 1 ml de EtOH 100% y mezclar con vórtex 10''.
6. Centrifugar a 16 500 g por 5' y desechar el sobrenadante.
7. Repetir los pasos 5-6.
8. Secar los tubos abiertos en campana por 20-30'.
9. Añadir 350 µl de tampón de lisis de parafinas (100 mM NaCl, 10 mM EDTA pH 8, 10 mM Tris, 0.5% Sarkosyl, pH 7) y 15-50 µl proteinasa K 20 mg/ml y mezclar.
10. Incubar a 56°C en baño térmico entre 3 y 8 horas hasta la lisis completa.
11. Incubar a 80°C durante 15'.
12. Enfriar la muestra a TA, añadir 2 µl de RNasa A 100 mg/ml e incubar 1 hora a TA.
13. Añadir 1 V de fenol: cloroformo: isoamilalcohol y mezclar con vórtex 30''.
14. Centrifugar a 16 500 g por 5'. Recoger la fase acuosa.
15. Añadir 150 µl de TE 1X a la fase orgánica y mezclar con vórtex 30''.
16. Centrifugar a 16 500 g por 5' min. Recoger la fase acuosa y mezclar con la similar del paso 14.
17. Añadir 2 Vs de EtOH 100%, 1/25 Vs de NaCl 5M y 1/100 Vs de MgCl<sub>2</sub> 1M e invertir.
18. Incubar toda la noche a -20°C.
19. Centrifugar a 16 900 g a 4°C durante 20'. Desechar el sobrenadante.
20. Lavar con 1 ml de EtOH 70% frío y mezclar con vórtex 30''.
21. Centrifugar a 16 900 g a 4°C por -5'. Desechar el sobrenadante.
22. Secar a temperatura ambiente.
23. Adicionar 15-50 µl de TE 1X, de acuerdo al pellet, y disolver.

## 2.7. Extracción de ADN de sangre

De algunos pacientes, en lugar de saliva se colectaron muestras de sangre total, en tubos con EDTA y se conservaron a  $-80^{\circ}\text{C}$ . El protocolo de purificación de ADN usado fue el método de salificación (*salting-out*)<sup>160</sup>:

1. Descongelar, mezclar la sangre, pasar 3 ml a un tubo de 15 ml y adicionar 9 ml de tampón de lisis de eritrocitos (TLE) 1X (0.155M  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , 0.1mM EDTA, pH 7.4).
2. Incubar 10' a TA y mezclar por inversión. Centrifugar a 4 000 rpm por 5' a  $15^{\circ}\text{C}$ . Eliminar el sobrenadante.
3. Repetir 2 veces los pasos 1-2.
4. Añadir 1 310  $\mu\text{l}$  de tampón SE (75mM NaCl, 25mM EDTA, pH 8.0), 150  $\mu\text{l}$  de SDS al 10% y 40  $\mu\text{l}$  de proteinasa K.
5. Incubar en baño térmico a  $60^{\circ}\text{C}$  hasta la lisis completa.
6. Adicionar 1.5 ml de tampón SE, 750  $\mu\text{l}$  NaCl 6M y 3.75 ml de cloroformo.
7. Mezclar con vórtex 20'' y durante 30' mezclar por inversión.
8. Centrifugar a 2 000 rpm por 10' a TA.
9. Transferir la fase superior a un tubo de 15 ml y añadir 1V de isopropanol.
10. Centrifugar a 4 000 rpm por 10' a TA y desechar el sobrenadante.
11. Añadir 1 ml de EtOH 70% frío y mezclar con vórtex.
12. Centrifugar a 4000 rpm por 5' a TA y eliminar el sobrenadante.
13. Secar a temperatura ambiente.
14. Adicionar 100-200  $\mu\text{l}$  de TE 1X y disolver.

## 2.8. Selección de genes candidatos

A partir de búsquedas bibliográficas intensivas realizadas en PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) usando como filtro de selección *Title/Abstract* y combinaciones de términos de búsqueda (*gastric cancer/adenocarcinoma, atrophy/atrophic gastritis, intestinal metaplasia, dysplasia, genetic association, Helicobacter pylori signaling/infection/mechanism, mouse/murine model*) se colectaron un conjunto de artículos científicos sobre modelos murinos de carcinogénesis gástrica, mecanismos de infección de *Helicobacter pylori* en la mucosa gástrica humana y SNPs previamente asociados con CG por estudios GWAS y con la evolución de las LPGs. Siendo las dos primeras las fuentes para la selección de genes candidatos de carcinogénesis gástrica, y la última una fuente para replicar SNPs previamente asociados.

Los modelos murinos de carcinogénesis gástrica han sido establecidos mediante la sobre-expresión transgénica o delección de genes relevantes en la fisiología gastrointestinal unido a la infección por *Helicobacter sp.*, causando la aparición de LPGs y CG en el ratón<sup>41</sup> (Tabla 6). La segunda fuente de selección de genes candidatos fue a partir del mecanismo de infección de *Helicobacter pylori*, constituido por proteínas humanas que interactúan con factores de virulencia y otras proteínas de esta bacteria induciendo la expresión de citoquinas pro-

inflamatorias y mitógenos, responsables de la inflamación crónica y la desregulación de la proliferación epitelial causada por la infección bacteriana (Tabla 7).

Debido a la presencia de una gran cantidad de genes en nuestra lista inicial se estableció un criterio de prioridad teniendo en cuenta la plausibilidad biológica. En el caso de los modelos murinos se les asignó prioridad 1 a aquellos genes cuya sobre-expresión o delección reproducía la cascada de Correa completamente, manifestando al unísono alguna LPGs y CG en el mismo modelo animal. En los casos en que solamente alguna LPG se encontraba presente pero no progresaba a cáncer gástrico se le asignó prioridad 2, con excepción de los genes *CDX1* y *CDX2* que poseen numerosos estudios funcionales y de expresión en LPGs y CG en humanos, por lo que se le confirió prioridad 1.

Con respecto a la vía de señalización de *Helicobacter pylori*, en el caso de observar repetidamente artículos con evidencia de la participación de dicho gen/proteína en la interacción con *Helicobacter pylori* se le asignó prioridad 1 al gen. En el caso de resultados contradictorios acerca del rol de la proteína en dicha vía (Ej. TLR4) o si se planeaba estudiar otra proteína con función similar se le asignó prioridad 2. En todos los casos, solamente se seleccionaron genes con prioridad 1 para el genotipado (Tablas 6 y 7). Para añadir otra evidencia al proceso de selección de estos genes se visitó el catálogo de estudios de asociación de genoma completo (<http://www.genome.gov/gwastudies/>) y se comprobó si existían resultados significativos de asociación publicados para los mismos, obteniéndose una confirmación adicional para su selección.

Además, se seleccionaron genes conteniendo SNPs significativos de estudios GWAS en cáncer gástrico y otros procedentes de estudios de asociación de genes candidatos; así como, de un análisis previo de asociación resultado de un estudio de seguimiento de carcinogénesis gástrica realizado por nuestro grupo en la provincia de Soria<sup>57</sup> (Tabla 8).

Tabla 6. Genes candidatos seleccionados a partir de estudios de modelos murinos de carcinogénesis gástrica.				
Gen y referencia	Función	Características de la expresión en modelo murino <sup>a</sup>	Fenotipo del modelo murino	Asociación en GWAS
<i>IL1B</i> <sup>41</sup>	Citoquina pro-inflamatoria que induce hipoclorhidria, maduración y proliferación de células B, diferenciación y apoptosis	Transgénico de IL1B humana expresada en estómago murino bajo el promotor H-K+	Atrofia-Metaplasia-Displasia-Adenocarcinoma	-----
<i>RUNX3</i> <sup>161</sup>	Factor de transcripción supresor tumoral en estómago. Se encuentra frecuentemente silenciado transcripcionalmente en CG	Knock-out	Hiperplasia-Metaplasia-Adenocarcinoma	Enfermedad de Crohn
<i>GAST</i> <sup>162</sup>	Estimula la secreción de HCl por las células parietales y mitógeno del epitelio gástrico	Transgénico y Knock-out	Atrofia-Metaplasia-Displasia-Adenocarcinoma	-----
<i>TFPI</i> <sup>163</sup>	Estabiliza el mucus gástrico posiblemente por estabilización de glicoproteínas y contribuye a la cicatrización epitelial	Knock-out	Atrofia-Hiperplasia-Displasia- Carcinoma Intramucosal	Cáncer de páncreas
<i>WNT1</i> <sup>164</sup>	Ligando de la vía de señalización Wnt, implicada en oncogénesis, regulación de la localización sub-celular durante la embriogénesis	Transgénico bajo el promotor queratina 19	Epitelio indiferenciado con infiltración de macrófagos	-----
<i>PTGS2 (COX2)</i> <sup>164</sup>	Síntesis de prostaglandinas implicadas en inflamación y mitogénesis. Blanco de AINES	Transgénico bajo el promotor queratina 19	Metaplasia-Displasia-Adenocarcinoma	Osteoartritis
<i>PTGES</i> <sup>165</sup>	Cataliza la oxido reducción de endoperóxido de prostaglandina (PGH2) a prostaglandina (PGE2)	Transgénico en mucosa gástrica	Metaplasia-Hiperplasia-Adenocarcinoma	-----
<i>CDX1</i> <sup>166</sup>	Factor de transcripción que regula la expresión de genes intestinales, diferenciación de los enterocitos y diferenciación del intestino	Transgénico en mucosa gástrica	Metaplasia Intestinal	-----
<i>CDX2</i> <sup>167</sup>	Factor de transcripción que regula el desarrollo embrionario temprano del tracto intestinal. Mantenimiento del revestimiento epitelial de los intestinos delgado y grueso	Transgénico en mucosa gástrica	Metaplasia Intestinal	-----

<sup>a</sup> Especifica si el gen se sobre-expresa o silencia en la mucosa gástrica murina.

Tabla 7. Genes candidatos seleccionados a partir del análisis de los mecanismos de infección de <i>Helicobacter pylori</i> .		
Gen y referencia	Función	Asociación en GWAS
<i>CD14</i> <sup>168 169</sup>	Receptor del lipopolisacárido bacteriano expresado en la superficie de monocitos y macrófagos. Cooperación con otras proteínas para inducir la respuesta inmune	-----
<i>MAP3K14</i> <sup>170</sup>	Quinasa de tipo serina/treonina constituyente de vías de señalización ERK y JNK. Activa la vía NFκB mediante la activación de IKKα (CHUK) e IKKβ (IKBKB) por interacción con TRAF2 y TRAF6	Esclerosis múltiple
<i>MAPK3</i> <sup>171</sup>	Controla la expresión y actividad de oncogenes Fos y Jun, es activada por MEK1/2	Rasgos antropométricos
<i>NFKB1</i> <sup>172</sup>	Factor de transcripción que induce expresión de citoquinas pro-inflamatorias. Su activación inapropiada está asociada con enfermedades inflamatorias	Enfermedad inflamatoria intestinal, Esquizofrenia
<i>NFKBIA</i> <sup>170</sup>	Inhibidor de NFKB1	Psoriasis
<i>NOD1</i> <sup>170</sup>	Receptor citosólico de peptidoglicano, el cual inicia una respuesta inflamatoria ante su detección. Potencia la apoptosis e induce actividad de la vía NFκB	-----
<i>PTPN11</i> <sup>21</sup>	Forma complejo con factor de virulencia de <i>H. pylori</i> cagA, que trae como consecuencia la dispersión, elongación y separación de las células epiteliales	Artritis reumatoide, Tetralogía de Fallot, Megacariopoesis y formación de plaquetas, Presión sanguínea, Diabetes tipo 1

<b>Tabla 8. Genes candidatos seleccionados a partir de estudios de asociación genética con cáncer gástrico (GWAS) y con lesiones precursoras gástricas (genes candidatos).</b>		
<b>Gen y referencia</b>	<b>Función</b>	<b>Fuente de selección</b>
<i>LRFN2</i> <sup>173</sup>	Promueve el crecimiento de neuritas en las neuronas del hipocampo. Potencia la expresión de los receptores N-metil-D-aspartato GRIN1 y GRIN2	GWAS en CG
<i>MUC1</i> <sup>174</sup>	Constituyente del mucus protector de la mucosa gástrica contra bacterias, enzimas y ácido clorhídrico. Su sobre-expresión, localización intracelular y cambios en su glicosilación han sido asociados con diversos carcinomas	GWAS en CG
<i>MUC2</i> y <i>MUC6</i> <sup>140</sup>	Idéntico a anterior	Asociación con evolución de LPGs
<i>PSCA</i> <sup>118</sup>	Glicoproteína de membrana tipo glicosil-fosfatidil-inositol sobre expresada en CG. Participa en la regulación de la proliferación celular	GWAS en CG
<i>PRKAA1</i> <sup>121</sup>	Sensor de estado energético celular y regulador de la polaridad celular por remodelación del citoesqueleto de actina	GWAS en CG
<i>DNAH11</i> <sup>173</sup>	ATPasa que participa en movimiento de cilios respiratorios	GWAS en CG
<i>ZBTB20</i> <sup>121</sup>	Posiblemente es un factor de transcripción con funciones en hematopoiesis, oncogenesis y respuesta inmune	GWAS en CG
<i>CDH1</i> (Cadherina E)	Participa en la adhesión y movilidad de epitelio gástrico. Su pérdida de función contribuye a la progresión tumoral por incremento de la proliferación, invasión y metástasis	Asociación con evolución LPGs <sup>140</sup> (datos no publicados)
<i>IL1RN</i>	Antagonista del receptor de IL1A e IL1B, modulando respuestas inflamatorias e inmunes mediadas por estas citoquinas	Asociación con evolución LPGs <sup>140</sup> (datos no publicados)



## 2.9. Selección de tagSNPs en los genes candidatos y análisis bioinformáticos

Haciendo uso del navegador del sitio web del proyecto HapMap (<http://hapmap.ncbi.nlm.nih.gov/>) en población de origen europeo (CEU) *release* #28 (*Phases 1, 2 & 3 - merged genotypes & frequencies, db SNP b 126*) fueron descargados los datos genotípicos de cada gen candidato seleccionado; así como 5 y 3 kilobases adicionales de sus regiones 5' y 3'. Posteriormente, dichos datos genotípicos fueron cargados en el programa informático Haploview versión 4.2 y seleccionados los SNPs con  $MAF \geq 0.05$ . Para la definición de bloques en LD (haplotipos) se aplicó el método de intervalos de confianza de Gabriel *et al.*<sup>92</sup>, utilizando los parámetros por defecto en Haploview v.4.2. Los tagSNPs fueron seleccionados utilizando el método *Aggressive Tagger* con un umbral de  $r^2 \geq 0.8$  entre cada par de tagSNPs y SNPs etiquetado; así como, haplotipos etiquetados con  $MAF \geq 0.05$ <sup>93</sup>. Para los polimorfismos significativos extraídos de GWAS y de estudios de asociación con la evolución de las LPGs de Soria<sup>140</sup> no se realizó selección de tagSNPs. Los análisis funcionales *in silico* de los SNPs genotipados fueron realizados como se describió en el apartado 1.5 de esta sección. La cobertura de la variabilidad de tagSNPs en las regiones genéticas fue obtenida por Haploview eliminando los polimorfismos que no eran tagSNPs del total de los existentes en los bloques de haplotipos (Tabla 9).

## 2.10. Genotipado de SNPs

La plataforma de genotipado usada fue la Sequenom, en el CEGEN del Instituto Carlos III. De todos los tagSNPs inicialmente seleccionados, todos los presentes en el gen *TNF* (N=13), y algunos en los genes *MAP3K14* (rs17686001, rs4792849, rs7216796) e *IL1B* (rs3136558) no pudieron ser genotipados debido a que no se pudo diseñar ningún ensayo para los mismos. Además, 7 SNPs inicialmente seleccionados no pudieron ser genotipados debido a generar clusters continuos o solapantes. Así, el total de SNPs fue de 141 en 28 genes en 559 pacientes. De los cuales, 131 fueron tagSNPs y 10 SNPs previamente asociados a CG o a la evolución de lesiones precursoras (Tabla 9). Como controles de calidad fueron duplicadas al azar 30 muestras de ADN, además de un trío de muestras Coriell (Na10830, Na10831 y Na12147) como controles positivos de la técnica.

## 2.11. Análisis estadísticos de los resultados de genotipado de SNPs

Los pacientes se dividieron en tres grupos: de progresión, regresión y estables; de acuerdo a si el diagnóstico histológico entre la biopsia de reclutamiento y el del fin del seguimiento se incrementó, decrementó o se mantuvo igual respecto a la severidad de las LPGs (Figura 10). Se utilizó la prueba de  $X^2$  de Pearson para examinar diferencias entre las variables al reclutamiento de la población entre los grupos de progresión, regresión y estables. Se utilizó la prueba  $X^2$  de bondad de ajuste para chequear si los SNPs genotipados se encuentran en equilibrio de Hardy-

Weinberg. Los OR con IC95% se calcularon mediante regresión logística no condicional sin incluir covariables en el ajuste, método implementado en la librería SNPassoc del programa informático R<sup>148</sup>. Se calcularon OR en dos tipos de análisis: (i) comparación del grupo de pacientes cuyas lesiones han progresado hacia una lesión de mayor severidad con respecto a aquellos cuyas lesiones se mantuvieron estables o experimentaron regresión (grupo de referencia) y (ii) aquellos pacientes que han regresado a una lesión de menor severidad respecto a los se mantuvieron estables o experimentaron progresión (grupo de referencia).

Se realizaron todos los análisis usando los modelos de herencia codominante, dominante, recesivo y log-aditivo (por alelo). Posteriormente, se aplicó la corrección para comparaciones múltiples *False Discovery Rate* (FDR) basada en la tasa de falsos positivos que genera un q-valor corregido<sup>175</sup>. Debido a que los portadores de la combinación de factores de virulencia de *H. pylori* cagA+vacAs1/m1 tienen mayor riesgo de CG se realizó un análisis estratificado de asociación con la evolución de las LPGs según el estatus cagA+vacAs1/m1 (N=131) respecto al resto de combinaciones de cagA y vacA (N=230).

El desequilibrio de ligamiento entre polimorfismos en cada región génica fue explorado mediante el programa Haploview. Las frecuencias haplotípicas fueron estimadas mediante el método de expectación-maximización implementado en el paquete estadístico haplo.stats de R. La asociación entre la progresión o regresión y cada haplotipo con una frecuencia de 0.01 se analizó utilizando el haplotipo más frecuente entre los controles como referencia. En todos los casos, un resultado de  $p < 0.05$  se consideró estadísticamente significativo.

Tabla 9. SNPs genotipados con su información asociada.

Selección <sup>a</sup>	Gen	SNP <sup>b</sup>	Cromosoma <sup>c</sup>	Posición <sup>d</sup>	Alelos <sup>e</sup>	MAF <sup>f</sup>	Variant Effect Predictor <sup>g</sup>	PupaSuite <sup>h</sup>	HWE <sup>i</sup>	Cobertura <sup>j</sup>
A	CDX1	rs6894617	5	149521614	C/T	0,259	Región 5', RR	TFBS/Conservada	0,397	92 %
A	CDX1	rs2302275	5	149527021	C/G	0,47	Cambio de a.a	Región conservada	0,852	
A	CDX1	rs10040224	5	149527350	C/G	0,062	Intrónico	Región conservada	1,000	
A	CDX1	rs887343	5	149529016	G/A	0,366	Intrónico	Región conservada	0,550	
A	CDX1	rs2282812	5	149533940	G/A	0,366	Intrónico	-----	1,000	
A	CDX1	rs10063514	5	149546502	C/T	0,068	Región 3'	Región conservada	0,712	
A	CDX2	rs2481952	13	27434901	C/T	0,498	UTR 3', RR	Región conservada	0,309	100 %
A	CDX2	rs2504212	13	27435537	C/A	0,211	Intrónico	Región conservada	0,008	
A	CDX2	rs4503658	13	27442919	G/T	0,08	UTR 5', RR	Región conservada	0,006	
A	CDX2	rs3812863	13	27443268	A/G	0,398	UTR 5', RR	Región conservada	0,441	
A	GAST	rs12453761	17	37120839	A/G	0,48	Región 5'	Región conservada	0,405	30 %
E	IL1A	rs17561	2	113253684	G/T	0,289	aa deletéreo	Región conservada	0,178	No es tagSNP
A	IL1B	rs3917366	2	113300351	G/T	0,246	Región 3'	Región conservada	0,381	88 %
A	IL1B	rs1143633	2	113306938	G/A	0,382	Intrónico	Región conservada	0,168	
A	IL1B	rs1143627	2	113310858	T/C	0,338	Región 5', RR	Región conservada	0,409	
A	PTGES	rs12001450	9	131536460	T/C	0,192	Región 3'	-----	0,073	100 %
A	PTGES	rs4636306	9	131536838	G/A	0,209	Región 3'	-----	0,567	
A	PTGES	rs2302821	9	131541702	A/C	0,073	UTR 3'	Región conservada	0,294	
A	PTGES	rs10760634	9	131543692	A/G	0,078	Intrónico	Región conservada	1,000	
A	PTGES	rs4837404	9	131545480	A/G	0,369	Intrónico, RR	Región conservada	1,000	
A	PTGES	rs10739757	9	131547731	T/C	0,129	Intrónico	-----	1,000	
A	PTGES	rs7872802	9	131555551	A/G	0,134	Región 5'	-----	0,841	
A	PTGS2	rs2206593	1	184909052	G/A	0,052	UTR 3'	Región conservada	1,000	100 %
A	PTGS2	rs5275	1	184909681	T/C	0,319	UTR 3'	Región conservada	0,454	
A	PTGS2	rs4648276	1	184912111	T/C	0,172	Intrónico, NMD	Región conservada	1,000	

Selección <sup>a</sup>	Gen	SNP <sup>b</sup>	Cromosoma <sup>c</sup>	Posición <sup>d</sup>	Alelos <sup>e</sup>	MAF <sup>f</sup>	Variant Effect Predictor <sup>g</sup>	PupaSuite <sup>h</sup>	HWE <sup>i</sup>	Cobertura <sup>j</sup>
A	PTGS2	rs5277	1	184914820	G/C	0,192	Sinónimo, NMD	Región conservada	0,373	
A	PTGS2	rs2745557	1	184915844	G/A	0,204	Intrónico, NMD,	Región conservada	0,475	
A	PTGS2	rs689466	1	184917374	A/G	0,175	Región 5', RR	Región conservada	0,201	
A	PTGS2	rs12042763	1	184918499	G/T	0,273	Región 5', RR	Triple hélice, conservada	0,295	
A	RUNX3	rs6663310	1	25094966	T/C	0,334	Región 3'	-----	0,145	91 %
A	RUNX3	rs9438858	1	25096492	C/A	0,391	Región 5'	Región conservada	0,098	
A	RUNX3	rs7553295	1	25096597	G/T	0,238	Región 3'	Región conservada	0,799	
A	RUNX3	rs9438859	1	25096908	T/C	0,232	Región 3', RR	Conservada, RR	0,297	
A	RUNX3	rs7514332	1	25098262	G/C	0,054	Region 3'	Región conservada	0,384	
A	RUNX3	rs2236850	1	25112928	T/C	0,408	Intrónico, RR	Región conservada	0,503	
A	RUNX3	rs2282718	1	25113643	G/A	0,335	Intrónico, RR	Región conservada	1,000	
A	RUNX3	rs9438876	1	25113703	A/G	0,489	Intrónico, RR	Región conservada	0,116	
A	RUNX3	rs2236852	1	25116354	A/G	0,486	Intrónico	-----	0,579	
A	RUNX3	rs7517302	1	25126904	T/C	0,362	Intrónico, RR	Región conservada	0,616	
A	TFF1	rs225352	21	42650501	C/T	0,086	Region 3'	-----	1,000	100 %
A	TFF1	rs178740	21	42650741	T/C	0,198	Region 3'	-----	0,012	
A	TFF1	rs1547374	21	42651964	A/G	0,312	Region 3'	-----	0,516	
A	TFF1	rs4919984	21	42653915	A/G	0,222	Region 3'	Región conservada	1,000	
A	TFF1	rs225355	21	42657551	T/C	0,492	Intrónico	-----	0,114	
A	TFF1	rs225357	21	42659382	C/T	0,431	Intrónico	Región conservada	0,638	
A	TFF1	rs4920094	21	42659402	G/A	0,428	Intrónico	Región conservada	0,158	
A	TFF1	rs225358	21	42659581	T/C	0,303	Intrónico, región de empalme	Región conservada	0,319	
A	TFF1	rs2156310	21	42659675	G/A	0,151	UTR 5'	TFBS/Triple hélice/ conservada	1,000	
A	TFF1	rs225359	21	42660505	A/G	0,327	Región 5'	-----	0,833	

Selección <sup>a</sup>	Gen	SNP <sup>b</sup>	Cromosoma <sup>c</sup>	Posición <sup>d</sup>	Alelos <sup>e</sup>	MAF <sup>f</sup>	Variant Effect Predictor <sup>g</sup>	PupaSuite <sup>h</sup>	HWE <sup>i</sup>	Cobertura <sup>j</sup>
A	<i>TFE1</i>	rs1788413	21	42661759	A/C	0,162	Region 3'	-----	0,864	
A	<i>TFE1</i>	rs9976977	21	42662543	G/A	0,369	Region 3'	-----	0,112	
A	<i>TFE1</i>	rs4242694	21	42662841	C/T	0,413	Region 3'	-----	0,445	
A	<i>TFE1</i>	rs13047838	21	42665938	C/T	0,257	UTR 3'	Conservada, Triple hélice	0,183	
D	<i>TFE2</i>	rs1079380	21	42643342	A/G	0,489	Intrónico	Región conservada	0,309	No es tagSNP
A	<i>WNT1</i>	rs4760663	12	47654900	C/T	0,429	Region 3'	-----	0,571	100 %
B	<i>MAP3K</i>	rs4247364	17	40692470	C/G	0,349	Region 3'	-----	0,476	81 %
B	<i>MAP3K</i>	rs1047841	17	40696414	G/A	0,055	UTR 3'	Región conservada	1,000	
B	<i>MAP3K</i>	rs16939926	17	40699032	T/C	0,055	Intrónico	Región conservada	1,000	
B	<i>MAP3K</i>	rs2074293	17	40699856	T/C	0,434	Intrónico	Región conservada	0,005	
B	<i>MAP3K</i>	rs3785803	17	40716996	C/T	0,114	Intrónico	Región conservada	0,356	
B	<i>MAP3K</i>	rs2074292	17	40717274	C/T	0,495	Intrónico	Región conservada	0,009	
B	<i>MAP3K</i>	rs9908076	17	40731893	C/T	0,067	Intrónico	-----	0,711	
B	<i>MAP3K</i>	rs2867316	17	40732230	C/T	0,362	Intrónico	-----	0,035	
B	<i>MAPK3</i>	rs7698	16	30033301	C/T	0,13	UTR 3', NMD	Región conservada	1,000	100 %
B	<i>MAPK3</i>	rs11865086	16	30037994	A/C	0,491	Intrónico, NMD	-----	0,516	
B	<i>NFKB1</i>	rs980455	4	103637989	A/G	0,391	Region 3'	-----	0,173	98 %
B	<i>NFKB1</i>	rs13117745	4	103697738	C/T	0,122	Intrónico	Región conservada	0,085	
B	<i>NFKB1</i>	rs1598861	4	103705733	A/C	0,365	Intrónico	Región conservada	0,003	
B	<i>NFKB1</i>	rs1610152	4	103725412	G/C	0,046	Intrónico, RR	-----	0,614	
B	<i>NFKB1</i>	rs11722146	4	103743667	G/A	0,268	Intrónico, RR	-----	0,906	
B	<i>NFKB1</i>	rs4648090	4	103746106	G/A	0,152	Intrónico	Región conservada	0,858	
B	<i>NFKB1</i>	rs4648127	4	103754951	C/T	0,033	Intrónico	-----	0,402	
B	<i>NFKB1</i>	rs230547	4	103755307	C/T	0,105	Intrónico, RR	-----	0,623	
B	<i>NFKB1</i>	rs4648135	4	103755716	A/G	0,051	Intrónico	-----	0,624	
B	<i>NFKB1</i>	rs7674640	4	103759828	T/C	0,485	Region 3', RR	Región conservada	0,228	

Selección <sup>a</sup>	Gen	SNP <sup>b</sup>	Cromosoma <sup>c</sup>	Posición <sup>d</sup>	Alelos <sup>e</sup>	MAF <sup>f</sup>	Variant Effect Predictor <sup>g</sup>	PupaSuite <sup>h</sup>	HWE <sup>i</sup>	Cobertura <sup>j</sup>
B	NFKBIA	rs3138056	14	34938265	G/A	0,319	Región 3', RR	Región conservada	0,670	100 %
B	NFKBIA	rs4982269	14	34939618	T/C	0,338	Región 3'	Región conservada	0,536	
B	NFKBIA	rs696	14	34940844	G/A	0,372	UTR 3', NMD	Región conservada	0,321	
B	NFKBIA	rs3138054	14	34942058	G/A	0,154	Intrónico, RR	Región conservada	0,859	
B	NFKBIA	rs3138053	14	34944605	A/G	0,284	Región 3'	Región conservada	0,820	
B	NOD1	rs11536450	7	30434089	C/G	0,13	Intrónico, NMD	-----	1,000	100 %
B	NOD1	rs2907749	7	30452266	A/G	0,295	Intrónico, RR	-----	0,074	
B	NOD1	rs7789045	7	30460547	T/A	0,463	Intrónico, NMD	Región conservada	0,641	
B	NOD1	rs2906766	7	30466100	T/C	0,302	UTR 5', RR	Región conservada	0,661	
B	NOD1	rs3823773	7	30466486	A/G	0,087	Intrónico, RR	-----	0,768	
B	NOD1	rs4272257	7	30470374	C/T	0,157	Intrónico, NMD,	-----	1,000	
B	NOD1	rs2709803	7	30477059	C/T	0,217	Intrónico, NMD	Región conservada	1,000	
B	NOD1	rs4720004	7	30478744	T/C	0,144	Intrónico, NMD	Región conservada	0,706	
B	NOD1	rs2256023	7	30481516	T/C	0,426	Intrónico, NMD	Región conservada	0,924	
B	NOD1	rs17770244	7	30482337	A/G	0,061	Intrónico, NMD	Región conservada	1,000	
B	PTPN1	rs17822304	12	111373075	C/A	0,049	Intrónico	Región conservada	1,000	100 %
B	PTPN1	rs11614544	12	111389338	A/G	0,154	Intrónico, RR	-----	0,285	
B	PTPN1	rs11066320	12	111390798	G/A	0,437	Intrónico	-----	0,778	
B	PTPN1	rs11066322	12	111406912	G/A	0,189	Intrónico	-----	0,128	
B	PTPN1	rs11066323	12	111407744	G/A	0,103	Intrónico	-----	0,613	
B	PTPN1	rs7958372	12	111419936	C/T	0,085	Intrónico	-----	0,561	
B	PTPN1	rs7953150	12	111425954	G/A	0,286	Intrónico, RR	-----	0,821	
B	SRC	rs6017901	20	35401507	T/C	0,043	Región 3'	-----	0,583	88 %
B	SRC	rs6094373	20	35402488	T/C	0,221	Región 3'	Región conservada	0,418	
B	SRC	rs6017916	20	35404692	A/C	0,251	Región 3', RR	Región conservada	0,712	
B	SRC	rs12106024	20	35412762	C/G	0,203	Intrónico, RR	Región conservada	0,115	
B	SRC	rs1547836	20	35414365	G/C	0,188	Intrónico, RR	Región conservada	0,878	
B	SRC	rs16986606	20	35417842	A/G	0,146	Intrónico, RR	Región conservada	0,263	

Selección <sup>a</sup>	Gen	SNP <sup>b</sup>	Cromosoma <sup>c</sup>	Posición <sup>d</sup>	Alelos <sup>e</sup>	MAF <sup>f</sup>	Variant Effect Predictor <sup>g</sup>	PupaSuite <sup>h</sup>	HWE <sup>i</sup>	Cobertura <sup>j</sup>	
B	SRC	rs6017996	20	35420067	G/A	0,144	Región 3', RR	-----	0,851	88%	
B	SRC	rs6018027	20	35424206	T/C	0,275	Intrónico	-----	0,488		
B	SRC	rs6063022	20	35426741	C/T	0,16	Intrónico	-----	1,000		
B	SRC	rs6018088	20	35432039	C/A	0,136	Intrónico	-----	0,692		
B	SRC	rs6018148	20	35439194	C/T	0,204	Intrónico	-----	0,670		
B	SRC	rs7269342	20	35441575	T/C	0,146	Intrónico, RR	-----	0,710		
B	SRC	rs3790150	20	35448513	A/G	0,163	Intrónico	Región conservada	0,395		
B	SRC	rs754626	20	35450754	A/C	0,223	Intrónico	-----	0,503		
B	SRC	rs754625	20	35450938	T/C	0,288	Intrónico	-----	0,821		
B	SRC	rs6018257	20	35455953	T/C	0,122	Intrónico, RR	Región conservada	0,385		
B	SRC	rs1570209	20	35462245	T/C	0,102	Intrónico	-----	1,000		
B	SRC	rs17785475	20	35466307	C/T	0,067	UTR 3'	Región conservada	0,711		
C	LRFN2	rs2494938	6	40644106	A/G	0,462	Intrónico	-----	0,263		No es tagSNP
C	MUC1	rs2070803	1	153424339	T/C	0,43	Región 3'	Región conservada	0,089		No es tagSNP
C	PSCA	rs2294008	8	143758933	C/T	0,437	Intrónico, UTR	Región conservada	0,349		No es tagSNP
C	PRKAA	rs13361707	5	40827641	C/T	0,268	Intrónico, RR	Región conservada	0,156		No es tagSNP
C	DNAH1	rs2285947	7	21550613	G/A	0,488	Intrónico	Región conservada	0,194	No es tagSNP	
C	ZBTB20	rs9841504	3	115845454	C/G	0,059	Intrónico, RR	Región conservada	0,209	No es tagSNP	
B	CD14	rs11167532	5	139968081	G/A	0,45	Región 5'	-----	0,075	88 %	
B	CD14	rs778588	5	139987195	T/C	0,294	Región 3'	-----	0,181		
B	CD14	rs2569190	5	139993100	A/G	0,494	Intrónico, UTR 5', RR	Conservado/Triple hélice	0,459		
B	CD14	rs5744455	5	139993491	C/T	0,2	Región 5', RR	Región conservada	0,384		
B	CD14	rs1583005	5	140011721	C/T	0,418	Intrónico, NMD	-----	0,390		
B	CDH1	rs16260	16	67328535	C/A	0,275	Región 5'	Región conservada	0,907	60 %	
B	CDH1	rs1125557	16	67367100	A/G	0,433	Intrónico, NMD	Región conservada	0,345		
B	CDH1	rs12597188	16	67372327	G/A	0,317	Intrónico, NMD	-----	1,000		

Selección <sup>a</sup>	Gen	SNP <sup>b</sup>	Cromosoma <sup>c</sup>	Posición <sup>d</sup>	Alelos <sup>e</sup>	MAF <sup>f</sup>	Variant Effect Predictor <sup>g</sup>	PupaSuite <sup>h</sup>	HWE <sup>i</sup>	Cobertura <sup>j</sup>
B	CDH1	rs9929218	16	67378447	G/A	0,292	Intrónico, NMD,	-----	0,576	
B	CDH1	rs7186053	16	67396794	G/A	0,423	Intrónico, NMD	-----	0,705	
B	CDH1	rs4783573	16	67398089	A/G	0,356	Intrónico, NMD	-----	0,419	
D	IL1RN	rs2637988	2	113593250	A/G	0,389	Intrónico, Región 3'	-----	0,041	No es tagSNP
D	MUC2	rs10902073	11	1050934	C/A	0,336	Intergénico, RR	-----	0,097	34 %
D	MUC2	rs10794281	11	1053149	T/C	0,397	Intergénico	Región conservada	0,176	
D	MUC2	rs2071174	11	1063712	T/C	0,29	Región 5', RR	-----	1,000	
D	MUC2	rs7944723	11	1083710	C/G	0,218	Exónico	-----	0,105	
D	MUC2	rs10794293	11	1088939	C/T	0,347	Intrónico,	-----	0,358	
D	MUC2	rs3924453	11	1095806	G/A	0,263	Región 3',	-----	0,341	
D	MUC2	rs4077759	11	1095976	T/C	0,353	Región 3',	-----	0,361	
D	MUC6	rs6597947	11	1027029	G/T	0,079	Región 3'	Región conservada	0,005	No es tagSNP

<sup>a</sup> Se refiere a la procedencia de los genes candidatos. A: modelos murinos de carcinogénesis gástrica, B: Mecanismo de infección por *H. pylori*, C: Genes asociados a GWAS en CG, D: Genes asociados a la evolución de LPGs en estudio de seguimiento de Soria <sup>140</sup>. <sup>b</sup> Los SNPs se ordenan en sentido 5'-3'. <sup>c-f</sup> Información de HapMap. <sup>e</sup> Alelo común/alelo variante. <sup>f</sup> MAF: Frecuencia del alelo menor en la muestra total estudiada. <sup>g-h</sup> Efecto funcional predicho de programas Variant Effect Predictor y PupaSuite respectivamente. NMD significa que el SNP se encuentra localizado en transcrito sujeto a mecanismo de degradación de ARNm *Nonsense mediated mRNA decay*. Región conservada se refiere a una secuencia de ADN conservada evolutivamente entre ratón y humano. TFBS se refiere a sitios de unión de factores de transcripción. Triple hélice se refiere a la formación de esa estructura en el ADN como consecuencia del alelo variante. Región regulatoria se refiere a aquella que afecta la expresión génica. ----- Significa que la herramienta no predijo dicho SNP. <sup>i</sup> p-valor de equilibrio de Hardy-Weinberg. <sup>j</sup> Cobertura de variabilidad de tagSNPs en la región genética.



### 3. Estudio de expresión génica por microarray en metaplasia intestinal que progresa a cáncer gástrico

#### 3.1. Pacientes y muestras

Se analizaron muestras de un estudio de seguimiento de lesiones precursoras gástricas realizado en la provincia de Soria, España, por nuestro grupo de trabajo <sup>57</sup>. El diseño de estudio es equivalente al descrito en el apartado 2.1 de esta sección; pero en este caso restringido a la provincia de Soria, en lugar de toda España. En resumen, los pacientes se sometieron a gastroscopia con colecta de biopsia gástrica durante el periodo de 1988-1994 y aquellos con diagnóstico histológico de GNA, GCA, MIC o MII fueron invitados a someterse a una nueva biopsia durante 2005-2007. De esta manera se obtuvo la información de seguimiento de los cambios histológicos en la mucosa gástrica a lo largo de un intervalo de tiempo de 12.5 años como promedio. Los criterios de exclusión fueron: úlcera péptica prevalente, edad fuera del rango de 25 a 69 años, cáncer gástrico o esófago de Barrett, gastrectomía previa y falta de información. Todos los pacientes firmaron un consentimiento informado accediendo a los procedimientos del estudio, cuyo protocolo fue aprobado por el Comité Ético del hospital Virgen del Mirón de Soria. Un cuestionario sobre la historia del consumo de tabaco, el uso de AINES, los antecedentes familiares de cáncer gástrico y la prevalencia de infección por *Helicobacter pylori* se administró a todos los pacientes. Adicionalmente, se analizaron muestras de MIC y MII que progresaron a cáncer gástrico provenientes del estudio multicéntrico de seguimiento de LPGs realizado en toda España descrito en el apartado 2.1 de esta sección. Además, se utilizaron muestras de MIC y MII que progresaron y que no progresaron a cáncer gástrico (posteriores a 1994 y con seguimiento >1 año) procedentes de los hospitales Gregorio Marañón de Madrid y Virgen del Mirón de Soria. Por último, fueron utilizadas como controles muestras de mucosa gástrica sana provenientes de este último hospital.

La clasificación de la MI (G1-G4) presente en las muestras se realizó combinando el sistema de clasificación de subtipos histológicos de Filipe *et al.* <sup>50</sup> y la descripción de la MII del sistema de clasificación histológico de Sydney <sup>47</sup>. Debido a que en este último no quedan reflejadas las escalas visuales del primero se hizo una adaptación en 4 niveles separando la tipo II según predominara la completa (G2) o la incompleta (G3), como se describe a continuación:

**G1 o MIC:** Exclusiva presencia de células caliciformes. Equivalente al estadio I de la clasificación de Filipe *et al.* <sup>50</sup>.

**G2 o MI predominantemente completa:** Presencia de células intermedias o híbridas entre caliciformes y enterocitos. Predominio de la MIC, aunque pueden existir focos

de MII que no alcanzan la superficie foveolar. Equivalente al estadio II de la clasificación de Filipe *et al.*<sup>50</sup>.

**G3 o MI predominantemente incompleta:** Presencia de focos de MII compuestos por células columnares y caliciformes alcanzando la superficie foveolar. Equivalente al estadio III de la clasificación de Filipe *et al.*<sup>50</sup>.

**G4 o MII:** La MII afecta la casi totalidad de los fragmentos, aunque cabe la presencia de áreas de MIC. Equivalente al estadio III de la clasificación de Filipe *et al.*<sup>50</sup>.

Cuando existió duda sobre el tipo de MI predominante en una biopsia, se consideró que predominaba la MI incompleta si la mitad o más de los fragmentos presentaban MI, bien fuera completa o incompleta. La extensión (E) de la MI se refirió de acuerdo al porcentaje de glándulas metaplásicas presentes en la muestra (E3 para >50% de la muestra y E4 para >75% de la muestra). Tanto el diagnóstico histológico como el de la extensión de la MI fue realizado por patólogos participantes en el estudio de Soria.

Se realizó una prueba piloto de comparación de las plataformas de microarrays Human Gene 2.0 ST y Almac-Xcel con el objetivo de conocer la eficiencia de esta técnica en las muestras de mucosa gástrica parafinada de varios años a emplear (media de 12.5 años). Para ello se utilizaron 3 muestras con MI que progresaron a CG<sup>57</sup> y 3 controles de mucosa gástrica sana (Tabla 10).

**Tabla 10. Muestras usadas para prueba piloto de comparación de microarrays de expresión.**

Código	Tipo de MI al reclutamiento <sup>a</sup>	Extensión de la MI <sup>b</sup>	Localización anatómica	Diagnóstico al final del seguimiento	Estatus <sup>c</sup>	Infección por <i>H. pylori</i>	Sexo <sup>d</sup>
A1	No presenta	No	Cuerpo	Sanos	Sanos	SI	H
A2	No presenta	No	Cuerpo	Sanos	Sanos	NO	H
A3	No presenta	No	Cuerpo	Sanos	Sanos	NO	M
A4	G1	E4	Antro	CG intestinal	MIC-CG	SI	H
A5	G3	E4	Antro	CG intestinal	MII-CG	SI	M
A6	G3	E4	Antro	CG mixto	MII-CG	NO	H

<sup>a</sup> Diagnóstico descrito en el apartado 3.1 de esta sección. <sup>b</sup> E4 significa extensión de la MI  $\geq 75\%$ . <sup>c</sup> MII/MIC-CG: Muestra al reclutamiento de MII o MIC que progresa a CG al final del seguimiento.

Después de confirmar con este estudio piloto que las muestras eran válidas para su análisis mediante microarrays de expresión, se procedió al análisis de los genes y mecanismos funcionales desregulados en la progresión de MI a CG en la totalidad de los casos a estudiar (Tabla 11).

**Tabla 11. Muestras usadas en el estudio de expresión diferencial por microarrays de los subtipos histológicos de metaplasia intestinal que progresan a cáncer gástrico, respecto a los que no progresan y a la mucosa sana.**

Código	Diagnóstico al reclutamiento <sup>a</sup>	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Diagnóstico al final del seguimiento	Años de seguimiento <sup>c</sup>	Estatus <sup>d</sup>	Infección por <i>H. pylori</i>	Sexo <sup>e</sup>	Edad <sup>f</sup>	Precedencia <sup>g</sup>
A7	G4	E4	Antro	CG intestinal	10	MII-CG	NO	M	52	H.Soria
A8	G4	E4	Antro	CG intestinal	8.4	MII-CG	SI	H	69	H.Soria
A9	G4	E3	Antro	CG Difuso	7.7	MII-CG	NO	H	45	H.Soria
A10	G4	E3	Antro	CG intestinal	2	MII-CG	NO	H	73	H.Soria
A11	G4	E3	Antro	CG intestinal	1	MII-CG	NO	H	60	H.Soria
A12	G4	E3	Antro	CG intestinal	11.8	MII-CG	SI	M	66	H.Soria
A13	G3	E4	Antro	CG	13.4	MII-No CG		H	53	H.G Marañoñ
A14	G4	E4	Antro	MII	15	MII-No CG	SI	H	64	H.Soria
A15	G3-G4	E3-E4	Antro	MII	11	MII-No CG	NO	M	64	H.Soria
A16	G3-G4	E4	Antro	MII	ND	MII-No CG	NO	H	64	H.Soria
A17	G3-G4	E3-E4	Antro	MII	14	MII-No CG	SI	H	65	H.Soria
A18	G3-G4	E4	Antro	MII	ND	MII-No CG	NO	M	74	H.Soria
A19	G3-G4	E4	Antro	MII	ND	MII-No CG	NO	H	71	H.Soria
A20	G1	E4	Antro	CG no cardias	2	MIC-CG	ND	H	57	Multicéntrico
A21	G1	E4	Cuerpo	CG intestinal	8	MIC-CG	SI	H	57	Multicéntrico
A22	G1	E4	Antro	CG intestinal	1	MIC-CG	ND	H	44	Multicéntrico
A23	G1	E4	Antro	MIC	15.37	MIC-CG	NO	H	56	H.G Marañoñ
A24	G2	E4	Antro	MIC	14	MIC-CG	SI	M	59.7	Multicéntrico
A25	G1	E4	Antro	MIC	8.24	MIC-CG	NO	M	57	H.G Marañoñ
A26	G1	E4	Antro	MIC	5.95	MIC-CG	NO	M	78	H.G Marañoñ
A27	G1	E4	Antro	CG no cardias	6.2	MIC-CG	NO	H	76	H.G Marañoñ
A28	G1	E4	Transicional	MIC	10.6	MIC-No CG	NO	H	67	H.G Marañoñ
A29	G2	E4	Antro	MIC	10.5	MIC-No CG	SI	H	63	H.G Marañoñ
A30	G2	E4	Transicional	MIC	12.31	MIC-No CG	NO	H	74	H.G Marañoñ
A31	G2	E4	Antro	MIC	13	MIC-No CG	NO	H	56.9	Multicéntrico

Código	Diagnóstico al reclutamiento	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Diagnóstico al final del seguimiento	Años de seguimiento <sup>c</sup>	Estatus <sup>d</sup>	Infección por <i>H. pylori</i>	Sexo <sup>e</sup>	Edad <sup>f</sup>	Procedencia <sup>g</sup>
A32	G1	E4	Antro	MII	14	MIC-No CG	SI	M	54	H.Soria
A33	G1	E4	Antro	MII	15	MIC-No CG	SI	H	57	H.Soria
A34	G1	E4	Antro	MIC	18	MIC-No CG	SI	H	57	H.Soria
A35	G1	E4	Antro	MIC	14	MIC-No CG	SI	H	62	H.Soria
A36	G1	E4	Antro	MIC	15	MIC-No CG	NO	H	62	H.Soria
A37	-----	-----	Antro	-----	-----	Sanos	SI	M	16	H.Soria
A38	-----	-----	Antro	-----	-----	Sanos	NO	H	42	H.Soria
A39	-----	-----	Antro	-----	-----	Sanos	NO	H	39	H.Soria
A40	-----	-----	Antro	-----	-----	Sanos	NO	H	41	H.Soria
A41	-----	-----	Antro	-----	-----	Sanos	NO	M	72	H.Soria
A42	-----	-----	Antro	-----	-----	Sanos	NO	M	86	Multicéntrico
A43	-----	-----	Antro	-----	-----	Sanos	NO	M	18	H. Soria
A44	-----	-----	Cuerpo	-----	-----	Sanos	NO	H	27	H. Soria
A45	-----	-----	Antro	-----	-----	Sanos	NO	H	68	H. Soria
A46	-----	-----	Antro	-----	-----	Sanos	NO	M	31	H. Soria
A47	-----	-----	Antro	-----	-----	Sanos	NO	M	19	H. Soria
A48	-----	-----	Cuerpo	-----	-----	Sanos	NO	H	73	H. Soria
A49	-----	-----	Cuerpo	-----	-----	Sanos	NO	M	58	H. Soria
A50	-----	-----	Antro	-----	-----	Sanos	NO	M	11	H. Soria
A51	-----	-----	Cuerpo	-----	-----	Sanos	SI	H	69	H. Soria

<sup>a</sup> Tipos histológicos de MI (G1-G4) descritos en el apartado 3.1 de esta sección. <sup>b</sup> E3>50%, E4>75% se refiere a la extensión de la MI en la muestra; en el caso de que existan dos diferentes en <sup>a</sup> ó <sup>b</sup> en la misma muestra se corresponden a los cortes inicial y final. <sup>c</sup> Se refiere al intervalo de tiempo entre el reclutamiento y el final del seguimiento. <sup>d</sup> MII/MIC-CG: Muestra de MII o de MIC al reclutamiento que progresa a CG al final del seguimiento. MII/MIC-No CG son controles de MII y MIC que no progresan a CG desde el reclutamiento hasta el final de seguimiento. <sup>e</sup> M: Mujer, H: Hombre. ND: No disponible. <sup>f</sup> Edad al reclutamiento. <sup>g</sup> Hospital o proyecto de procedencia de las muestras. Multicéntrico hace referencia al proyecto de seguimiento de LPGs realizado en varios hospitales de España.

Se seleccionaron 7 muestras de MII tipo G4 incluidas en tejido gástrico parafinado que progresaron a CG (MII-CG) durante un seguimiento de entre 1 y 12.5 años, con localización anatómica en el antro y con una extensión de la MI entre E3 y E4. Como controles se seleccionaron 6 muestras de MII que no progresaron a CG (MII-No CG). Posteriormente, para estudiar ambos subtipos histológicos de MI, se seleccionaron 8 muestras de MIC que después de un seguimiento medio de 10 años progresaron a CG (MIC-CG) y 7 que no progresaron (MIC-No CG). Estas muestras presentaron diferentes localizaciones anatómicas y ambos tipos de MIC (G1 o G2), pero todas fueron de extensión E4. En ambos casos, los controles que no progresaron a CG de cada subtipo histológico fueron seleccionados con características similares de clasificación histológica (G1-G4), extensión (E3-E4) y localización anatómica de la MI, sexo y edad como variables de ajuste. Así, el total de muestras analizadas fue de 45: MII-CG (N=6), MII-No CG (N=7), MIC-CG (N=8), MIC-No CG (N=9) y mucosa gástrica sana (N=15).

### **3.2. Corte de bloques de parafina y evaluación histológica**

Los bloques de tejido parafinado fueron enfriados en nieve carbónica y cortados en micrótopo. Se realizaron cortes iniciales y finales a 3 µm de espesor para realizar tinción con hematoxilina-eosina y confirmar el diagnóstico histológico.

El tejido entre ambos cortes fue cortado a 5 µm de espesor para la extracción de ARN total. Para evitar la contaminación se cambió la cuchilla entre cada bloque y se limpió el micrótopo con Histoclear (HS-202, National Diagnostics, EUA), los guantes también se cambiaron frecuentemente. Un control negativo de tejido parafinado murino de pulmón se cortó a intervalos regulares de cada tanda (inicio, medio y final).

### **3.3. Extracción de ARN total**

La extracción de ARN total se realizó utilizando el kit Recover All Total Nucleic Acid (AM1975, Ambion, EUA) de acuerdo al protocolo recomendado por los fabricantes, con modificaciones menores:

1. Limpiar cuidadosamente el área de trabajo con RNaseZAP, después limpiar 2 veces con H<sub>2</sub>O destilada, usar pipetas específicas para ARN y limpiarlas antes de su uso con EtOH 70%, cambiar con frecuencia los guantes durante el protocolo.
2. Añadir 1 ml de xileno 100% a las muestras, mezclar con vórtex brevemente.
3. Golpe de centrifuga.
4. Incubar a 50°C durante 3' en baño térmico.
5. Centrifugar a 16 500 g por 2' y desechar el sobrenadante.
6. Añadir 1 ml de EtOH 100%, mezclar con vórtex.
7. Centrifugar a 16 500 g por 2' y descartar el sobrenadante.
8. Repetir los pasos 6 y 7.
9. Incubar los tubos abiertos 45' a TA.

10. Añadir 100 µL de tampón de digestión y 4-20 µL Proteasa, mezclar pipeteando.
11. Incubar a 50°C en baño térmico hasta la lisis completa.
12. Incubar las muestras a 80°C durante 15' en bloque térmico.
13. Golpe de centrifuga.
14. Añadir 120 µl de Isolation additive y 275 µl de EtOH 100%, mezclar pipeteando.
15. Echar 700 µl de la muestra en la columna de separación.
16. Centrifugar a 10 000 g por 30'' para que el líquido atravesase la columna.
17. Desechar el líquido filtrado y conservar la columna.
18. Añadir 700 µL de solución Wash 1 a la columna.
19. Centrifugar a 10 000 g por 30'' para que el líquido atravesase la columna.
20. Desechar el líquido y conservar la columna.
21. Añadir 500µL de solución Wash 2/3 a la columna.
22. Centrifugar a 10 000 g por 30'' y descartar el líquido.
23. Centrifugar nuevamente a 10 000 g por 30''.
- 26- Preparar una *master mix* de DNAsa (6 µl de tampón DNAsa 10X, 4 µl de DNAsa y 50 µl de H<sub>2</sub>O destilada libre de nucleasas).
27. Añadir 60 µL de la *master mix* de DNAsa al centro de la columna.
28. Incubar 30' a TA.
29. Añadir 700 µL de solución Wash 1 a la columna.
30. Incubar 1' a TA.
31. Centrifugar a 10 000 g por 30'' y desechar el líquido.
32. Añadir 500 µL de solución Wash 2/3 a la columna.
33. Centrifugar a 10 000 g por 30'' y descartar el líquido.
34. Repetir los pasos 32 y 33.
35. Centrifugar a 10 000 g por 1' para eliminar el líquido residual.
36. Transferir la columna a un tubo nuevo.
37. Añadir 15-30 µl de Elution Solution en el centro de la columna e incubar 1' a TA.
38. Centrifugar a 16 500 g por 1'.
39. Colectar el ARN y mantener en hielo hasta guardar a -80°C.

#### 3.4. Medición de cantidad y calidad del ARN

La concentración del ARN total se estimó mediante espectrofotometría usando el equipo Nanodrop. Adicionalmente, se emplearon métodos de cuantificación por fluorescencia más precisos según el protocolo de los fabricantes: el Qubit® RNA BR Assay Kit (Q10210, Life Technologies,) para el ARN y el Qubit® dsDNA BR Assay Kit (Q32850, Life Technologies, EUA) para estimar el ADN genómico residual en la muestra. La calidad del ARN extraído fue evaluada a partir de la estimación del número de integridad del ARN (RIN, del inglés *RNA Integrity Number*), obtenido de analizar las muestras en el Agilent 2100 Bioanalyzer (EUA), mediante electroforesis en chips microfluídicos utilizando el Agilent RNA 6000 Nano Kit (5067-1511, Agilent, EUA) según el protocolo de los fabricantes.

### 3.5. Prueba piloto de comparación de microarrays Human Gene 2.0 ST vs Almac Xcel

El procesamiento de las muestras e hibridación del ADNc se realizó en el servicio de microarrays del Instituto Hospital del Mar de Investigaciones Médicas (IMIM) de Barcelona. El ARN procedente de tejido parafinado se encontraba degradado por la acción del formaldehído y presentaba valores de RIN entre 1-2.5, bajos en contraste con la calidad del ARN de muestras no parafinadas. Teniendo en cuenta que la mayoría de las muestras presentaron un RIN aproximado de 2, se escogieron solamente muestras con  $RIN \geq 2$  y se excluyeron aquellas con  $RIN < 2$ .

Se realizó una prueba piloto con el objetivo de comparar el microarray Human Gene 2.0 ST (Affymetrix, EUA) respecto al Almac Xcel (Affymetrix, EUA), el cual es específico de muestras parafinadas. El Human Gene 2.0 ST es un array en el que las sondas hibridan a lo largo de todo el transcrito interrogando 40 716 de los mismos, mediante 21 sondas/gen. En el Almac-Xcel las sondas hibridan en el extremo 3' del ARNm, interrogando alrededor de 97 000 transcritos y es equivalente en un 70% al Human Genome U133 Plus 2.0 array, el más usado de Affymetrix<sup>176</sup>. Pevio a la hibridación en el chip se amplificó el ADNc usando el WT-Ovation™ FFPE System V2 (Nugen, EUA), específico para ARN degradado de muestras parafinadas. En esta prueba piloto se emplearon las muestras descritas en la tabla 10 y posteriormente se procedió a hibridar muestras clínicas (Tabla 11), con el objetivo de conocer procesos fisiológicos causales involucrados en la progresión desde MI a CG y que caracterizan a la MI.

### 3.6. Análisis de calidad y estadísticos

Los análisis de calidad consistieron en la exploración de imágenes y la evaluación de la consistencia de los datos. Respecto a los análisis estadísticos, se normalizaron los datos de intensidad de fluorescencia aplicando el método *Robust Multi-array Average*<sup>177</sup> implementado en el programa Affymetrix® Expression Console™ para corregir el ruido de fondo, aplicar la normalización cuantil y expresar todas las sondas de un transcrito en base logarítmica 2. A continuación se corrigió el efecto *batch* presente en los diferentes experimentos, ya que las muestras fueron hibridadas en fechas diferentes, mediante el método *Combat*<sup>178</sup>. Una vez corregido este efecto, de las 81 804 sondas iniciales, se seleccionaron aquellas con desviación estándar (DE)  $< 0,34$  (30% de la DE global de todas las sondas), obteniendo un total de 57 263 sondas a emplear en los análisis. Se calculó el log *Fold Change* como la media de la intensidad de las muestras del primer grupo dividida entre la media de la intensidad de las muestras del segundo grupo a comparar y se aplicó la fórmula  $2^{\log \text{Fold Change}}$  para transformar los valores a *Fold Change*. Con el objetivo de comparar la distribución de las muestras de los distintos grupos (MI-CG, MI-No CG,

MII-CG, MII-No CG, MIC-CG, MIC-No CG y Sanos) se construyeron dendrogramas usando la distancia euclídeana y el método *Complete*.

Los genes diferencialmente expresados entre los grupos MII-CG vs MII-No CG, MIC-CG vs MIC-No CG, MI-No CG vs Sanos se obtuvieron aplicando la prueba t moderada, específica para datos de microarrays (*Linear Models for Microarray Data*, Limma), que tiene en cuenta la media, desviación estándar y N de los grupos a comparar<sup>179</sup>, y se ajustó para las comparaciones múltiples por la tasa de falsos positivos (FDR)<sup>175</sup>. Se consideraron las diferencias estadísticamente significativas si  $p < 0.05$  y el *Fold Change*  $\geq 2$  (genes sobre-expresados) ó  $\leq 0.5$  (genes sub-expresados). Los análisis se realizaron en el programa estadístico R (v3.1.1)<sup>180</sup> con sus paquetes estándar y, adicionalmente, los paquetes Biobase y Limma de Bioconductor<sup>181</sup>. Se comprobó si los genes diferencialmente expresados en las comparaciones MII-CG vs MII-No CG y MIC-CG vs MIC-No CG también estaban diferencialmente expresados en MI-No CG vs Sanos.

### 3.7. Análisis de enriquecimiento funcional por IPA y GSEA

El análisis individual de genes diferencialmente expresados solo permite dilucidar parcialmente mecanismos funcionales implicados en procesos biológicos, razón por la que es mucho más informativo realizar análisis de vías de señalización celular y otros procesos moleculares enriquecidos en los grupos en estudio.

Con este objetivo se utilizó el programa bioinformático *Ingenuity Pathway Analysis* (IPA, [www.ingenuity.com](http://www.ingenuity.com)) para anotar funciones biológicas e identificar vías canónicas de señalización celular a partir de los genes diferencialmente expresados en el microarray de expresión mediante la comparación de los mismos con la base de datos Ingenuity® Knowledge Base<sup>182</sup>. La misma consta de >2.44 millones de artículos científicos revisados por expertos y mediante procesamiento de lenguaje natural, integrando información de múltiples fuentes sobre alrededor de 16 900 genes humanos y la interacción de los mismos a nivel génico y proteico. A las vías canónicas identificadas les es asignado un p-valor y se grafican sus componentes, así como el nivel de sub o sobre-expresión de los mismos en los datos. Además, construye redes moleculares no redundantes y optimizadas respecto a la interconectividad y la maximización de su número de componentes, las cuales permiten conocer cómo se relacionan los genes diferencialmente expresados. A dichas redes se les asigna un p-valor y un *score*, el cual indica la probabilidad de que los mismos pertenezcan a dicha red en comparación con un resultado azaroso<sup>182</sup>.

Usando dicho programa se cargaron los genes diferencialmente expresados de cada comparación (MII-GC vs MII-No CG, MIC-CG vs MIC-No CG y MI-No CG vs Sanos) junto con su *Fold Change* y mediante la opción *Core Analysis* se obtuvieron las vías canónicas significativas, y las funciones moleculares y celulares sobre-representadas



en cada grupo. Por otra parte se construyeron redes moleculares de alta puntuación (score>15) que incluyen genes diferencialmente expresados interconectados con otras moléculas. Un score  $\geq 3$  significa que es significativa (IC 99.9%) la probabilidad de que dichos genes interactúen entre sí.

El programa bioinformático *Gene Set Enrichment Analysis* (GSEA)<sup>183</sup> identifica si los transcritos de un experimento de microarrays están significativamente enriquecidos en vías de señalización celular u otros procesos biológicos mediante la comparación con conjuntos génicos diferencialmente expresados y representativos de los anteriores, previamente depositados en la base de datos *Molecular Signatures Database* (MSigDB). Para determinar los conjuntos génicos significativamente enriquecidos calcula el *Enrichment Score* (ES) definido como la máxima desviación respecto a 0 encontrada entre los conjuntos génicos de bases de datos y los resultados del experimento, y su p-valor nominal asociado no ajustado, así como ajustado para comparaciones múltiples mediante FDR.

Los valores de expresión normalizados de las comparaciones MII-CG vs MII-No CG, MIC-CG vs MIC-No CG, y MI-No CG vs Sanos fueron cargados en el programa GSEA y buscados conjuntos génicos diferencialmente expresados en nuestros resultados usando los catálogos c2.all.v5.0 y c3.tft.v5.0. El catálogo c2.all.v5.0 está compuesto por vías de señalización celular y perturbaciones químicas y genéticas procedente de la base de datos MSigDB v5.0. Por su parte c3.tft.v5.0 se encuentra compuesto por conjuntos génicos que contienen genes que comparten un sitio de unión a un factor de transcripción definido en TRANSFAC v7.4. Solamente aquellos conjuntos génicos compuestos por 15 o más genes fueron analizados. Los conjuntos génicos de perturbaciones químicas y genéticas poseen información de genes inducidos o reprimidos debido a un efecto biológico y la mayoría han sido extraídos de publicaciones científicas.

Se emplearon los parámetros por defecto del programa, con excepción de la comparación MII-CG vs MII-No CG en la que la opción de permutación aplicada fue *Geneset*, debido a que el tamaño de la muestra en esta comparación es  $< 7$ . Se consideró un conjunto génico significativo si su p-valor nominal fue  $< 0.05$  y su q-valor de FDR  $< 0.25$  para las comparaciones MIC-CG vs MIC-No CG y MI-No CG vs Sanos<sup>184</sup>. En la comparación MII-CG vs MII-No CG se consideró un resultado significativo si su p-valor nominal  $< 0.01$  y su q-valor de FDR  $< 0.05$ . Se usó este criterio ya que el tipo de permutación *Geneset* realizada, en lugar de fenotipo, es un criterio menos estricto de significación por lo que se debe aplicar un valor de corte más riguroso de q-valor de FDR<sup>184 185</sup>. En ambos casos se seleccionaron vías con plausibilidad biológica demostrada en la metaplasia intestinal, la carcinogénesis gástrica o la carcinogénesis en general.

En el análisis del GSEA no todos los componentes del conjunto génico contribuyen de igual forma a la señal de enriquecimiento sino que existe un subconjunto que lo hace de manera más significativa, denominados *leading edge genes*<sup>185</sup>, siendo el núcleo que contribuye a la señal de enriquecimiento ES. Teniendo esto en cuenta, para las comparaciones con conjuntos génicos significativos MII-CG vs MII-No y MI-No CG vs Sanos se determinaron los genes *leading edge* de los conjuntos génicos significativos y se compararon con los genes sobre-expresados de dichas comparaciones.

Adicionalmente, un conjunto de datos humanos de expresión génica de metaplasia intestinal y mucosa gástrica sana<sup>186</sup> con número de identificador GSE47797 se descargaron de la base de datos Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/info/linking.html>), y analizadas vías y mecanismos funcionales usando el programa GSEA. Se siguió idéntica metodología a la descrita previamente escogiendo el tipo de permutación *Geneset* debido a  $N < 7$ . Los datos están constituidos por valores de expresión normalizados de 34 127 sondas obtenidas con el microarray Whole Human Genome Microarray 4x44K v2 (Affymetrix, EUA).

Además, usando el sitio web <http://geneontology.org/> se realizó un análisis de enriquecimiento de los genes diferencialmente expresados de las comparaciones estudiadas (MII-CG vs MII-No CG, MIC-CG vs MIC-No CG y MI-No CG vs Sanos) en los términos de la ontología génica (GO, del inglés *Gene Ontology*) Procesos Biológicos, Funciones Moleculares y Componentes Celulares, empleando solamente información de evidencias experimentales. Adicionalmente, se realizaron los mismos análisis anteriores pero en conjuntos de datos provenientes de Panther GO-Slim empleando el PANTHER Overrepresentation Test *release* 20150430 (<http://pantherdb.org/tools/compareToRefList.jsp>).

### 3.8. Validación de expresión génica por qRT-PCR

Algunos de los genes diferencialmente expresados en el microarray se validaron mediante qRT-PCR usando arrays dinámicos de expresión 96.96 Biomark HD (Fluidigm, EUA) y sondas UPL (Roche, Suiza). Para esto se seleccionaron genes sobre y sub-expresados procedentes principalmente de las comparaciones MII-CG vs MII-No CG y MII-No CG vs Sanos, y en menor medida de MIC-CG vs MIC-No CG, teniendo en cuenta la novedad de los hallazgos y/o funciones relevantes en la metaplasia intestinal, fisiología digestiva y/o carcinogénesis gástrica. Los genes de referencia utilizados fueron *ACTB*, *GAPDH*, *G6PD*, *RPL29* y *B2M*, seleccionados a partir de reportes previos<sup>187</sup> y el hecho de que no son significativos en el microarray de expresión.

Las sondas UPL y los cebadores específicos fueron seleccionados mediante el programa informático Universal Probe Library Assay Design Center (<http://lifescience.roche.com>)<sup>188</sup>, seleccionando *Homo sapiens* e introduciéndole el (los) identificador(es) de ARNm (Ej NM\_002123.4) de los genes de interés obtenidos del sitio web <http://www.ncbi.nlm.nih.gov/gene>, en el caso de que no aparecieran identificador(es) de ARNm se buscaron en Ensembl. Se seleccionó automáticamente un ensayo que abarcara el intrón para no detectar expresión génica debido a contaminación con ADN genómico en el ARN total. Siempre que hubiera varios conjuntos de sondas y cebadores para los transcritos de un mismo gen se seleccionó un ensayo común para evaluarlos al unísono mediante una sonda y un par de cebadores únicos. Si existían amplicones muy grandes (HLA-C[135pb], HLA-DQA1[107pb]) se diseñaron nuevos cebadores, más internos, haciendo uso de la secuencia de la sonda UPL en el programa informático Primer3 (<http://primer3.ut.ee/>). En la tabla 12 aparece la información correspondiente a los genes desregulados y de referencia seleccionados, transcritos interrogados, sondas UPLs y secuencia de los cebadores utilizados para la validación por qRT-PCR.

Las muestras de ARN usadas para la validación provienen de mucosa gástrica parafinada de MII-CG, MII-No CG, MIC-CG, MIC-No CG y Sanos. Una parte de las mismas había sido usada en el microarray de expresión mientras que otras constituyen una serie independiente (Tabla 13). En estas últimas la extensión de la MIC y la MII en los cortes inicial y final no siempre es mayor del 50%; sin embargo, las muestras fueron seleccionadas de manera que se garantizara la mayor extensión posible con vistas a que el ARN tuviera un componente mayoritario de tejido metaplásico.

**Tabla 12. Genes significativos del microarray de las comparaciones MII-CG vs MII-No CG, MIC-CG vs MIC-No CG y MI-No CG vs Sanos a ser validados por qRT-PCR.**

Gen	Transcrito(s)	Comparación	Sonda UPL	Tamaño amplicon (pb)	Secuencia de los cebadores
ACTB	NM_001101.3	Genes de referencia	64	97	ccaaccgcgagaagatga y ccagagggtacagggatag
B2M	NM_004048.2	Genes de referencia	42	86	ttctggcctggaggctatc y tcaggaaattgactttcattc
G6PD	NM_000402.4, NM_001042351.2	Genes de referencia	22	91	ctgcagatcgtgtctcgtg y tgcattttcaacaccttgacc
GAPDH	NM_002046.3	Genes de referencia	24	70	cctcctctaagatgggtctcg y gaccgatcgctccaaatc
RPL29	NM_000992.2	Genes de referencia	53	81	caggctcccaaacgtacc y gcaccagtcctctctgtcctc
AFAP1-AS1	NR_026892.1	MI-CG vs MII-No CG	14	60	tcaccgatttcaaaagcttcc y ttcaggcgtgagtgattgt
ATP5A1	NM_001001937.1, NM_004046.4	MI-CG vs MII-No CG	61	93	tgctattggtcataaagagatcca y gtagccgacaccacaatgg
ATP6V0E1	NM_003945.3	MI-CG vs MII-No CG	45	112	cctcactgfcctctcattg y caccaacatggtaataataactcc
GNAS	NM_000516.4, NM_001077488.2, NM_001077489.2, NM_001077490.1, NM_016592.2, NM_080425.2, NM_080426.2	MI-CG vs MII-No CG	3	72	aaggacaacagcaggctaccg y ggtgcttttaccagattctcca
GSTA1	NM_145740.3	MI-CG vs MII-No CG	53	90	acggtagacagcgittaaaca y ccgtgcattgaagtagtgga
HLA-C	NM_001243042.1, NM_002117.5	MI-CG vs MII-No CG	32	116	accggactcactctcc y gcggtgcgaaatacctcat
HLA-DQA1	NM_002122.3	MI-CG vs MII-No CG	68	101	accaaggccatttgaat y aatcggggccagagaatagtg
HLA-DRB3	NM_022555.3	MI-CG vs MII-No CG, MIC-No CG vs MIC-CG	41	113	gggctgttcatctactctcagga y caaagctggggcagaagat
HLA-DRB4	NM_021983.4	MI-CG vs MII-No CG, MIC-No CG vs MIC-CG	41	98	gggacagggctgttctacta y ccttgaatgggtcatctgc
IGHG4	ENST00000390543	MI-CG vs MII-No CG	79	106	tcctccatcgcagaaaaacat y ggctgacctgttcttgggt
IK	NM_006083.3	MI-CG vs MII-No CG	11	103	agctgaccagatcctttca y tcagcctcaggagggtttctt
IL1R2	NM_173343.1, NM_004633.3	MI-CG vs MII-No CG, MIC-No CG vs MIC-CG	72	89	cacatagagagccctacc y ggcaacttcaatgtagtctcattatt
MGAM	NM_004668.2	MI-CG vs MII-No CG	3	93	ccaggtagatgggacataga y gcagtgacatttctggcattac
MIME	NM_000902.3, NM_007287.2, NM_007288.2, NM_007289.2	MI-CG vs MII- No CG	67	78	ggggaggcttattatgggaag y tctcggatctgtgcaatcaa
PGC	NM_001166424.1, NM_002630.3	MI-CG vs MII- No CG	21	61	cttgggggcccctctgt y gttgaagcgggagtgactg
PI3	NM_002638.3	MI-CG vs MII- No CG	34	95	tgatcgtgggtggttctt y acggcctttgacagtcttt
PPIA	NM_021130.3	MI-CG vs MII- No CG	48	97	atgctggraccaccaacacaaat y tctttcactttgccaacacc

Gen	Transcrito(s)	Comparación	Sonda UPL	Tamaño amplicon (pb)	Secuencia de los cebadores
<i>RHOA</i>	NM_001664.2	MI- CG vs MI- No CG	8	83	gggagtagccaagatgaag y gaccacaaaagcgccaatc
<i>TMEM25</i>	NM_001144034.1, NM_001144035.1, NM_001144036.1, NM_001144037.1, NM_001144038.1, NM_032780.3	MI- CG vs MI- No CG	79	109	cgctctgtcatctttaatgt y caccagggcaaacaggac
<i>ANPEP</i>	NM_001150.2	MI- No CG vs Sanos	18	75	catcatcagagatggcagac y tgctgaagagatcgttctgg
<i>APOB</i>	NM_000384.2	MI- No CG vs Sanos	55	78	gacgacttttctaaatggaactctac y
<i>CDH17</i>	NM_001144663.1,	MI- No CG vs Sanos	2	72	cgtcaccagtaaacgctatttga y gtgcagtaagggcccgata
<i>CDX1</i>	NM_001804.2	MI- No CG vs Sanos	70	76	agccctacgagatggatg y tgccttggccgagcttac
<i>CDX2</i>	NM_001265.3	MI- No CG vs Sanos	34	82	atcaccatccggaggaaaag y tgcggttctgaaaccagatt
<i>CLCA1</i>	NM_001285.3	MI- No CG vs Sanos	62	78	ccaatggaagaatacaagcagtaa y
<i>CPS1</i>	NM_001122633.2, NM_001122634.2, NM_001875.4	MI- No CG vs Sanos	8	64	caagtttgcagtggaatcg y actggtagccaatgggtgc
<i>DMBT1</i>	NM_004406.2, NM_007329.2,	MI- No CG vs Sanos	21	77	accacttacggcattgac y cgacacctgtcactccatt
<i>FABP1</i>	NM_001443.2	MI- No CG vs Sanos	78	90	tgatccaaaacgaattcacg y caccttccaactgaaccactg
<i>MUC12</i>	NM_001164462.1	MI- No CG vs Sanos	72	68	cctggaacccttagcaccag y gacagacgattgtttccat
<i>MUC17</i>	NM_001040105.1	MI- No CG vs Sanos	17	76	ggggtgaacatcacaagcta y
<i>MUC3A</i>	ENST00000319509, ENST00000422757, ENST00000414964	MI- No CG vs Sanos	18	102	gtggagatcctgtccctgag y cactgtctatactcgtctc
<i>OLFM4</i>	NM_006418.4	MIC- CG vs MIC- NO CG, MI- No CG vs Sanos	24	74	atcaaaacaccctctgctc y gctgatgttcaccacaccac
<i>PRDM10</i>	NM_020228.2, NM_199437.1, NM_199438.1, NM_199439.1	MI- No CG vs Sanos	36	88	gtgaggacacggatctgga y gcgttatgtcacctctcaca
<i>SI</i>	NM_001041.3	MI- No CG vs Sanos	86	65	ttttggcatccagattcgcac y atccaggcagccaagaatc
<i>SLC26A3</i>	NM_000111.2	MI- No CG vs Sanos	2	74	ccatcatctgtctgattgtc y agctgccaggacggactt
<i>TFF3</i>	NM_003226.3	MI- No CG vs Sanos	4	87	gctctgctttgactccag y tggaggtgcctcagaaggt
<i>CYP3A4</i>	NM_001202855.2, NM_017460.5	MI- No CG vs Sanos	2	96	gatggctctcatccagactt y agtccatgtgaatgggttcc
<i>SLFN5</i>	NM_144975.3	IHQ	84	86	agcaagcctgtgtcattc y accactctctgaaataactgga
<i>SLFN12</i>	NM_018042.3	IHQ	2	76	aacctggacactcttcactggtt y atgcagttccaagcagaaa
<i>SLFN12L</i>	NM_001195790.1	IHQ	67	96	ttgaccgagaaggaatggat y ctggagacaggtgatgatgta

Tabla 13. Serie independiente de muestras usadas en la validación de expresión por qRT-PCR.								
Código	Tipo de MI al reclutamiento <sup>a</sup>	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Estatus <sup>c</sup>	Infección por <i>H. pylori</i>	Sexo <sup>d</sup>	Edad <sup>e</sup>	Procedencia <sup>f</sup>
A49	I-G3 F-G3	I-E2 F-E2	ND	MII-CG	ND	Hombre	60	La Princesa
A50	I-G3 F-G3	I-E2 F-E2	ND	MII-CG	Negativo	Mujer	71	La Princesa
A51	I-G3 F-G3	I-E2 F-E2	Antro	MII-CG	Positivo	Hombre	75	Soria
A52	I-G1 F-G1	I-E2 F-E2	Antro	MIC-No	ND	Hombre	61	FIS Multicéntrico
A53	I-G3 F-G4	I-E1 F-E4	Antro	MII-CG	Negativo	Mujer	80	Soria
A54	I-G1 F-G1	I-E2 F-E1	Antro	MIC-CG	Positivo	Hombre	49	FIS Multicéntrico
A55	I-G1 F-G1	I-E1 F-E1	Cuerpo	MIC-No	Positivo	Mujer	47	FIS Multicéntrico
A56	I-G2 F-G2	I-E4 F-E4	Antro	MIC-No	Positivo	Mujer	68	FIS Multicéntrico
A57	I-G2 F-G2	I-E2 F-E2	Antro	MIC-No	Negativo	Hombre	69	FIS Multicéntrico
A58	I-G2 F-G2	I-E2 F-E2	Antro	MIC-No	ND	Hombre	57	FIS Multicéntrico
A59	I-G2 F-G2	I-E3 F-E3	Antro	MIC-No	Positivo	Mujer	54	FIS Multicéntrico
A60	I-G1 F-G1	I-E2 F-E2	Cuerpo	MIC-No	Negativo	Mujer	54	FIS Multicéntrico
A61	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	Negativo	Hombre	61	FIS Multicéntrico
A62	I-G2 F-G2	I-E4 F-E3	Antro	MIC-No	Negativo	Mujer	49	FIS Multicéntrico
A63	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	Positivo	Hombre	49	FIS Multicéntrico
A65	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	Positivo	Mujer	46	FIS Multicéntrico
A66	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	Positivo	Hombre	35	FIS Multicéntrico
A67	I-G1 F-G1	I-E2 F-E2	Antro	MIC-No	Negativo	Hombre	55	FIS Multicéntrico
A68	I-G1 I-G1	I-E2 I-E1	Antro	MIC-No	Positivo	Mujer	44	FIS Multicéntrico
A69	I-G1 F-G1	I-E1 F-E1	Cuerpo	MIC-No	Negativo	Hombre	53	FIS Multicéntrico
A70	I-G1 F-G1	I-E2 F-E1	Antro	MIC-No	Negativo	Hombre	44	FIS Multicéntrico
A72	I-G1 F-G3	I-E1 F-E3	Cuerpo	MII-No	Negativo	Mujer	52	FIS Multicéntrico
A74	I-G1 F-G1	I-E1 F-E1	Cuerpo	MIC-CG	Positivo	Hombre	82	Soria
A75	I-G1 F-G1	I-E3 F-E2	Incisura	MIC-CG	Negativo	Mujer	75	Soria
A76	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	Negativo	Hombre	67	Soria
A77	I-G1 F-G1	I-E2 F-E2	Cuerpo	MIC-CG	Negativo	Hombre	49	FIS Multicéntrico

Código	Tipo de MI al reclutamiento <sup>a</sup>	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Estatus <sup>c</sup>	Infección por <i>H. pylori</i>	Sexo <sup>d</sup>	Edad <sup>e</sup>	Procedencia <sup>f</sup>
A78	I-G1 F-G1	I-E2 F-E2	Antro	MIC-No	Positivo	Mujer	64	Soria
A79	I-G1 F-G1	I-E1 F-E2	Antro	MIC-No	ND	Mujer	53	FIS Multicéntrico
A80	I-G1 F-G2	I-E3 F-E3	Cuerpo	MIC-No	ND	Mujer	49	FIS Multicéntrico
A81	I-G1 F-G1	I-E3 F-E3	Cuerpo	MIC-No	ND	Hombre	55	FIS Multicéntrico
A82	I-G3 F-G3	I-E3 F-E2	Antro	MII-CG	Negativo	Mujer	81	La Princesa
A83	I-G1 F-G1	I-E1 F-E1	Antro	MIC-CG	Negativo	Hombre	60	La Princesa
A84	I-G1 F-G1	I-E1 F-E1	Antro	MIC-CG	Negativo	Mujer	53	La Princesa
A85	I-G1 F-G1	I-E1 F-E1	Cuerpo	MIC-CG	Positivo	Hombre	63	FIS Multicéntrico
A86	I-G1 F-G1	I-E1 F-E1	Cuerpo	MIC-CG	Positivo	Hombre	63	FIS Multicéntrico
A87	I-G1 F-G1	I-E3 F-E3	Cuerpo	MIC-CG	Negativo	Hombre	52	FIS Multicéntrico
A88	I-G1 F-G1	I-E1 F-E1	Incisura	MIC-CG	Positivo	Hombre	54	FIS Multicéntrico
A89	I-G1 F-G1	I-E3 F-E3	Antro	MIC-CG	Negativo	Mujer	69	FIS Multicéntrico
A90	I-G1 F-G1	I-E1 F-E1	Antro	MIC-CG	Negativo	Mujer	58	FIS Multicéntrico
A91	I-G1 F-G1	I-E3 F-E3	Antro	MIC-CG	Negativo	Hombre	68	FIS Multicéntrico
A92	I-G1 F-G1	I-E2 F-E2	Antro	MIC-No	Positivo	Hombre	60	Soria
A93	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	Positivo	Mujer	41	Soria
A94	I-G1 F-G1	I-E1 F-E1	Antro	MIC-No	ND	Mujer	51	Soria
A95	I-G0 F-G0	I-E0 F-E0	Incisura	Sanos	Negativo	Hombre	34	Soria
A96	I-G1 F-G1	I-E2 F-E2	Incisura	MIC-CG	Negativo	Mujer	51	FIS Multicéntrico
A97	I-G1 F-G1	I-E1 F-E1	Antro	MIC-CG	Positivo	Hombre	72	Soria
A98	I-G1 F-G1	I-E2 F-E2	Antro	MIC-CG	Positivo	Hombre	81	FIS Multicéntrico
A99	I-G1 F-G1	I-E4 F-E4	Transicional					Marañón

<sup>a</sup> Tipos histológicos de MI (G1-G4) descritos en apartado 3.1. <sup>b</sup> E3>50%, E4>75%: Extensión de la MI en la muestra. <sup>c</sup> MII-CG, MIC-CG: Muestra de MII o de MIC al reclutamiento que progresa a CG al final del seguimiento. MII o MIC-No CG son controles de MII y MIC que no progresan a CG desde el reclutamiento hasta el final de seguimiento. <sup>d</sup> M: Mujer, H: Hombre, ND: No disponible. <sup>e</sup> Edad en el reclutamiento. <sup>f</sup> Hospital o proyecto de procedencia de las muestras. Multicéntrico hace referencia al proyecto de seguimiento de LPGs realizado en varios hospitales de España.

**3.8.1. Protocolo de qRT-PCR por sondas UPL y cebadores en plataforma Biomark**

Se trabajó en condiciones de esterilidad, se descongelaron las muestras y reactivos en hielo y se mezclaron suavemente. Todos los volúmenes se describen para placas de 96 pocillos.

1. Mezclar los cebadores 5' y 3' de cada gen a 20  $\mu\text{M}$  de c/u (mezcla de cebadores).
2. Preparar la mezcla para qRT-PCR en la placa.

Componente	Volumen ( $\mu\text{l}$ )
2X Assay Loading Reagent (Fluidigm,	12
Mezcla de cebadores 20 $\mu\text{M}$	9.6
Sonda UPL 10 $\mu\text{M}$	2.4

Cubrir la placa con adhesivo y poner en la nevera hasta su utilización.

3. Diluir el ARN en  $\text{H}_2\text{O}$  libre de RNAsas a concentración 70 ng/ $\mu\text{l}$  partiendo de las concentraciones de ARN medidas mediante Qubit.
4. Síntesis de ADN complementario (ADNc).

Componente	Volumen / reacción	Volumen 96 + 10% ( $\mu\text{l}$ )
Reverse Transcription Master Mix (PN 100-	1.0	105.6
$\text{H}_2\text{O}$	3.0	316.8
ARN (2,5 pg- 250 ng/ $\mu\text{l}$ )	1.0	
Total	5.0	422.4

5. Mezclar y dar un golpe de centrifuga a la placa.
6. Reacción en termociclador con el programa 25°C 5', 42°C 30', 85°C 5', 4°C  $\infty$ .
7. Usar inmediatamente o guardar a -20°C.
8. Preamplificación del ADNc (*Specific Target Amplification*, STA).

Mezcla de cebadores STA	
Componente	Volumen ( $\mu\text{l}$ )
Mezcla de cebadores (gen 1....N) 20 $\mu\text{M}$	2
$\text{H}_2\text{O}$	104
Total	200

Componente	Volumen / reacción ( $\mu\text{l}$ )	Volumen 96 + 10% ( $\mu\text{l}$ )
PreAmp Master Mix (Fluidigm, 1005876B1)	1.0	105.6
Mezcla de cebadores STA	1.25	132.0
$\text{H}_2\text{O}$	1.5	158.4
ADNc	1.25	
Total	5.00	



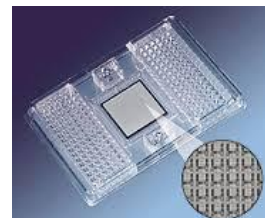
9. En la placa añadir 3.75 µl de mezcla de cebadores y 1.25 µl de ADNc específico, mezclar y dar un golpe de centrifuga.
10. Reacción en termociclador con programa 95°C 2', 16 ciclos 95°C 2', 60°C 4', 4°C.
11. Diluir las muestras 1:5 con tampón TE 1X para un volumen final de 25 µl.
12. Unir la pre-mezcla de muestras y ADNc preamplificado (STA).

Componente	Volumen / reacción (µl)	Volumen 96 + 10% (µl)
FastStart Universal Probe Master ROX (Roche 04913949001 2,5 ml)	4	384
20X GE Sample Loading Reagent (Fluidigm 85000746)	0.4	38.4
ADNc preamplificado (STA)	3.6	
Total	7	

Pre-mezcla

13. Mezclar en placa 4.4 µl de pre-mezcla con 3.6 µl de ADNc preamplificado (STA).

Preparación del chip GE 96x96 (PN 68000130 D1, Fluidigm, EUA)



14. Encender Fluidigm HX (96x96).
15. Comprobar las válvulas del chip con la jeringa e inyectar el contenido manteniendo el chip en un ángulo de 45°.
16. Colocar el chip en IFC Controller HX y correr el programa Prime (138X).
17. Encender el equipo de PCR en tiempo real Biomark y la lámpara.
18. Adicionar 5 µl de “Pre-mezcla + ADNc preamplificado (STA)” en los pocillos de la izquierda del chip. Adicionar 5 µl de “Mezcla para qRT-PCR” a la derecha del chip.
19. Correr el programa Load Mix (138x) en el IFC Controller HX.
20. Introducir el chip en el equipo Biomark, abrir el programa Data Collection y seleccionar parámetros Gene Expression, ROX, FAM-TAMRA, Auto Exposure y seleccionar programa GE 96x96 UPL v1.pcl.
21. Salvar el archivo de resultados Chip Run.bml.

**3.8.2. Procesamiento de los resultados de qRT-PCR por sondas UPL y cebadores en plataforma Biomark**

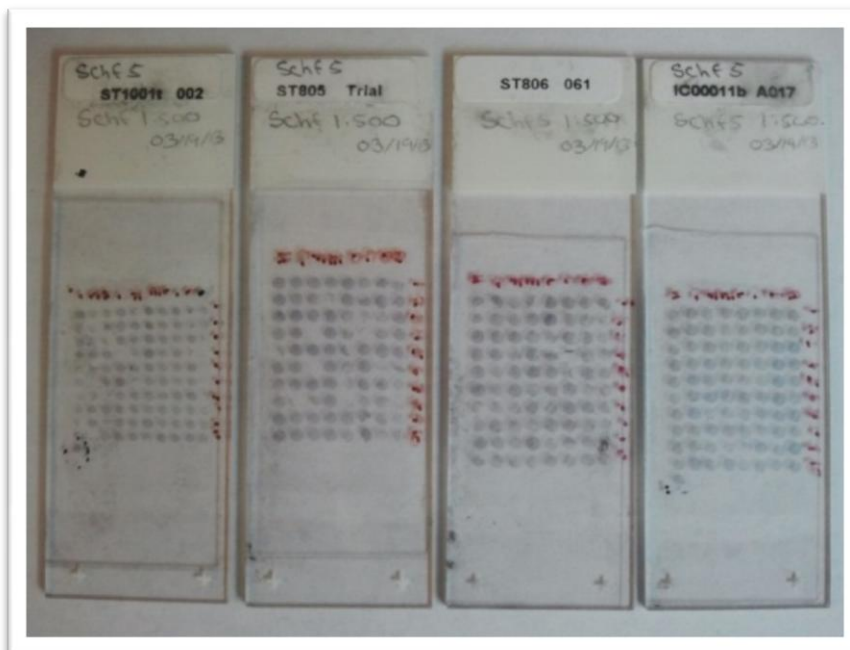
1. Abrir el archivo Chip Run.bml en el programa Fluidigm Real-Time PCR Analysis v 4.0.1.
2. Ir a Sample y Detector Setup para indicar al programa los nombres de las muestras y genes estudiados así como el patrón de pipeteo usado.

3. Establecer los valores umbrales: Quality Threshold (0.3), Baseline Correction (Lineal), Ct Threshold method (Autodetectors).
4. Pinchar el botón *Analyze* y exportar los resultados a Excel.
5. Seleccionar aquellos que pasan el control de calidad (Pass en columna Ct Call).
6. Chequear del mejor gen de referencia mediante análisis de varianza (ANOVA) de los valores duplicados de Ct de genes *ACTB*, *B2M*, *G6PD*, *GAPDH* y *RPL29* entre los grupos MI-CG (casos MI) vs MI-No CG (controles MI) vs Sanos. Cada gen fue evaluado independientemente y la corrección para comparaciones múltiple de Bonferroni fue aplicada posteriormente al ANOVA. Se usó el programa GraphPad Prism v 5.01 y un valor de  $\alpha < 0.05$  fue considerado estadísticamente significativo.
7. Cálculo del promedio de Ct para cada gen duplicado en una misma muestra.
8. Cálculo de  $dCt$  (MI-CG) =  $Ct\%(MI-CG) - Ct\%ACTB$  (MI-CG) y  $dCt$  (MI-No CG) =  $Ct\%(MI-No CG) - Ct\%ACTB$  (MI-MI).
9. Identificación de valores extremos (*outliers*) de Ct% y  $dCt$  independientemente para cada gen en los grupos MI-CG y MI-No CG. Para ello se empleó la prueba de Grubs del programa QuickCalc (<http://www.graphpad.com/quickcalcs/>) y un valor de  $\alpha < 0.05$  fue considerado estadísticamente significativo. En caso de existir un resultado significativo el *outlier* fue eliminado, aplicada nuevamente la prueba y eliminado nuevamente el *outlier*.
10. Cálculo de  $2^{-dCt}$  para cada gen en las muestras de cada grupo, MI-CG y MI-No CG.
11. Introducción de  $2^{-dCt}$  para cada gen de los grupos MI-CG y MI-No CG al programa Bootsratio (<http://regstattools.net/br>)<sup>189</sup>. El mismo determina la significación estadística de datos de expresión génica relativa a través de un test de permutación sin tener en cuenta el tipo de distribución estadística de los datos. De esta forma compara si la razón de la media de casos/controles (razón de medias [RM] o *Mean Ratio*) es significativamente diferente de 1. En caso de que la razón de medias sea distinta de la razón del *bootsratio* de la mediana el análisis puede ser no válido debido a una gran variabilidad, por lo que no se tuvieron en cuenta aquellos genes cuya razones entre los anteriores parámetros fueran  $> 1.5$  según recomendación personal del autor<sup>189</sup>. Los valores de RM, desviación estándar de RM, tamaño de muestra y p-valor fueron seleccionados para cada gen y un valor de  $\alpha < 0.05$  fue considerado significativo.

## 4. Estudio de expresión génica e inmuno-histoquímica de Schlafen5 en lesiones precursoras y cáncer gástrico

### 4.1. Tissue arrays comerciales de tejido gástrico

En una primera aproximación, se analizaron las muestras de cuatro *tissue arrays* comerciales de tejido gástrico (ST805, ST806, ST1001t, IC00011b, US Biomax Inc., <http://www.biomax.us/tissue-arrays>) que contienen mucosa gástrica normal (N=15), gastritis superficial (N=43), inflamación crónica (N=34), gastritis crónica atrófica (N=30), gastritis crónica atrófica ligera-moderada (N=27), úlcera (N=10), displasia (N=11), cáncer gástrico (N=106) y muestras no analizadas (tejido muscular, vasos sanguíneos, muestras dañadas, tejidos muy pequeños, sin muestra, N=66) (Figura 18). La mucosa gástrica normal es adyacente al CG en el 80% de las muestras. El 82% de las muestras de gastritis atrófica poseen metaplasia intestinal.



**Figura 18.** Microarrays de tejido usados para inmunohistoquímica de Schlafen5.

### 4.2. Pacientes y muestras analizadas

Para los estudios comparativos de la tinción de Schlafen5 en diferentes LPGs, las muestras de pacientes usadas pertenecen al proyecto de seguimiento de LPGs de Soria y los controles de mucosa gástrica sana descritos en apartado 3.1 de esta sección. Además de las mismas, fueron usadas muestras de cáncer gástrico (subtipos intestinal y difuso) obtenidas mediante gastroscopia en el Hospital Virgen del Mirón, Soria, con un mínimo de 50% del tejido tumoral libre de necrosis tumoral o ulceración. El tamaño de muestra, el tipo de lesión y la información clínica de los pacientes se presentan en la tabla 14.

**Tabla 14. Muestras de lesiones precursoras y cáncer gástrico usadas en el estudio inmuno-histoquímico de Schlafen5.**

Diagnóstico Histológico <sup>a</sup>	N <sup>b</sup>	Edad <sup>c</sup> (% ± DE)	Sexo (% ♂) <sup>d</sup>	Infección por <i>H. pylori</i> (% +) <sup>e</sup>
Normal	41	42.5 (18.8)	51.2	0
GNA	26	44.2 (12.3)	46.1	88.5
GCA	49	48.5 (12.6)	38.8	49.0
MIC	44	53.1 (10.5)	45.4	63.6
MIC-CG	6	62.0 (6.6)	50	50
MIC- No CG	38	51.7 (10.4)	44.7	44.7
MII	54	58.3 (12.8)	50	51.8
MII-CG	19	66.5 (13.6)	63.2	63.2
MII-No CG	35	53.8 (9.9)	42.9	42.9
MI	98	55.9 (12.1)	48	57.1
MI-CG	25	65.4 (12.4)	60	60
MI-No CG	73	52.7 (10.1)	43.8	43.8
CG	67	72.6 (11.0)	61.2	25.4

<sup>a</sup> Normal se refiere a mucosa gástrica sana, GNA: Gastritis no atrófica, GCA: Gastritis crónica atrófica multifocal, MIC/MII-CG: MIC/MII que progresa a CG, MIC/MII-No CG: MIC/MII que no progresa a CG, CG: Cáncer Gástrico. <sup>b</sup> Tamaño de muestra. <sup>c</sup> Edad en el diagnóstico al reclutamiento. <sup>d</sup> Porcentaje de hombres. <sup>e</sup> Porcentaje de infectados por *H. pylori*.

### 4.3. Histopatología

El tejido parafinado obtenido mediante gastroscopia fue cortado con micrótopo en laminillas de 4 micras de espesor que fueron teñidas mediante técnicas de hematoxilina-eosina, azul de Alcian, ácido peryódico de Schiff (PAS, pH 2.5) y Giemsa, para su diagnóstico histológico. Además, en algunas biopsias con metaplasia intestinal incompleta, se tiñeron secciones con hierro-diamina y azul de Alcian para detectar mucinas sulfatadas. El diagnóstico fue realizado según lo descrito en el apartado 3.1 de esta sección.

### 4.4. Infección por *Helicobacter pylori*

La infección por *Helicobacter pylori* se determinó por la tinción de Giemsa y mediante la revisión de las historias clínicas (prueba de ureasa en la biopsia y test de aliento con urea) y el resultado de la entrevista a los sujetos sobre la presencia anterior de *Helicobacter pylori*, el tratamiento y sus resultados. La presencia de al menos una infección positiva en cualquiera de estas fuentes de información se tomó como un resultado positivo.

### 4.5. Inducción de la expresión de Schlafen5 por IFN $\alpha$

Empleando modelos celulares Se exploró si la expresión de Schlafen5 es inducida por IFN $\alpha$  en monocitos y células T humanas. La línea celular mieloide HL-60

(leucemia promielocítica humana, ATCC®, CCL-240) fue cultivada a 37°C y 5% de CO<sub>2</sub> en medio Dulbecco modificado (IMDM, Life Technologies, EUA) suplementado con suero fetal bovino al 20% (Hyclone, EUA). La línea celular Jurkatt (leucemia aguda de células T, ATCC®, TIB-152) fue cultivada en medio RPMI-1640 (Gibco/Invitrogen, EUA) suplementado con suero fetal bovino al 10%. Ambas líneas celulares fueron tratadas con IFN $\alpha$  humano recombinante (IFN2 $\alpha$ , 1.000 UI/ml, # 11105-1, Interferon Source, EUA) durante 8h (HL-60) o 4h (Jurkat), se colectaron por centrifugación y el ARN total se extrajo usando Tryzol® (Life Technologies, EUA) según el protocolo de los fabricantes. Se emplearon 500 ng de ARN total para sintetizar ADNc y la expresión génica fue cuantificada en un termociclador Bio-Rad C1000 con el sistema de detección en tiempo real CFX96TM y el programa CFX Manager software v.1.5 (Hercules, CA) utilizando la sonda TaqMan® (Hs00288058\_m1) y el kit TaqMan Gene Expression Assay (Life Technologies, EUA). El nivel de expresión fue normalizado respecto al ARNr 18S (Hs 99999901\_s1). Se empleó como control negativo ARN total proveniente de las líneas celulares cultivadas pero no tratadas con IFN $\alpha$ . Los valores de expresión normalizados fueron comparados entre las muestras tratadas con IFN $\alpha$  (9 réplicas) respecto a las no tratadas (9 réplicas) usando el programa Bootsratio (<http://regstattools.net/br>)<sup>189</sup>.

#### 4.6. Silenciamiento de Schlafen5 y cytospin en células Jurkat

Con el objetivo de validar la especificidad del reconocimiento de Schlafen5 por una técnica alternativa, la línea celular Jurkat se cultivó en medio RPMI-1640 con suero fetal bovino al 10%, penicilina 100 U/ml, estreptomycin 100  $\mu$ g/ml en un ambiente de CO<sub>2</sub> 5% a 37 °C. Las células Jurkat se transfectaron con siRNA para Schlafen5 (hSlfn5, SC94178, EUA) o su siRNA control, utilizando el kit Lipofectamine RNAiMAX (Life Technologies, EUA) según las instrucciones del fabricante. Dos días después de la transfección, las células fueron tratadas con IFN2 $\alpha$  (800 U/ml) durante 24 horas, colectadas por centrifugación y depositadas en portaobjetos de vidrio (631-0108, VWR, EUA). La expresión de Schlafen5 se determinó mediante inmunohistoquímica, tal y como se expone en el apartado siguiente a partir del paso con el anticuerpo primario y obviando la contra tinción con hematoxilina.

#### 4.7. Detección inmunohistoquímica de Schlafen5

La tinción inmunohistoquímica de Schlafen5 se realizó en los *tissue arrays* comerciales de tejido gástrico (apartado 4.1 de esta sección) y en las muestras de LPGs de estudios de seguimiento descritas en los apartados 3.1 y 4.2 de esta sección (Tabla 14).

El tejido parafinado fue cortado en micrótopo a 4  $\mu$ m de espesor, depositado en portaobjetos de vidrio (631-0108, VWR, EUA) e incubado toda la noche a 60°C. Las secciones fueron desparafinadas en incubaciones sucesivas de xileno, EtOH

absoluto, EtOH 90%, EtOH 70%, e hidratadas en PBS 1X (NaCl 137 mM, KCl 2.7 mM, Na<sub>2</sub>HPO<sub>4</sub> 10 mM, KH<sub>2</sub>PO<sub>4</sub> 1.8 mM, pH 7.4). La recuperación del antígeno se realizó por ebullición en tampón citrato de sodio 10mM pH 6 durante 45 minutos. Después, las muestras se incubaron durante 20' a TA y se lavaron dos veces con tampón TBS-T 1X (50 mM Tris, 150 mM NaCl, Triton X-100 0.001%, pH 7.5). A continuación las secciones se sumergieron por 10' en H<sub>2</sub>O<sub>2</sub> para inactivar las peroxidasas endógenas y se realizaron tres lavados en TBS-T 1X. Los sitios de unión no específicos fueron bloqueados con solución Protein Block durante 10'.

El anticuerpo primario contra Schlafen5 humano (ab121537, Abcam, Reino Unido) se añadió en una dilución 1:500 en TBS-T 1X y se incubó durante 90', después de los cuales se realizaron dos lavados con TBS-T 1X. Se añadió el anticuerpo secundario adecuado biotinilado (ab64261, Abcam, Reino Unido), se incubó durante 10' y se realizaron dos lavados con TBS-T 1X. A continuación, las muestras se incubaron durante 10' en estreptavidina-peroxidasa y se hicieron dos lavados con TBS-T 1X. Finalmente, el producto de la reacción se visualizó por incubación durante 10' en una mezcla de 2 ml de H<sub>2</sub>O destilada, 500 µl de sustrato de diaminobenzidina (DAB) y 1 gota de cromógeno DAB. El desarrollo de color se detuvo incubando las secciones en H<sub>2</sub>O destilada durante 2 minutos y se realizó una contra tinción utilizando hematoxilina de Gill. Se emplearon como controles positivos tejido de nasofaringe humana (cavum), y colón e íleon afectados con enfermedad de Crohn; el control negativo se realizó reemplazando el anticuerpo primario por TBS-T 1X. Se validaron los resultados de la tinción de Schlafen5 en un subgrupo de muestras mediante otro anticuerpo que también reconoce a Schlafen5 (HPA 017760, Sigma-Aldrich, EUA)<sup>88</sup>.

#### **4.8. Interpretación de la tinción inmunohistoquímica**

La interpretación de la tinción se realizó por microscopía óptica con un aumento de 200X. Cuando la tinción de Schlafen5 mostró localización nuclear se consideró verdadero positivo y una tinción citoplasmática en glándulas gástricas se interpretó como un falso positivo debido a la localización nuclear de esta proteína<sup>190</sup>. Las muestras fueron cuantificadas visualmente teniendo en cuenta el % de células positivas (0-100%) con diferentes intensidades de la tinción (negativa (0), baja (1), media (2) y alta (3) para cada campo analizado. Cada campo fue cuantificado mediante el score  $H = \%Células \text{ con intensidad Baja} \times 1 + \%Células \text{ con intensidad Media} \times 2 + \%Células \text{ con intensidad Alta} \times 3$ <sup>191</sup> independientemente para el estroma (lámina propia) y las glándulas gástricas, con el objetivo de diferenciar el componente inflamatorio con respecto al epitelial. En el caso de que linfocitos intra-epiteliales estuvieran presentes se cuantificaron como parte del estroma. Un promedio de la intensidad de la señal (%) se calculó teniendo en cuenta todos los campos observados. Para evaluar la variabilidad inter-observador en la

interpretación inmuno-histoquímica, una submuestra aleatoria del 15% de los casos y controles fue examinada cualitativamente por un patólogo.

#### 4.9. Análisis estadísticos

El gráfico de dispersión de puntos *scatter dot-plot* mostrando la media  $\pm$  DE se empleó para mostrar la distribución de la cuantificación de Schlafen5 en el estroma gástrico. Para probar si la misma se ajusta a una distribución normal, se aplicó la prueba de normalidad ómnibus de D'Agostino y Pearson. Se realizó la prueba de Kruskal-Wallis para evaluar las diferencias en la cuantificación de la expresión de Schlafen5 entre los grupos GNA, GCA, MIC, MII y CG. Se aplicó la prueba U de Mann-Whitney para comparar la expresión de Schlafen5 en el grupo de MI que progresa a cáncer gástrico (MI-GC) con respecto a MI que no progresa a cáncer gástrico (MI-No CG), así como respecto a sus subtipos histológicos (MIC-CG vs MIC-No CG y MII-CG vs MII-No CG).

Se categorizó el score H de Schlafen5 estromal según  $10 > \text{scoreH} \geq 0$  (Negativo),  $100 \geq \text{scoreH} \geq 10$  (Débil Positivo),  $200 \geq \text{scoreH} \geq 100$  (Moderadamente Positivo) y  $300 \geq \text{scoreH} \geq 200$  (Fuerte Positivo)<sup>191</sup>. Para el grupo de metaplasia intestinal se realizaron dos modelos de regresión logística incluyendo o no la cuantificación de Schlafen5 estromal categorizado y siempre ajustando por sexo, edad y tipo de MI, considerando como variable dependiente la progresión a CG. Se calcularon OR con IC95% para la cuantificación del Schlafen5 estromal considerando el grupo Negativo como la categoría de referencia. Se construyeron curvas de sobrevivencia *Receiver Operating Characteristic* (ROC) para ambos modelos de regresión logística y se calculó el área bajo la curva (AUC, del inglés *area under curve*). Para probar la mejora en la predicción del modelo añadiendo la cuantificación del Schlafen5 estromal, se utilizó el test de la razón de verosimilitud *Log Likelihood*. El programa informático GraphPad v5.0 (GraphPad software, EUA, [www.graphpad.com](http://www.graphpad.com)) se utilizó para la gráfica y la prueba U de Mann-Whitney, mientras que el programa SAS OnlineDoc v9.1.3 (SAS Institute Inc., EUA) se utilizó para el resto de las pruebas estadísticas. La significación estadística se estableció en  $p < 0.05$ .

#### 4.10. Doble tinción inmunohistoquímica para Schlafen5 y CD2, CD20 o Mac2

Teniendo en cuenta que Schlafen5 se expresa en las células T murinas y humanas (Jurkat) y en la línea celular de células mieloides HL-60 pero no se sabe su localización celular precisa en la mucosa gástrica humana, se realizó doble tinción inmuno-histoquímica para Schlafen5 y CD20, CD2 y MAC2, marcadores de superficie de células B, T y mieloides, respectivamente.

El corte y la desparafinación se realizaron de la manera previamente descrita para Schlafen5. La recuperación de antígeno se llevó a cabo en tampón citrato de sodio

10 mM, pH 6 durante 15', se lavó con TBS 1X (50 mM Tris, 150 mM NaCl, pH 7.5) y los sitios de unión no específicos fueron bloqueados con suero de cabra al 20% disuelto en TBS-T 1X. El anticuerpo primario  $\alpha$ CD20 (ab9475, Abcam, Reino Unido) se incubó en dilución 1:50 a 4°C durante la noche. Después de lavar con TBS 1X, el anticuerpo secundario biotinilado obtenido en cabra (BA-9200, Vector Laboratories, EUA) fue incubado en una dilución 1:200 durante 30' y se lavó con TBS 1X. A continuación, las muestras fueron incubadas con reactivo ABC-AP (Vectastain, Vector Laboratories, EUA) durante 30' y lavadas con TBS 1X. El producto de la reacción se visualizó por incubación con solución de sustrato de fosfatasa alcalina roja durante 20' (Red AP Substrate Kit SK-5100, Vector Laboratories, EUA) disuelta en 100 mM Tris-HCl pH8.5. Posteriormente, las muestras fueron lavadas con agua del grifo y TBS-T 1x. Después de este paso se llevó a cabo el mismo protocolo de detección de Schlafen5 previamente descrito a partir del bloqueo de peroxidasas con H<sub>2</sub>O<sub>2</sub>. Se utilizó TBS 1X para todos los lavados finales antes del secado al aire.

Para la colocación con el marcador de células T, el anticuerpo CD2 (ab131276, Abcam, Reino Unido) se incubó a una dilución 1:100 en TBS 1X con 1% de BSA a 4 °C durante toda la noche. Las secciones fueron incubadas con el anticuerpo secundario biotinilado obtenido en cabra durante 10' y luego se lavaron con TBS 1X. A continuación se procedió a la detección de Schlafen5.

Para detectar células mieloides, el anticuerpo MAC2 (14-5301, eBioscience, USA) se incubó a una dilución 1:200 en TBS 1X con 1% de BSA durante 1 hora a TA. Las secciones fueron incubadas con el anticuerpo secundario biotinilado anti-rata (BA-9400, Vector Laboratories, EUA) a una dilución de 1:200 durante 30'. Posteriormente, se incubaron con el reactivo ABC-AP durante 30' y el desarrollo de color se realizó por incubación en oscuridad durante 30' usando el sustrato BCIP/NBT (SK-5400, Vector Laboratories, EUA) disuelto en Tris-HCl 100 mM con 0.1% de Tween pH 9. Las secciones fueron lavadas en H<sub>2</sub>O y deshidratadas, no se realizó contra-tinción con hematoxilina. A continuación se procedió a la detección de Schlafen5.





# **RESULTADOS**



## 1. Estudio de asociación de genes de señalización de *Helicobacter pylori* y riesgo de cáncer gástrico

### 1.1. Análisis estadísticos de asociación genética

Se emplearon en este estudio muestras de 365 casos de CG y 1 284 controles; sus principales variables analizadas se describen en la tabla 15. En ambos grupos los hombres representan alrededor del 59% de la muestra y las mujeres el 41%, siendo la edad media de 58.4 años para ambos, indicando una alta homogeneidad de dichos grupos respecto a estas variables. La infección por *H. pylori* es la única variable con diferencias estadísticamente significativas, estando el 80.5% de los casos infectados por *H. pylori* respecto al 60.7% de los controles.

Tabla 15. Principales características de la población analizada en el estudio de asociación genética con cáncer gástrico.			
Variables analizadas		Controles (N=1284) N (%)	Casos (N=365) N (%)
Sexo	Hombres	759 (59.1)	214 (58.6)
	Mujeres	525 (40.9)	151 (41.4)
Edad al reclutamiento (años)	Media $\pm$ DE <sup>b</sup>	58.4 $\pm$ 7.69	58.4 $\pm$ 7.93
País	Francia	3 (0.23)	2 (0.55)
	Italia	206 (16.0)	56 (15.3)
	España	134 (10.4)	41 (11.2)
	Inglaterra	135 (10.5)	41 (11.2)
	Países bajos	99 (7.71)	26 (7.12)
	Grecia	88 (6.85)	24 (6.58)
	Alemania	186 (14.5)	48 (13.2)
	Suecia	220 (17.1)	64 (17.5)
	Dinamarca	205 (16.0)	61 (16.7)
	Noruega	8 (0.62)	2 (0.55)
	Infección por <i>H. pylori</i>	Ausencia	492 (38.3)
Presencia		779 (60.7)	294 (80.5)
Desconocido		13 (1)	3 (0.9)
cagA en casos no cardias y controles infectados por <i>H. pylori</i> <sup>a</sup>	+	584 (74.97)	151 (92.6)
	-	195 (25.03)	12 (7.4)
Localización anatómica de CG	Cardias	----	107 (29.3)
	No cardias	-----	181 (49.6)
	Ambos		6 (1.64)
	Desconocido		71 (19.5)
Subtipo histológico de CG	Intestinal	----	126 (34.5)
	Difuso	-----	128 (35.1)
	Mixtos		8 (2.19)
	Desconocido		103 (28.2)

<sup>a</sup> El % de individuos cagA + o - se calculó con respecto a 163 casos de CG no cardias y 779 controles infectados por *H. pylori*. \* indica valor de p<0,0001 para la comparación entre los casos y controles mediante la prueba de chi cuadrado. <sup>b</sup> DE indica desviación estándar.

Los 30 tagSNPs analizados (6 en *CD14*, 5 en *NOD2*, 7 en *TLR4* y 12 en *NFKB1*) se encuentran en equilibrio de Hardy-Weinberg en los controles. Las características de los SNPs analizados y su asociación con el CG como fenotipo global (intestinal+difuso o cardias+no cardias) se describe en la tabla 16. Se encontró una asociación directa entre el SNP rs1583005 de *CD14* y asociaciones inversas entre los SNPs de *NOD2* (rs7202124, rs2111235) y *TLR4* (rs10491851) con el riesgo de CG global. Las asociaciones de *NOD2* rs7202124 (OR=0.74, IC95% 0.61-0.89, p=0.002) y rs2111235 (OR=0.77, IC95% 0.64-0.93, p=0.005) fueron las más significativas y tras corregir las comparaciones múltiples sólo la región *NOD2* (p=0.009) permaneció significativa (Tabla 16).

Cuando los casos se agruparon teniendo en cuenta la localización anatómica del CG (Tabla 17) los SNPs de *NOD2* rs7202124 y rs2111235 se asociaron con el CG no cardias, y son los más significativos del estudio. La prueba de heterogeneidad demostró diferencias para el SNP rs2111235 entre ambas localizaciones anatómicas. *NOD2* fue la única región genética que se mantuvo asociada con CG no cardias, ya que a pesar de la existencia de SNPs significativos de *NFKB1*, este gen no fue significativo después del test de permutaciones.

Cuatro SNPs de *CD14* se asociaron con el CG del cardias, y la región permaneció significativa después de corregir por comparaciones múltiples. El test de heterogeneidad confirmó una diferencia significativa entre las localizaciones anatómicas de CG cardias y no cardias para las variantes rs11167532 y rs2569190, asociadas negativamente con el CG tipo cardias.

Al estratificar los CG no cardias y los controles infectados por la presencia de anticuerpos anti-cagA los SNPs de *NOD2* más significativos del estudio rs7202124 y rs2111235 se mantuvieron asociados con el CG no cardias sólo en los individuos cagA+. Sin embargo, rs5743289 (*NOD2*) sólo fue significativo en el subgrupo cagA-, aunque el tamaño de la muestra de los casos es muy bajo para emitir alguna conclusión (Tabla 18).

Respecto a los subtipos histológicos de CG intestinal y difuso (Tabla 19) la variante de *CD14* rs778588 se asoció con el subtipo intestinal con tests de heterogeneidad y permutaciones significativos, confirmando que es diferente su efecto atendiendo al subtipo histológico. El SNP rs7202124 de *NOD2* tuvo una asociación negativa con ambos subtipos mientras que rs2111235 solamente con el subtipo intestinal, pero el test de heterogeneidad no fue significativo en ninguno de estos casos. Diversos SNPs de *NFKB1* se asociaron con el subtipo difuso, pero ninguno fue significativo después de aplicarse la corrección para comparaciones múltiples y el test de heterogeneidad.

El análisis haplotípico (Tabla 20) reveló asociaciones directas entre la combinación GCCCGT de *CD14* con el riesgo de los CG cardias e intestinal. Dicho haplotipo está formado por los alelos de riesgo de rs778588 (alelo C asociado positivamente con CG cardias e intestinal) y rs1583005 (alelo T asociado positivamente con CG cardias); así como, por los alelos comunes de rs11167532 y rs2569190, cuyos alelos menores fueron asociados inversamente con CG cardias. Otro resultado a destacar es la asociación inversa con el CG no cardias e intestinal del haplotipo de *NOD2* GTCGC. Este haplotipo contiene el alelo G de rs7202124 y el alelo T de rs2111235, los cuales en el análisis de asociación de cada SNP se hallaron inversamente asociados con el CG global y su subtipo no cardias. La variabilidad genética cubierta por los haplotipos de *NOD2* y *CD14* es del 81 y 38%, respectivamente.

El análisis haplotípico aporta nueva información de asociación en los casos de *TLR4* y *NFKB1* como factores de riesgo del CG de cardias y no cardias, respectivamente; ya que estos genes no se asociaron a nivel de SNP individual después de aplicar el test de comparaciones múltiples. La variabilidad genética que cubren dichos haplotipos es del 40 y 70%, respectivamente.

## 1.2. Análisis bioinformáticos de efectos funcionales

Al evaluar los posibles efectos funcionales de los SNPs asociados a CG encontramos que el SNP más significativo, rs2111235 de *NOD2*, así como rs7202124 y rs5743289, se localizan en transcritos afectados por el mecanismo de degradación del ARNm *Nonsense Mediated mRNA Decay* (NMD). El SNP más significativo rs2111235 se encuentra, además, en una región predicha como regulatoria por el programa Pupasuite v3.1 (Tabla 16).

En el caso de los SNPs de *CD14* asociados a CG, el rs11167532 se encuentra en región 5', el rs2569190 en una región conservada triple hélice UTR 5', y rs1583005 es transcrito sujeto a NMD del gen *IK*, un regulador negativo de la expresión de HLA-DR en células hematopoiéticas<sup>192</sup>. Los SNPs no asociados a CG también se encontraron en potenciales elementos funcionales como regiones regulatorias (rs4914, rs1329057) y de triple hélice (rs1585214). Al comparar la distribución de las localizaciones entre los SNPs asociados y no asociados a CG se observó un incremento significativo de SNPs localizados en transcritos sujetos a mecanismo NMD en los SNPs asociados (4) respecto a los no asociados (0) (test exacto de Fisher  $p=0.0026$ ). La búsqueda de potenciales eQTLs entre los SNPs asociados a CG identificó a todos los SNPs de *CD14* asociados a CG (rs11167532, rs778588, rs1583005, y rs2569190), así como a rs7202124 y rs5743289 de *NOD2* (Tabla 21).

Tabla 16. Características de los SNPs analizados y su asociación estadística con el CG global (cardias + no cardias/intestinal + difuso).									
Gen	SNP <sup>a</sup>	Localización <sup>b</sup>	Análisis bioinformáticos <sup>c</sup>	Genotipos <sup>d</sup>	Número de controles con cada genotipo <sup>e</sup>	Número de casos con cada genotipo <sup>f</sup>	OR (IC 95%) <sup>g</sup>	p-valor <sup>h</sup>	p-valor perm <sup>i</sup>
CD14 <sup>b</sup>	rs11167532	Región 5'	Región 5'	GG/GA/AA	382/582/198	112/158/54	0.97 (0.81-1.16)	0.720	0.188
	rs7711117	No codificante	RR, Conservado	CC/CT/TT	770/439/67	203/142/19	1.14 (0.94-1.38)	0.202	
	rs778588	Región 3'	Región 3'	TT/TC/CC	667/526/89	189/143/32	1.05 (0.87-1.26)	0.614	
	rs4914*	Sinónimo. L367L	RR, Cons	CC/CG/GG	957/261/13	80/19/1	0.91 (0.69-1.19)	0.479	
	rs2569190	UTR 5'	RR, Triple hélice, Cons	GG/GA/AA	307/621/214	51/79/42	0.92 (0.77-1.09)	0.341	
NOD2	rs1583005	IK. intrónico	Transcripto NMD	CC/CT/TT	244/249/90	53/71/27	<b>1.20 (1.02-1.42)</b>	<b>0.032</b>	
	rs7202124	SNX20, intrónico	NMD, Conservado	AA/AG/GG	313/225/45	97/48/5	<b>0.74 (0.61-0.89)</b>	<b>0.002</b>	<b>0.009</b>
	rs2111235	Intrónico	NMD, RR	CC/CT/TT	631/507/134	205/129/26	<b>0.77 (0.64-0.93)</b>	<b>0.005</b>	
	rs5743289*	Intrónico	NMD	CC/CT/TT	478/190/19	15/11/1	1.05 (0.85-1.31)	0.634	
	rs3135500	UTR 3'	Conservado	GG/GA/AA	472/598/214	42/66/20	1.13 (0.96-1.33)	0.150	
TLR4	rs3785142	CYLD, intrónico	Conservado	TT/TC/CC	343/614/317	39/56/33	0.92 (0.78-1.08)	0.324	
	rs1329061	Intergénico	Intergénico	TT/TC/CC	572/580/129	49/64/15	1.16 (0.98-1.39)	0.089	0.24
	rs1329060	Intergénico	Intergénico	GG/GA/AA	948/309/26	87/36/3	1.13 (0.90-1.42)	0.302	
	rs1329057*	Intergénico	RR, intergénico, Conservado	AA/AG/GG	825/410/48	62/41/4	1.13 (0.92-1.39)	0.234	
	rs10491851	Región 5'	Conservado	GG/GT/TT	678/516/87	217/126/21	<b>0.81 (0.67-0.99)</b>	<b>0.037</b>	
NFKB1	rs4986790*	D299G	Benigna, Cons	AA/AG/GG	1134/133/3	316/45/0	1.22 (0.85-1.75)	0.273	
	rs4986791*	T399I	Deletérea, Cons	CC/CT/TT	1124/134/5	309/45/0	1.23 (0.86-1.77)	0.156	
	rs11536889*	UTR 3'	RR, Conservado	GG/GC/CC	940/308/35	258/98/9	1.09 (0.87-1.36)	0.457	
	rs2085549*	UTR 5'	Intergénico	TT/TC/CC	754/450/79	70/30/6	1.00 (0.82-1.21)	0.965	0.577
	rs2085548	Intergénico	Intergénico	GG/GA/AA	642/531/111	57/42/8	1.00 (0.83-1.20)	1.000	
	rs747559	Intergénico	Intergénica, Cons	AA/AG/GG	417/610/177	39/50/15	1.04 (0.87-1.24)	0.669	
	rs3774934*	Intrónico	RR, Cons	GG/GA/AA	1045/225/14	145/35/1	0.99 (0.75-1.31)	0.972	
	rs1585214	Intrónico	Triple hélice	GG/GA/AA	387/635/250	119/182/61	0.89 (0.75-1.05)	0.178	
	rs4648022	Intrónico	RR	CC/CT/TT	1068/206/8	298/59/8	1.23 (0.93-1.62)	0.146	
	rs4648055	Intrónico	Conservado	GG/GA/AA	656/511/117	183/155/27	0.99 (0.82-1.18)	0.896	
rs9790601	rs9790601	Intrónico	Intrónico	AA/AG/GG	624/532/125	170/161/32	1.03 (0.86-1.23)	0.747	
	rs1609798	Intrónico	Intrónico, Cons	CC/CT/TT	628/528/124	49/46/12	1.02 (0.85-1.22)	0.830	
	rs7677509	Región 3'	RR, Región 3'	CC/CT/TT	323/627/327	32/47/26	0.88 (0.74-1.03)	0.113	
	rs765789	Intrónico	Cons, Intrónico	TT/TA/AA	845/395/45	103/66/11	1.22 (0.99-1.49)	0.062	
	rs228611	Intrónico	RR, Conservado	GG/GA/AA	341/639/297	114/175/75	0.86 (0.73-1.01)	0.073	

<sup>a</sup> Los SNPs están orientados en la dirección 5'-3' de cadena codificante. SNPs marcados con \* no son fiables debido a N<5 en el genotipo de riesgo.  
<sup>b</sup> El gen *CD14* está orientado en la dirección 5'-3' en la cadena no codificante. La localización de los tagsSNPs fue buscada mediante el programa Pupasuite v3.1. <sup>c</sup> Resultado de programas Variant Effect Predictor y Pupasuite v3.1, RR significa región regulatoria, NMD significa que el SNP reside en transcritto sujeto a mecanismo *Non sense-mediated mRNA decay*, Cons significa conservado, que consiste en que el SNP se localiza en región conservada evolutivamente entre ratón y humano. <sup>d</sup> Los genotipos son reportados como alelo común/alelo menor. <sup>e,f</sup> Número de controles y casos con los genotipos indicados en <sup>d</sup>. <sup>g</sup> OR (IC 95%) calculados para el modelo log-aditivo. En el caso de rs4986790 and rs4986791, solo el modelo codominante fue estimado. <sup>h</sup> p-valor asociado al OR del análisis de asociación por regresión logística. <sup>i</sup> p-valor del test de permutaciones por regiones genéticas.

Tabla 17. Asociación estadística de SNPs con respecto a la localización anatómica del cáncer gástrico.										
CG tipo no cardíacas (N=181)										
Gen	SNP	Número de casos con cada genotipo <sup>a</sup>	OR (IC 95%) <sup>b</sup>	p-valor <sup>b</sup>	p-valor de permutación <sup>c</sup>	Número de casos con cada genotipo <sup>a</sup>	OR (IC 95%) <sup>b</sup>	p-valor <sup>b</sup>	p-valor de permutación <sup>c</sup>	p-valor de heterogeneidad <sup>d</sup>
<i>CD14</i>	rs11167532	40/46/10	<b>0.69 (0.49-0.95)</b>	<b>0.023</b>	<b>0.010*</b>	51/74/33	1.13 (0.89-1.44)	0.328	0.907	<b>0.007</b>
	rs778588	44/52/11	<b>1.50 (1.09-2.05)</b>	<b>0.013</b>		94/70/16	1.02 (0.79-1.31)	0.881		0.060
	rs2569190	36/55/15	<b>0.69 (0.51-0.93)</b>	<b>0.013</b>		51/79/42	0.99 (0.78-1.25)	0.939		<b>0.025</b>
	rs1583005	25/55/27	<b>1.61 (1.20-2.14)</b>	<b>0.001</b>		65/83/33	1.05 (0.84-1.32)	0.645		0.056
<i>NOD2</i>	rs7202124	58/40/9	0.92 (0.67-1.26)	0.595	0.786	117/58/5	<b>0.60 (0.46-0.79)</b>	<b>0.00019</b>	<b>0.0003*</b>	0.090
	rs2111235	53/39/13	0.99 (0.74-1.34)	0.961		115/56/8	<b>0.58 (0.44-0.76)</b>	<b>3.664e-05</b>		<b>0.021</b>
<i>NFKB1</i>	rs765789	71/31/5	1.09 (0.76-1.56)	0.655	0.617	103/66/11	<b>1.42 (1.09-1.85)</b>	<b>0.012</b>	0.131	0.154
	rs228611	34/47/26	0.93 (0.70-1.24)	0.635		61/85/35	<b>0.79 (0.63-0.99)</b>	<b>0.041</b>		0.243

<sup>a</sup> Reportados para los homocigotos del alelo común/heterocigotos/homocigotos del alelo menor. <sup>b</sup> OR (IC95%) y p-valores son el resultado del análisis de regresión logística con el modelo log-aditivo ajustado. <sup>c</sup> p-valores resultantes del test de permutaciones para las localizaciones cardíacas y no cardíacas. La significación después de la corrección de Bonferroni fue fijada en 0.05/4 genes=0.0125 (\*). <sup>d</sup> p-valor resultante del test de heterogeneidad para saber si existen diferencias en el efecto del SNP entre los grupos.



Tabla 18. Asociación estadística de SNPs previamente asociados con cáncer gástrico no cardíacos en individuos infectados por <i>Helicobacter pylori</i> con respecto al estatus de caga.						
Infectados por <i>H. pylori</i> y caga+ (controles N=584, CG no cardíacos N=151) <sup>a</sup>						
Gen	SNP	Número de controles con cada genotipo <sup>b</sup>	Número de casos con cada genotipo <sup>b</sup>	OR (IC 95%) <sup>c</sup>	p-valor <sup>c</sup>	Número de controles con cada genotipo <sup>b</sup>
<i>NOD2</i>	rs7202124	313/225/45	97/48/5	<b>0.66 (0.47-0.91)</b>	<b>0.009</b>	92/77/25
	rs2111235	295/224/60	99/43/7	<b>0.57 (0.42-0.79)</b>	<b>0.0003</b>	86/83/23
	rs5743289	409/54/21	99/44/8	1.23 (0.90-1.70)	0.204	148/43/4
						9/3/0
						7/4/1
						4/7/1
						0.3 (0.08-1.08)
						0.68 (0.25-1.87)
						<b>3.76 (1.33-10.63)</b>
						0.035
						0.446
						<b>0.012</b>

<sup>a</sup> El estatus de caga fue calculado respecto a todos los casos de CG no cardíacos (163) y controles (779) infectados por *Helicobacter pylori*. <sup>b</sup> Reportados para los homocigotos del alelo común/heterocigotos/homocigotos del alelo menor. <sup>c</sup> OR (IC 95%) y p-valores resultantes del análisis de regresión logística con el modelo log-aditivo ajustado para país, sexo y edad.

Tabla 19. Asociación estadística de SNPs con respecto al subtipo histológico de cáncer gástrico.										
CG Intestinal (N=126)										
CG Difuso (N=128)										
Gen	SNP	Número de casos con cada genotipo <sup>a</sup>	OR (IC 95%) <sup>b</sup>	p-valor <sup>b</sup>	p-valor de permutación <sup>c</sup>	Número de casos con cada genotipo <sup>a</sup>	OR (IC 95%) <sup>b</sup>	p-valor <sup>b</sup>	p-valor de permutación <sup>c</sup>	p-valor de heterogeneidad <sup>d</sup>
<i>CD1</i>	rs778588	52/59/15	<b>1.49 (1.12-1.97)</b>	<b>0.006</b>	<b>0.041*</b>	76/41/10	0.80 (0.59-1.09)	0.151	0.652	<b>0.010</b>
<i>NOD</i>	rs7202124	74/43/7	<b>0.73 (0.54-0.99)</b>	<b>0.040</b>	0.077	78/44/5	<b>0.72 (0.53-0.99)</b>	<b>0.037</b>	0.204	0.922
	rs2111235	74/44/7	<b>0.68 (0.50-0.93)</b>	<b>0.011</b>		69/50/8	0.83 (0.62-1.10)	0.191		0.232
<i>NFK</i>	rs1585214	38/66/21	0.94 (0.72-1.23)	0.665	0.727	46/65/16	<b>0.76 (0.58-0.99)</b>	<b>0.044</b>	0.213	0.129
	rs228611	40/53/32	0.93 (0.72-1.21)	0.605		39/74/15	<b>0.74 (0.56-0.96)</b>	<b>0.024</b>		0.144
	rs765789	75/44/7	1.31 (0.96-1.80)	0.096		73/43/11	<b>1.46 (1.07-1.99)</b>	<b>0.018</b>		0.280
	rs7677509	36/54/35	0.97 (0.75-1.26)	0.819		36/72/19	0.77 (0.59-1.00)	0.047		0.107

<sup>a</sup> Reportados para los homocigotos del alelo común/ heterocigotos/homocigotos del alelo menor. <sup>b</sup> OR (IC 95%) y p-valores resultantes del análisis de regresión logística con el modelo log-aditivo ajustado para país, sexo y edad. <sup>c</sup> p-valores resultantes del test de permutaciones para localizaciones cardíacas y no cardíacas. La significación después de la corrección de Bonferroni fue fijada en 0.05/4 genes=0.0125, por lo que un p-valor<0.0125 fue considerado significativo (\*). <sup>d</sup> p-valor resultante del test de heterogeneidad realizado para saber si existen diferencias en el efecto del SNP entre los grupos.

Tabla 20. Asociación estadística de haplotipos respecto a los subtipos histológicos y anatómicos de cáncer gástrico.							
Gen	Haplotipo	SNPs que componen el haplotipo <sup>a</sup>	Frecuencia haplotípica en controles	Frecuencia haplotípica en casos	OR (IC 95%) <sup>b</sup>	p-valor <sup>c</sup>	Subtipo histológico o anatómico de CG
CD14	ACTCAC	rs11167532,rs7711117,	0.413	-----	1.00 (Referencia)	-----	-----
		rs778588,rs4914,rs2569190,					
		rs1583005					
CD14	GCCCGI	Idem	0.128	0.212	2.15 (1.41-3.27)	<b>0.0004</b>	Cardias
CD14	GCCCGI	Idem	0.128	0.193	1.71 (1.17-2.49)	<b>0.006</b>	Intestinal
NOD2	ACCAT	rs7202124, rs2111235, rs5743289,	0.355	-----	1.00	-----	-----
		rs3135500, rs3785142			(Referencia)		
NOD2	GTCGC	Idem	0.198	0.135	0.60 (0.43-0.86)	<b>0.005</b>	No cardias
NOD2	GTCGC	Idem	0.198	0.141	0.64 (0.43-0.96)	<b>0.031</b>	Intestinal
TLR4	TGATACG	rs1329061, rs1329060, rs1329057,		-----	1.00	-----	-----
		rs10491851, rs4986790,	0.267		(Referencia)		
		rs4986791, rs11536889					
TLR4	CGAGACG	Idem	0.074	0.112	1.98 (1.14-3.47)	<b>0.016</b>	Cardias
NFKB1	TGAGACGACTTA	rs2085549, rs2085548, rs747559,		-----	1.00	-----	-----
		rs3774934, rs1585214, rs4648022,			(Referencia)		
		rs4648055, rs9790601, rs1609798,	0.328				
		rs7677509, rs765789, rs228611					
NFKB1	TGAGGTGACCAG	Idem	0.066	0.095	1.78 (1.15-2.73)	<b>0.009</b>	No cardias

<sup>a</sup> Orden de los SNPs que componen los haplotipos. <sup>b</sup> Odds ratio (IC 95%) y <sup>c</sup> p-valores, ajustados por país, sexo y edad para haplotipos asociados significativamente (p<0.05) con los subtipos histológicos o anatómicos de CG, los cálculos fueron realizados para el modelo log-aditivo y el haplotipo de referencia es el común.

Tabla 21. Búsqueda de potenciales eQTLs entre los SNPs asociados y no asociados a CG.

SNPs asociados a CG	SNPs no asociados a CG	Gen	$r^2$ /p-valor o score <sup>a</sup>	Transcripto regulado	Tejido y referencia	Bases de datos <sup>b</sup>
rs11167532	-----	<i>CD14</i>	0.19/4.29 x 10 <sup>-8</sup>	<i>WDR55</i>	Cerebro <sup>153</sup>	A
rs1583005	-----	<i>CD14</i>	0.36/3.17 x 10 <sup>-15</sup>	<i>WDR55</i>	Cerebro <sup>153</sup>	A
rs2569190	-----	<i>CD14</i>	0.23/1.54 x 10 <sup>-9</sup>	<i>WDR55</i>	Cerebro <sup>153</sup>	A
rs1583005	-----	<i>CD14</i>	30.50	<i>WDR55</i>	Diversos <sup>157</sup>	B
rs778588	-----	<i>CD14</i>	34.51	<i>SRA1</i>	Monocitos <sup>156</sup>	B
rs7202124	-----	<i>NOD2</i>	42.18	<i>NOD2</i>	Monocitos <sup>156</sup>	B
rs7202124	-----	<i>NOD2</i>	16.77	<i>SNX20</i>	Monocitos <sup>156</sup>	B
rs5743289	-----	<i>NOD2</i>	3.903	<i>CYLD</i>	Linfoblastoide <sup>193</sup>	B
-----	rs228611	<i>NFKB1</i>	7.7	<i>MANBA</i>	Linfoblastoide <sup>193</sup>	B
-----	rs7677509	<i>NFKB1</i>	52.94	<i>MANBA</i>	Diversos <sup>157</sup>	B
	rs747559	<i>NFKB1</i>	27.318	<i>COL9A2</i>	Monocitos <sup>156</sup>	B
	rs4648055	<i>NFKB1</i>	23.828	<i>COL9A2</i>	Monocitos <sup>156</sup>	B
	rs1609798	<i>NFKB1</i>	0.012	<i>MANBA</i>	Linfoblastoide <sup>155</sup>	B

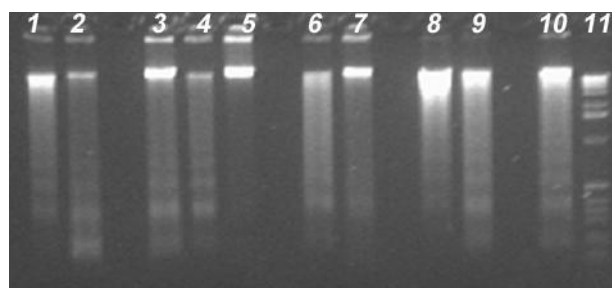
<sup>a</sup> Valor de  $r^2$  y p-valor asociado de la correlación entre la asociación del SNP con los niveles de expresión génica para la base de datos A y score para la B. <sup>b</sup> A: Base de datos eQTL browser (NCBI), B: Base de datos eQTL de Gilad/Pritchard.

## 2. Estudio de asociación genética en el seguimiento de lesiones precursoras gástricas

### 2.1. Prueba piloto de colecta, conservación y extracción de ADN

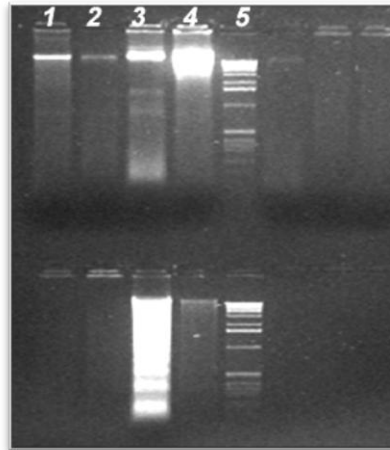
#### Hisopos bucales Isohelix Dri-Capsules y BuccalAmp™ DNA Extraction Kit

El ADN obtenido por Isohelix Dri-Capsules mostró una gran variación en la masa obtenida ( $5.53 \pm 5.17 \mu\text{g}$ ), relaciones de absorbancia 260/280 indicativas de contaminación proteica y con sales, y se observó degradación del mismo, como se evidencia en la siguiente electroforesis (Figura 19).



**Figura 19.** Electroforesis en gel de agarosa 1% de muestras de ADN extraídas de hisopos bucales Dri-Capsules (Isohelix, Inglaterra) conservados 30 días a temperatura ambiente. Leyenda: 1-10 muestras, 11 marcador de peso de ADN 1 Kb ladder (Invitrogen, EUA).

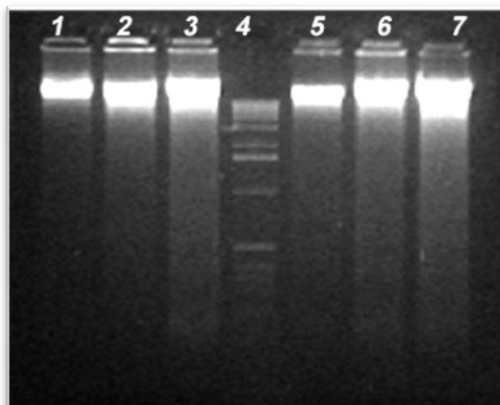
Mediante los hisopos bucales BuccalAmp™ DNA Extraction Kit el rendimiento fue menor y también mostró una gran variabilidad ( $3.01 \pm 3.62 \mu\text{g}$ ), pero menor degradación del ADN obtenido (Figura 20). Al igual que en el kit anterior las relaciones de absorbancia 260/280 fueron indicativas de contaminación proteica y con sales.



**Figura 20.** Electroforesis en gel de agarosa 1% de muestras de ADN extraídas de hisopos bucales BuccalAmp™ DNA Extraction Kit (Epicentre, EUA) conservados 30 días a temperatura ambiente. Leyenda: 1-4 muestras, 5 marcador de peso de ADN 1 Kb ladder (Invitrogen, EUA).

#### **Muestras de saliva (kit Oragene DNA, OG-500, DNA Genotek)**

El ADN de saliva obtenido mediante este kit mostró una degradación insignificante (Figura 21), relaciones de absorbancia 260/280  $>1.8$  demostrando poca contaminación proteica y con sales, y una elevada masa de ADN ( $125 \pm 55.9 \mu\text{g}$ ), suficiente para realizar el genotipado propuesto y que quedara material para futuros proyectos.



**Figura 21.** Electroforesis en gel de agarosa 1% de muestras de ADN extraídas de saliva conservada 30 días a temperatura ambiente empleando el kit Oragene DNA OG-500 (DNA Genotek, Canadá) . Leyenda: 1-3 y 5-7 muestras, 4 marcador de peso de ADN 1 Kb ladder (Invitrogen, EUA).

## 2.2. Diseño de solución estabilizadora de saliva y prueba de conservación

Teniendo en cuenta que en las comparaciones previas los mejores resultados fueron obtenidos en el ADN proveniente de saliva y que los kits eran de un coste elevado (13.9 euros/muestra, Abyntek distribuidor en Europa de DNA Genotek) para el presupuesto disponible, se planteó la posibilidad de confeccionar en el laboratorio un similar de la solución estabilizadora de saliva del kit Oragene DNA, basada en la descrita en la patente que dicho kit <sup>159</sup>. El lenguaje ambiguo de las patentes hace difícil saber cual es la mejor entre todas las soluciones estabilizadoras de saliva descritas, por lo que se realizó una prueba para evaluar el estado de conservación del ADN al cabo de 60 días de coleccionar la saliva en cada una de las soluciones descritas (Tabla 6 en la sección Materiales y Métodos).

Como resultado se observó una oxidación (color amarillo) de las soluciones 1 y 2 debido a la presencia de vitamina C; sin embargo, las soluciones 3 y 4 no presentaron dicho fenómeno (Tabla 22). La vitamina C es un agente reductor que provoca la ruptura de puentes di-sulfuro de mucinas para facilitar la extracción del ADN; sin embargo, experimenta auto-oxidación en presencia de aire y metales pesados con producción de radicales libres del O<sub>2</sub>, los cuales provocan mutaciones al ADN. Dicha auto-oxidación es evitada añadiendo un compuesto de composición desconocida, por lo que descartamos el uso de las soluciones 1 y 2.

Como se obtuvo una mayor masa de ADN con la solución 4 respecto a la 3 se optó por ella como la mejor. Su composición es CDTA 3.3 mM, Tris base 33 mM, SDS 0.6%, LiCl 0.67 M, Urea 0.67 M, EtOH 30%, pH 8, proteinasa K 150 ng/μl. El conocimiento de la acción de sus componentes contribuyó a dicha elección:

-Urea y SDS 0.67 M: Agentes caotrópicos, desnaturalizan DNAsas.

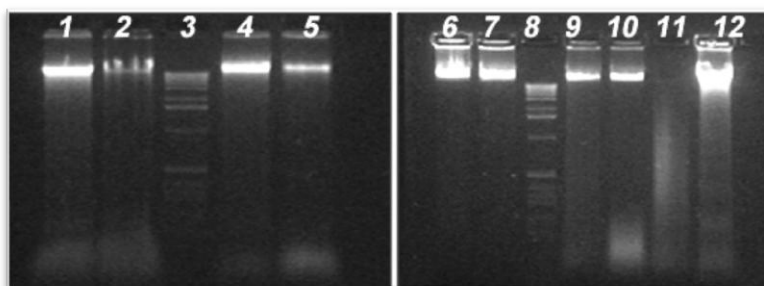
-CDTA: Agente quelante que forma complejos con iones metálicos inhibiendo la generación de radicales libres e inhibiendo DNAsas que usan como cofactores iones metálicos.

-Etanol: Agente antimicrobiano.

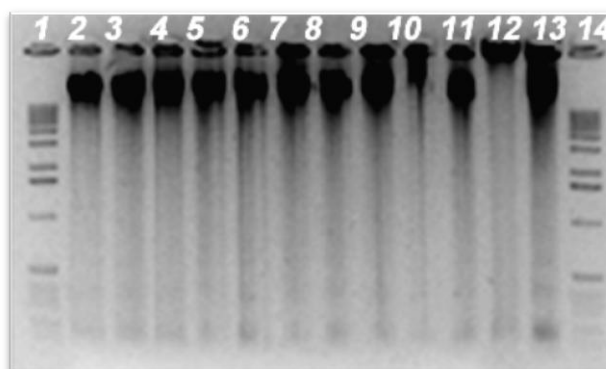
La electroforesis del ADN extraído usando esta solución evidenció que la mayoría de las muestras poseían un adecuado tamaño para ser utilizadas en el genotipado de SNPs (Figura 22). Al ser aplicado este método a la colecta de mayores cantidades de muestras de pacientes en condiciones no ideales respecto a las recomendadas, se observó que la masa ( $23.5 \pm 7.8 \mu\text{g}$ ) y el estado de conservación del ADN (Figura 23) seguía siendo adecuado para el genotipado de SNPs.

Tabla 22. Promedio de la masa de ADN obtenida y presencia de oxidación en las diferentes soluciones estabilizadoras de saliva.		
Soluciones	Masa ADN ( $\mu\text{g}$ ) <sup>a</sup>	Auto-oxidación
1	91,5	SI
2	114,5	SI
3	24,8	NO
4	28,15	NO

<sup>a</sup> Hace referencia a la purificación de un grupo de muestras con cada solución.



**Figura 22.** Electroforesis en gel de agarosa 1% del ADN extraído a partir de muestras de saliva conservadas durante 60 días en solución estabilizadora # 4 a temperatura ambiente. Leyenda: 1, 2, 4-7 y 9-12 muestras, 3 y 8 marcador de peso de ADN 1 Kb ladder (Invitrogen, EUA).



**Figura 23.** Electroforesis en gel de agarosa 1% del ADN extraído a partir de muestras de saliva colectadas por los hospitales en solución estabilizadora # 4. Leyenda: 2-13 muestras, 1 y 14 marcador de peso de ADN 1 Kb ladder (Invitrogen, EUA).

### 2.3. Genotipado de ADN de saliva. Prueba piloto

A continuación se realizó una prueba piloto de genotipado para evaluar la calidad de las muestras, usando sondas de hidrólisis tipo TaqMan para el SNP rs7578034 en el gen *IL1B*, previamente genotipado en el departamento<sup>194</sup>. Se observaron altos valores de score (Tabla 23) y una clara distinción entre los genotipos homocigotos y heterocigotos (Figura 24), indicativos de un adecuado genotipado. Se concluye que el método permite obtener ADN genotipable por sondas TaqMan, por lo que

también podría ser genotipado por otra plataforma diferente como la Sequenom, la cual había sido propuesta por el CEGEN como la adecuada de acuerdo a las características de nuestro estudio.

Tabla 23. Resultados del genotipado del SNP rs7578034 en gen <i>IL1B</i> en muestras de ADN extraído de saliva.		
Muestras	Call	Score
10001	Alelo X	0.94
10002	XY	0.99
10003	XY	0.98
10004	Alelo Y	0.9
10006	Alelo X	0.84
10007	XY	0.97
10008	Alelo X	0.87
10009	XY	0.93
10009	Alelo Y	0.9
10010	XY	0.95
Control +	Alelo Y	0.98
Control +	XY	0.98
Control +	XY	0.91
Control +	XY	0.98
Control +	Alelo X	0.98
Control +	Alelo Y	0.94
H <sub>2</sub> O	Negativo	0.00
H <sub>2</sub> O	Negativo	0.00

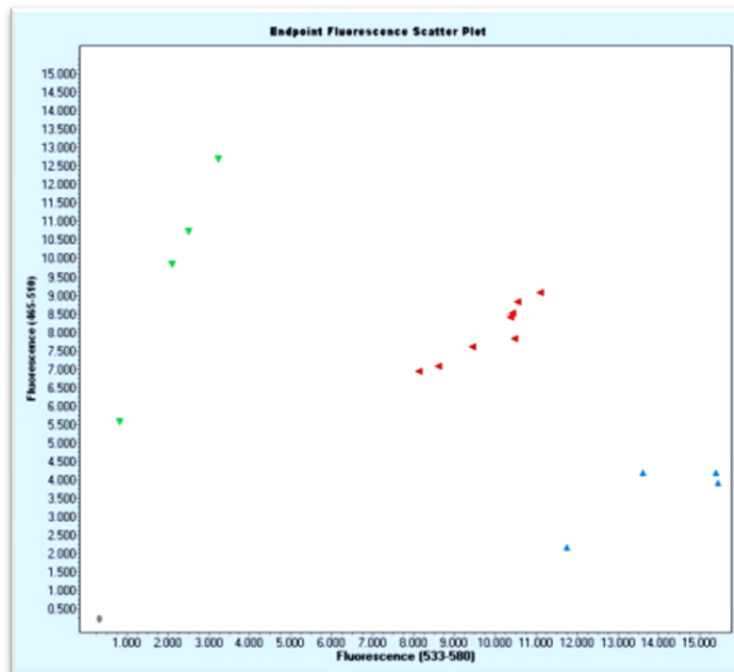


Figura 24. Genotipado del SNP rs7578034 en el gen *IL1B* empleando LightCycler II.

## 2.4. Genotipado de ADN en plataforma Sequenom

Siete (rs11079008, rs2236851, rs28362491, rs4072037, rs6090575, rs6688452 y rs6890699) de los 148 SNPs seleccionados fueron excluidos del análisis por generar clusters de genotipos continuos o solapantes. Los genotipos de otros 11 SNPs (rs11614544, rs1547836, rs225355, rs225357, rs2907749, rs4636306, rs4837404, rs6597947, rs6894617, rs7269342 y rs9976977) se separaban en clusters no perfectamente diferenciados; sin embargo, las réplicas y controles realizados confirmaron la reproducibilidad de los resultados obtenidos, por lo que fueron analizados (Figura 25). El resto de los SNPs poseían clusters con una adecuada diferenciación genotípica (Figura 26). Todos los SNPs seleccionados para el análisis se encuentran en equilibrio de Hardy-Weinberg ( $p > 0.001$ ) y fueron genotipados en más de un 95% de las muestras.

En relación a las muestras, 28 no superaron los controles de calidad por carecer de resultados en más del 20% de los SNPs analizados, por lo que no fueron analizadas estadísticamente. La concordancia entre los genotipos de 33 muestras duplicadas como controles internos de calidad del genotipado fue del 99%. Los genotipos del trío de muestras controles Coriell (Na10830, Na10831 y Na12147) coinciden con los datos publicados. La muestra final disponible para los análisis estadísticos, tras excluir muestras de calidad inadecuada y SNPs no genotipados, fue de 559 casos y 141 SNPs en 29 genes distintos.

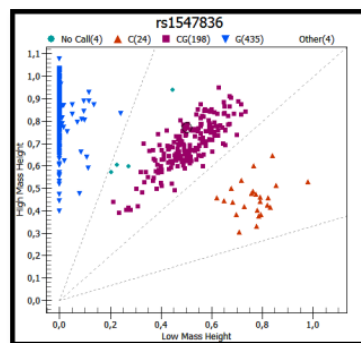


Figura 25. Cluster de calidad media en el genotipado Sequenom.

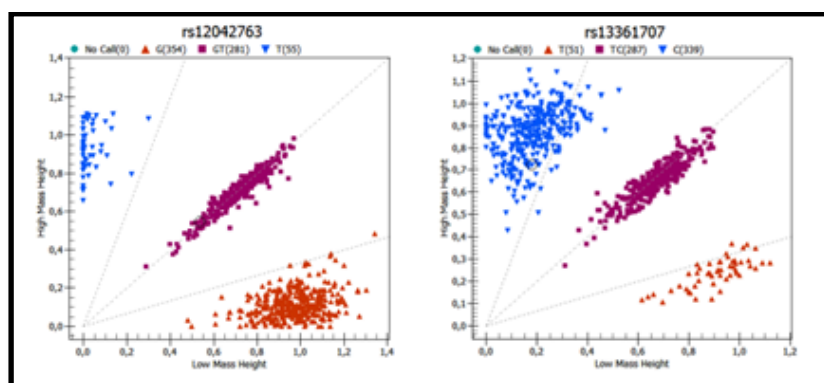


Figura 26. Clusters de calidad óptima en el genotipado Sequenom.



## 2.5. Análisis estadísticos de asociación genética

La evolución de las LPGs fue analizada teniendo en cuenta el diagnóstico histológico al reclutamiento y al final del seguimiento del mismo paciente, acorde a la severidad de las LPGs de la cascada de Correa (Figura 10). Basado en ello la muestra (N=559) se divide en los grupos regresión (N=204, 36.5%), estables (N=259, 46.3%), y progresión (N=96, 17.2%) si el diagnóstico disminuye en severidad, se mantiene o empeora, respectivamente. Las variables analizadas (sexo, edad, consumo de AINES, antecedentes familiares de CG, tabaquismo, infección por *H. pylori* y diagnóstico histológico) de la población al reclutamiento se muestran en la tabla 24. En la comparación estadística de la distribución de dichas variables entre pacientes que progresan, regresan o permanecen estables, sólo el diagnóstico histológico al reclutamiento mostró diferencias significativas ( $p < 0.0001$ ).

Los resultados del genotipado de los 141 SNPs analizados en cada uno de los grupos de evolución de las lesiones se indican en la tabla anexa 1. La asociación más significativa (OR=0.63, IC95% 0.48-0.81,  $p=0.0003$ , modelo aditivo) se observó en el SNP rs10902073 de *MUC2* asociado inversamente con la regresión de las LPGs; fue el único polimorfismo significativo después de aplicar la corrección para comparaciones múltiples mediante FDR. Los SNPs asociados con un incremento del riesgo de progresión se encontraron en los genes *CD14*, *DNAH11*, *MAPK3*, *PRKAA1*, *RUNX3* y *TFF1*. Otros SNPs localizados en los genes *CD14*, *CDH1*, *PTGS2*, *TFF1* y *TFF2*, se asociaron inversamente con la progresión de las lesiones (Tabla 25A). Los resultados del SNP rs10760634 no fueron considerados fiables debido al amplio rango de los límites de confianza del 95% en el cálculo del OR.

El análisis de la regresión mostró que SNPs localizados en los genes *CD14*, *IL1B*, *IL1RN*, *NFKB1*, *TFF1* y *TFF2* se asocian con una mayor regresión; mientras que SNPs en *CDH1*, *MUC1*, *MUC2*, *NFKB1* y *TFF1* se asocian con una menor regresión. También, existen algunos SNPs cuyos resultados no fueron considerados a causa del amplio rango de los límites de confianza del 95% en el cálculo del OR (Tabla 25B). A continuación, se exploró el sentido de la asociación de aquellos SNPs que se asociaron tanto a progresión como a regresión de las lesiones y se eliminó del análisis a aquellos en los que existía el mismo sentido en ambos análisis, rs1598861 y rs7674640 de *NFKB1*, por considerarlos no informativos respecto a la evolución de las LPGs (Tabla 26).

Tabla 24. Características al reclutamiento de la población estudiada.									
Características	Regresión <sup>a</sup>		Estables <sup>b</sup>		Progresión		Total		p-valor <sup>d</sup>
	N	%	N	%	N	%	N	%	
Sexo									
Hombres	108	19.32	137	24.51	57	10.2	302	54.03	0.5128
Mujeres	96	17.17	122	21.82	39	6.98	257	45.97	
Edad									
≤50	93	16.64	103	18.43	32	5.72	228	40.79	0.1184
>50	111	19.86	156	27.91	64	11.4	331	59.21	
Consumo de AINES									
SI	95	17.46	136	25.00	43	7.90	274	50.37	0.5827
NO	102	18.75	122	22.43	46	8.46	270	49.63	
Historia familiar de CG									
SI	18	3.35	40	7.45	16	2.98	74	13.78	0.0771
NO	174	32.40	215	40.04	74	13.7	463	86.22	
Tabaquismo									
Fumadores o ex-fumadores	112	20.44	149	27.19	53	9.67	314	57.30	0.8814
No fumadores	88	16.06	109	19.89	37	6.75	234	42.70	
Infección por <i>H. pylori</i>									
SI	168	30.38	211	38.16	72	13.0	451	81.56	0.6021
cagA+,VacAs1m1	90	68.7	18	13.74	23	17.5	131	36.29	
Otras combinaciones	93	40.43	105	45.6	32	13.9	230	63.71	
NO	34	6.15	48	8.68	20	3.62	102	18.44	
Diagnóstico									
GCA	54	9.66	52	9.30	35	6.26	141	25.22	<b>&lt; 0.0001</b>
MIC	56	10.02	116	20.75	43	7.69	215	38.46	
MII	83	14.85	83	14.85	18	3.22	184	32.92	
DISP	11	1.97	8	1.43	0	0.00	19	3.40	
Total	204	36.49	259	46.33	96	17.1	559	100	

<sup>a</sup>, <sup>b</sup>, <sup>c</sup> Indica pacientes que pasan a una lesión de menor, igual o mayor severidad respectivamente, desde el reclutamiento hasta el fin de seguimiento. <sup>d</sup> p-valor resultante del test de X<sup>2</sup> al comparar los grupos a vs b vs c.

Tabla 25. OR (IC 95%) de los SNPs asociados significativamente con progresión y regresión de las LPGs.

A) Progresión										
Gen	SNP <sup>a</sup>	Modelo genético <sup>b</sup>	Genotipos	Número de controles con cada genotipo <sup>c</sup>	Número de casos con cada genotipo <sup>d</sup>	OR (IC 95%)	p-valor	Variant Effect Predictor <sup>e</sup>	PupaSuite <sup>f</sup>	
CD14	rs11167532	log-aditivo	GG/AG/AA	118/211/104	42/42/11	0.63 (0.46-0.87)	0.004	Región 5'	-----	
CD14	rs2569190	log-aditivo	GG/AG/AA	113/224/123	34/47/15	0.65 (0.47-0.89)	0.006	Intrónico, UTR 5', RR	Conservada/ Triple hélice	
CD14	rs5744455	log-aditivo	CC/CT/TT	293/154/14	49/40/7	1.63 (1.13-2.36)	0.01	Región 5', RR	Conservada	
CD14	rs1583005	log-aditivo	CC/CT/TT	152/235/74	21/50/25	1.56 (1.13-2.16)	0.006	Intrónico, NMD	-----	
CDH1	rs4783573	recesivo	AA/AG/GG	196/205/60	42/48/5	0.37 (0.14-0.95)	0.02	Intrónico, NMD	-----	
DNAH11	rs2285947	dominante	GG/AG/AA	128/217/116	20/56/20	1.57 (1.01-2.46)	0.045	Intrónico	Conservada	
MAPK3	rs11865086	dominante	AA/AC/CC	123/222/116	16/57/23	1.82 (1.02-3.23)	0.033	UTR 3', NMD	Conservada	
NFKB1	rs1598861 *	recesivo	AA/AC/CC	201/184/77	46/42/8	0.45 (0.21-0.98)	0.028	Intrónico	Conservada	
NFKB1	rs7674640 *	recesivo	TT/CT/CC	115/217/130	30/51/14	0.46 (0.26-0.83)	0.004	Región 3', RR	Conservada	
PRKAA1	rs13361707	log-aditivo	CC/CT/TT	240/193/29	44/42/10	1.44 (1.02-2.02)	0.038	Intrónico, RR	Conservada	
PTGES	rs10760634 *	recesivo	AA/AG/GG	392/69/2	79/14/3	7.44 (1.23-45.11)	0.031	Intrónico	Conservada	
PTGS2	rs5275	log-aditivo	TT/CT/CC	216/194/50	58/32/6	0.64 (0.45-0.92)	0.013	UTR 3'	Conservada	
PTGS2	rs4648276	dominante	TT/CT/CC	315/131/14	77/19/0	0.52 (0.31-0.87)	0.03	Intrónico, NMD	Conservada	
RUNX3	rs6663310	dominante	TT/CT/CC	215/191/56	35/51/10	1.61 (1.03-2.5)	0.035	Región 3'	-----	
TFF1	rs9976977	recesivo	GG/AG/AA	176/231/54	29/47/20	1.46 (1.05-2.02)	0.023	Region 3'	-----	
TFF1	rs424694	log-aditivo	CC/CT/TT	162/217/84	44/41/11	0.69 (0.5-0.96)	0.025	Region 3'	-----	
TFF2	rs1079380	log-aditivo	AA/AG/GG	113/243/107	32/49/15	0.7 (0.51-0.98)	0.033	Intrónico	Conservada	

<sup>a</sup> \* Resultados para estos SNPs no considerados debido a que la dirección de asociación en el análisis de progresión y regresión tiene el mismo sentido (SNPs de NFKB1) o a un excesivo rango en los límites de confianza del 95% del OR (rs10760634). <sup>b</sup> Modelo genético con mayor significación estadística. <sup>c</sup> <sup>d</sup> Número de controles (regresión + estables) y de casos (progresión) de homocigotos para el alelo más común, heterocigotos y homocigotos para el alelo menos común, respectivamente. <sup>e</sup> <sup>f</sup> Predicción de efecto funcional en los programas indicados. Conservada se refiere a región conservada evolutivamente entre humanos y ratón.

B) Regresión										
Gen	SNP <sup>a</sup>	Modelo genético <sup>b</sup>	Genotipos	Número de controles con cada genotipo <sup>c</sup>	Número de casos con cada genotipo <sup>d</sup>	OR (IC 95%)	p-valor	Variant Effect Predictor <sup>e</sup>	PupaSuite <sup>f</sup>	
<i>CD14</i>	rs778588	Sobre-	TT/CT/CC	187/138/30	86/103/15	1.6 (1.13-2.27)	0.008	Región 3'	-----	
<i>CDH1</i>	rs16260	log-aditivo	CC/AC/AA	173/146/33	113/80/10	0.76 (0.57-1)	0.048	Región 5'	Conservada	
<i>CDH1</i>	rs12597188	log-aditivo	GG/AG/A	152/159/43	101/84/15	0.75 (0.58-0.98)	0.036	Intrónico, NMD	-----	
<i>CDX2</i>	rs4503658 *	recesivo	GG/GT/TT	312/41/2	169/30/5	1.54 (1.01-2.37)	0.047	UTR 5', RR	Conservada	
<i>IL1B</i>	rs1143627	dominante	TT/CT/CC	165/152/36	77/103/24	1.45 (1.02-2.06)	0.038	Región 5', RR	Conservada	
<i>IL1RN</i>	rs2637988	recesivo	AA/AG/G	139/165/51	80/80/44	1.64 (1.05-2.56)	0.031	Intrónico, Región	-----	
<i>MAP3K14</i>	rs16939926	dominante	TT/CT/CC	305/49/1	188/16/0	0.52 (0.29-0.94)	0.023	Intrónico	Conservada	
<i>MUC1</i>	rs2070803	dominante	TT/CT/CC	106/168/80	79/86/38	0.67 (0.47-0.96)	0.031	Conservada	Conservada	
<b><i>MUC2</i></b>	<b>rs10902073</b>	<b>log-aditivo</b>	<b>CC/AC/A</b>	<b>141/155/59</b>	<b>110/76/18</b>	<b>0.63 (0.48-0.81)</b>	<b>0.0003 *</b>	Intergénico, RR	-----	
<i>MUC2</i>	rs10794281	log-aditivo	TT/CT/CC	123/158/74	85/93/26	0.55 (0.34-0.9)	0.014	Intergénico	Conservada	
<i>MUC6</i>	rs6597947 *	recesivo	GG/GT/TT	307/46/2	172/25/6	5.38 (1.07-26.89)	0.025	Región 3'	Conservada	
<i>NFKB1</i>	rs980455	log-aditivo	AA/AG/G	144/153/55	99/97/41	1.29 (1.01-1.65)	0.038	Región 3'	-----	
<i>NFKB1</i>	rs1598861	recesivo	AA/AC/CC	153/137/64	94/89/21	0.52 (0.31-0.88)	0.012	Intrónico	Conservada	
<i>NFKB1</i>	rs7674640	recesivo	TT/CT/CC	87/164/103	58/104/41	0.62 (0.41-0.93)	0.019	Región 3', RR	Conservada	
<i>PTGS2</i>	rs5277	recesivo	GG/CG/C	225/109/21	139/61/4	0.32 (0.11-0.94)	0.021	Sinónimo, NMD	Conservada	
<i>SRC</i>	rs6017901 *	dominante	TT/CT/CC	314/39/1	193/11/0	0.45 (0.22-0.89)	0.016	Región 3'	-----	
<i>TFF1</i>	rs4920094	dominante	GG/AG/A	97/194/64	75/93/36	0.65 (0.45-0.93)	0.021	Intrónico	Conservada	
<i>TFF1</i>	rs225357	recesivo	CC/CT/TT	121/179/54	61/97/46	1.62 (1.04-2.51)	0.032	Intrónico	Conservada	
<i>TFF1</i>	rs424694	recesivo	CC/CT/TT	134/171/50	72/87/45	1.73 (1.11-2.7)	0.017	Región 3'	-----	
<i>TFF1</i>	rs13047838	dominante	CC/CT/TT	178/152/25	121/74/9	0.72 (0.54-0.97)	0.026	UTR 3'	Conservada	
<i>TFF2</i>	rs1079380	recesivo	AA/AG/G	99/189/67	46/103/55	1.59 (1.06-2.39)	0.027	Intrónico	Conservada	

<sup>a</sup> \* Resultados para estos SNPs no considerados debido a que el número de individuos con el genotipo de riesgo es muy bajo o el intervalo de confianza muy amplio. <sup>b</sup> Modelo genético con mayor significación estadística. <sup>c</sup> <sup>d</sup> Número de controles (regresión + estables) y de casos (progresión) de homocigotos para el alelo más común, heterocigotos y homocigotos para el alelo menos común, respectivamente. <sup>e</sup> <sup>f</sup> Localización del SNP y predicción de efecto funcional en los programas indicados. Conservada se refiere a región conservada evolutivamente entre humanos y ratón.

**Tabla 26. Asociación con la evolución de LPGs en el presente y en el estudio de seguimiento de LPGs de Soria.**

Selección	Gen	SNP	Presente estudio		Estudio de Soria	
			Progresión	Regresión	Progresión	Regresión
Mecanismo de infección de <i>H.pylori</i>	<i>CD14</i>	rs11167532	X			
	<i>CD14</i>	rs1583005	X			
	<i>CD14</i>	rs2569190	X		X	
	<i>CD14</i>	rs5744455	X			X
	<i>CD14</i>	rs778588		X		
	<i>NFKB1</i>	rs1598861	X	X		
	<i>NFKB1</i>	rs7674640	X	X		
	<i>NFKB1</i>	rs980455		X		
	<i>MAPK3</i>	rs11865086	X			
	<i>MAP3K14</i>	rs16939926		X		
	<i>SRC</i>	rs6017901		X		
	<i>NOD1</i>	rs11536450		X		
	<i>CDH1</i>	rs4783573	X			
	<i>CDH1</i>	rs12597188		X		
	<i>CDH1</i>	rs16260		X		
	<i>CDH1</i>	rs1125557	No asociado	No asociado		X
Modelos murinos de carcinogénesis gástrica	<i>PTGS2</i>	rs5275	X			
	<i>PTGS2</i>	rs4648276	X			
	<i>PTGS2</i>	rs5277		X		
	<i>PTGES</i>	rs10760634	X	X		
	<i>TFF1</i>	rs424694	X	X		X
	<i>TFF1</i>	rs9976977	X			X
	<i>TFF1</i>	rs4920094		X		
	<i>TFF1</i>	rs13047838		X		X
	<i>TFF1</i>	rs225357		X		
	<i>TFF1</i>	rs225358	No asociado	No asociado	X	
	<i>Cdx2</i>	rs4503658		X		
	<i>IL1B</i>	rs1143627		X		
	<i>RUNX3</i>	rs6663310	X			
GWAS y estudio de Soria	<i>TFF2</i>	rs1079380	X	X	X	
	<i>MUC1</i>	rs2070803		X		
	<i>MUC2</i>	rs10902073		X	X	X
	<i>MUC2</i>	rs10794281		X	X	X
	<i>MUC2</i>	rs7944723	No asociado	No asociado		X
	<i>MUC2</i>	rs10794293	No asociado	No asociado	X	
	<i>MUC2</i>	rs3924453	No asociado	No asociado	X	
	<i>MUC2</i>	rs4077759	No asociado	No asociado	X	
	<i>MUC6</i>	rs6597947		X	X	X
	<i>DNAH11</i>	rs2285947	X			
	<i>PRKAA1</i>	rs13361707	X			
	<i>IL1RN</i>	rs2637988		X	X	
<i>IL1A</i>	rs17561	No asociado	No asociado		X	

El color verde indica asociación inversa y el rojo asociación directa. Los SNPs sombreados poseen igual sentido de la asociación en los análisis de progresión y regresión de estudios individuales.

Cuando comparamos los resultados obtenidos en este estudio con los obtenidos con SNPs genotipados en un estudio previo de seguimiento de lesiones precursoras, realizado en la provincia de Soria por nuestro grupo, se confirmaron las asociaciones inversas de *CD14* (rs2569190 y rs5744455), *TFF2* (rs1079380) con la progresión. También que se mantuvo la asociación de *MUC2* (rs10902073 y rs10794281) y *TFF1* (rs424694) con la regresión (Tabla 26). Por el contrario, no se replicaron en el presente estudio las asociaciones de *CDH1* (rs1125557), *TFF1* (rs9976977, rs13047838 y rs225358), *MUC2* (rs7944723, rs10794293, rs3924453 y rs4077759), *IL1RN* (rs2637988) e *IL1A* (rs17561).

Al analizar genes que mostraron asociación a nivel de SNP y haplotipos, este análisis confirmó las asociaciones a nivel de SNP (Tabla 27), es decir, correspondiéndose con lo esperado atendiendo al sentido de asociación de sus SNPs componentes. Se encontró asociación con mayor riesgo de progresión de los haplotipos TGTG de *CD14* (OR=1.90, IC95% 1.26-2.87), así como CAAC (OR=1.82, IC95% 1.10-2.98) y CAGC (OR=1.89, IC95% 1.03-3.46) de *TFF1*. La combinación de *CD14* contiene al alelo de riesgo T de rs5744455 que se asocia positivamente con la progresión. El haplotipo GCCGAG de *PTGS2* (OR=0.55, IC95% 0.31-0.99) se asocia inversamente con la progresión, y el alelo C de rs5275 se asocia inversamente con la progresión.

Haplotipos de *CDH1* y *CDX2* se asocian con una mayor probabilidad de de regresión mientras que haplotipos de *PTGES*, *TFF1*, *MAP3K14*, *MUC2* y *PTPN11* se asocian inversamente con la regresión, observándose una buena concordancia en el sentido de la asociación haplotípica esperada teniendo en cuenta la de sus SNPs constituyentes. Por ejemplo, el alelo de riesgo G de rs4783573 de *CDH1* se asocia inversamente con la progresión y el haplotipo AG que lo contiene se asocia con una mayor probabilidad de regresión. Este análisis se realizó para cada haplotipo con significación estadística, confirmando la consistencia en la dirección de asociación en todos los casos. Sin embargo, el gen *PTPN11* sólo muestra una asociación inversa con la regresión a nivel haplotípico.

El análisis de asociación con la evolución de las LPGs en los pacientes portadores de la combinación cagA+vacAs1/m1 en comparación con el resto de combinaciones de cagA y vacA se expone en la tabla 28.

Tabla 27. Asociaciones significativas de haplotipos respecto a la evolución de las LPGs.							
Gen	Haplotipo	SNPs que componen el haplotipo	Frecuencia haplotípica en controles	Frecuencia haplotípica en casos	OR (IC 95%) <sup>a</sup>	p-valor <sup>b</sup>	Evolución
CD14	CATA	rs5744455, rs11167532 rs778588, rs2569190	0.326	0.434	1 (Referencia)	-----	Progresión
CD14	IGTG	Idem	0.196	0.281	1.90 (1.26-2.87)	0.002	Progresión
PTGS2	ATTGAG	rs2745557,rs5275,rs4648276,rs5277,689466, rs12042763	0.205	0.213	1 (Referencia)	-----	
PTGS2	GCCGAG	Idem	0.17	0.098	0.55 (0.31-0.99)	0.047	Progresión
TFF1	CAGT	rs13047838, rs1788413 rs9976977, rs424694	0.410	0.328	1 (Referencia)	-----	Progresión
TFF1	CAAC	Idem	0.110	0.156	1.82 (1.10-2.98)	0.019	Progresión
TFF1	CAGC	Idem	0.059	0.088	1.89 (1.03-3.46)	0.039	Progresión
CDH1	AA	rs7186053, rs4783573	0.392	0.34	1 (Referencia)	-----	Regresión
CDH1	AG	Idem	0.041	0.074	2.14 (1.09-4.19)	0.028	Regresión
PTGES	AAT	rs10760634, rs4837404, rs10739757	0.481	0.53	1 (Referencia)	-----	Regresión
PTGES	GAC	Idem	0.097	0.059	0.55 (0.33-0.90)	0.018	Regresión
TFF1	CAGT	rs13047838, rs1788413 rs9976977, rs424694	0.377	0.428	1 (Referencia)	-----	Regresión
TFF1	TAAC	Idem	0.284	0.225	0.69 (0.50-0.95)	0.025	Regresión
Cdx2	TCG	rs2481952, rs2504212, rs4503658	0.505	0.505	1 (Referencia)	-----	Regresión
Cdx2	CAI	Idem	0.051	0.095	1.78 (1.11-2.85)	0.017	Regresión
MAP3K14	TTTC	rs2074292, rs16939926 rs2074293,	0.473	0.51	1 (Referencia)	-----	Regresión
MAP3K14	CCTC	Idem	0.068	0.039	0.52 (0.28-0.95)	0.033	Regresión
MUC2	CT	rs10902073, rs10794281	0.564	0.644	1 (Referencia)	-----	Regresión
MUC2	AC	Idem	0.380	0.274	0.66 (0.50-0.86)	0.002	Regresión
PTPN11	GCAAGC	rs11066322, rs17822304 rs11614544, rs11066320 rs11066323, rs7958372	0.43	0.458	1 (Referencia)	-----	Regresión
PTPN11	GCAGGC	Idem	0.191	0.134	0.68 (0.47-0.97)	0.034	Regresión

<sup>a, b</sup> Odds ratio (IC 95%) y p-valores, respectivamente, asociados significativamente ( $p < 0.05$ ) con la evolución de las LPGs, los cálculos fueron realizados para el modelo log-aditivo y el haplotipo de referencia es el común.

**Tabla 28. Análisis de asociación en portadores de la combinación de los factores de virulencia de *H. pylori* cagA+vacAs1/m1 vs otras combinaciones, en individuos infectados por *H. pylori*.**

Progresión en cagA+vacAs1/m1		Progresión en el resto de combinaciones								
Gen	SNP <sup>a</sup>	Modelo genético <sup>b</sup>	# de controles con cada genotipo <sup>b</sup>	# de casos con cada genotipo <sup>c</sup>	OR (IC 95%) <sup>c</sup>	p-valor <sup>d</sup>	# de controles con cada genotipo <sup>b</sup>	# de casos con cada genotipo <sup>c</sup>	OR (IC 95%) <sup>c</sup>	p-valor <sup>d</sup>
CD14	rs5744455*	recesivo	72/35/1	13/7/3	16.05 (1.59-162.2)	<b>0.010</b>	129/64/4	17/14/1	1.56 (0.17-14.39)	0.708
CDX1	rs6894617	recesivo	52/49/7	8/10/5	4.01 (1.15-14.02)	<b>0.038</b>	110/75/13	21/9/2	0.95 (0.2-4.42)	0.946
GAST	rs12453761	log-aditivo	31/51/26	11/10/2	0.49 (0.25-0.97)	<b>0.034</b>	50/106/42	9/16/7	0.95 (0.55-1.65)	0.865
NFKB1	rs7674640*	recesivo	26/46/36	7/15/1	0.09 (0.01-0.7)	<b>0.001</b>	51/69/51	11/15/6	0.67 (0.26-1.71)	0.382
RUNX3	rs2282718	dominante	54/44/10	6/13/4	2.83 (1.04-7.73)	<b>0.033</b>	84/93/21	11/13/8	1.41 (0.64-3.07)	0.387
SRC	rs6017916	recesivo	58/43/7	11/6/6	5.09 (1.53-17)	<b>0.011</b>	112/68/18	16/12/4	1.43 (0.45-4.53)	0.557
SRC	rs6094373	recesivo	64/38/6	13/5/5	4.72 (1.3-17.12)	<b>0.024</b>	122/62/14	18/10/4	1.88 (0.58-6.11)	0.319
TFF1	rs225355	sobredominant	26/54/28	10/6/7	0.35 (0.13-0.96)	<b>0.033</b>	63/88/45	10/15/6	1.15 (0.54-2.46)	0.717
CD14	rs11167532*	recesivo	37/47/24	9/11/3	0.53 (0.14-1.92)	0.302	61/90/47	11/19/2	0.21 (0.05-0.93)	<b>0.012</b>
MAP3K14	rs16939926*	log-aditivo	96/12	18/5	2.22 (0.7-7.08)	0.194	177/21/0	24/8/0	2.81 (1.12-7.04)	<b>0.037</b>
MAP3K14	rs2074292	log-aditivo	24/63/21	5/15/3	0.86 (0.42-1.76)	0.684	42/108/46	11/17/3	0.53 (0.29-0.95)	<b>0.030</b>
MAP3K14	rs4247364	dominante	43/51/14	5/15/3	2.38 (0.82-6.89)	0.091	84/93/20	6/20/6	3.22 (1.27-8.18)	<b>0.007</b>
PTGS2	rs4648276*	dominante	73/34/1	17/6/0	0.74 (0.27-2.03)	0.547	129/61/6	27/5/0	0.36 (0.13-0.97)	<b>0.027</b>
PTGS2	rs5275*	dominante	53/46/9	14/7/2	0.62 (0.25-1.55)	0.302	85/88/25	24/7/1	0.25 (0.11-0.59)	<b>0.001</b>
PTGES	rs10760634	log-aditivo	93/14/1	22/1/0	0.29 (0.04-2.21)	0.562	164/33/1	22/8/2	2.34 (1.14-4.83)	<b>0.027</b>
RUNX3	rs2282718	recesivo	54/44/10	6/13/4	2.06 (0.59-7.27)	0.279	84/93/21	11/13/8	2.81 (1.12-7.04)	<b>0.037</b>
RUNX3	rs6663310*	dominante	56/41/11	12/9/2	0.99 (0.4-2.43)	0.978	88/80/29	8/19/5	2.42 (1.04-5.66)	<b>0.032</b>
NFKB1A	rs696	dominante	38/60/10	9/13/1	0.84 (0.33-2.13)	0.72	74/93/31	5/21/5	3.1 (1.14-8.43)	<b>0.015</b>
WNT1	rs4760663	log-aditivo	38/49/21	11/10/2	0.62 (0.31-1.21)	0.147	69/95/33	7/15/10	1.74 (1.02-2.96)	<b>0.041</b>

<sup>a</sup> El asterisco \* indica SNPs que fueron asociados con la evolución de las LPGs en el análisis general y los no marcados con \* indica que son nuevas asociaciones. <sup>b</sup> Modelo genético donde fue obtenida la asociación más significativa. <sup>c, d</sup> Controles (regresión + estables) y casos (progresión). <sup>d</sup> OR (IC 95%) y p-valor asociado del modelo genético con la asociación más significativa.



Regresión en cagaA+vacAs1/m1							Regresión en el resto de combinaciones			
Gen	SNP <sup>a</sup>	Modelo genético <sup>b</sup>	# de controles con cada genotipo <sup>b</sup>	# de casos con cada genotipo <sup>c</sup>	OR (IC 95%) <sup>c</sup>	p-valor <sup>d</sup>	# de controles con cada genotipo <sup>b</sup>	# de casos con cada genotipo <sup>c</sup>	OR (IC 95%) <sup>c</sup>	p-valor <sup>d</sup>
MUC2	rs10794281*	log-aditivo	31/37/22	23/13/5	0.53 (0.31-0.91)	<b>0.016</b>	53/56/28	35/45/13	0.9 (0.62-1.3)	0.576
MUC2	rs10902073*	dominante	38/36/16	27/10/4	0.38 (0.18-0.82)	<b>0.012</b>	57/56/24	47/38/8	0.7 (0.41-1.18)	0.181
CDH1	rs1125557	log-aditivo	21/48/21	18/18/5	0.5 (0.29-0.89)	<b>0.014</b>	38/76/23	31/48/14	0.85 (0.57-1.26)	0.411
CDH1	rs12597188*	log-aditivo	34/46/10	23/14/2	0.49 (0.26-0.93)	<b>0.024</b>	62/63/12	46/39/7	0.86 (0.57-1.31)	0.49
MUC1	rs2070803*	dominante	18/52/20	21/12/8	0.24 (0.11-0.53)	<b>0.0004</b>	45/67/25	35/42/16	0.81 (0.47-1.41)	0.455
NOD1	rs2709803	sobredominante	49/37/4	30/9/2	0.4 (0.17-0.94)	<b>0.029</b>	91/38/7	53/34/6	1.49 (0.85-2.61)	0.169
CDH1	rs4783573*	dominante	50/33/7	15/21/5	2.17 (1.01-4.63)	<b>0.043</b>	54/64/19	36/46/9	0.99 (0.58-1.71)	0.982
CDX1	rs6894617	sobredominante	46/35/9	14/24/3	2.22 (1.05-4.71)	<b>0.036</b>	81/46/10	50/38/5	1.37 (0.79-2.36)	0.261
SRC	rs754625	recesivo	45/40/5	20/12/9	4.78 (1.49-15.35)	<b>0.007</b>	77/50/10	51/33/9	1.36 (0.53-3.49)	0.523
RUNX3	rs9438859	recesivo	51/38/1	26/11/4	9.62 (1.04-89)	<b>0.022</b>	78/49/9	58/28/7	1.15 (0.41-3.2)	0.791
CD14	rs2569190*	log-aditivo	24/38/28	12/21/8	0.77 (0.47-1.27)	0.309	25/69/41	28/44/21	0.68 (0.46-0.99)	<b>0.042</b>
CD14	rs778588*	sobredominante	46/35/9	16/22/3	1.82 (0.86-3.84)	0.115	78/49/10	37/48/8	1.92 (1.12-3.27)	<b>0.017</b>
MAPK3	rs11865086*	recesivo	17/52/21	6/20/14	1.77 (0.78-3.99)	0.172	41/72/23	22/43/28	2.12 (1.13-3.98)	<b>0.019</b>
MAP3K14	rs2074292	log-aditivo	19/55/16	10/23/8	0.96 (0.54-1.72)	0.897	36/76/23	17/49/26	1.55 (1.03-2.32)	<b>0.031</b>
MAP3K14	rs3785803	log-aditivo	76/14/0	31/10/0	1.75 (0.7-4.36)	0.234	102/33/2	80/11/1	0.49 (0.25-0.95)	<b>0.027</b>
MUC6	rs225358	dominante	42/33/14	17/20/3	1.21 (0.57-2.57)	0.62	60/59/13	56/30/6	0.54 (0.31-0.92)	<b>0.023</b>
NFKB1	rs13117745	dominante	69/20/1	34/7/0	0.68 (0.26-1.75)	0.41	117/19/1	66/23/4	2.39 (1.25-4.59)	<b>0.008</b>
TFF1	rs13047838*	dominante	52/32/6	22/18/1	1.18 (0.56-2.48)	0.66	55/74/8	57/29/7	0.42 (0.25-0.73)	<b>0.002</b>
TFF1	rs2156310	dominante	65/22/3	29/12/0	1.08 (0.48-2.43)	0.86	90/44/3	74/16/2	0.47 (0.25-0.87)	<b>0.014</b>
TFF1	rs225359	log-aditivo	41/35/14	18/20/3	0.87 (0.51-1.49)	0.612	58/63/15	52/35/6	0.65 (0.42-0.99)	<b>0.039</b>
TFF1	rs424694*	recesivo	37/40/13	17/5/9	1.67 (0.65-4.28)	0.295	50/68/19	31/38/24	2.16 (1.1-4.23)	<b>0.024</b>
TFF1	rs4920094*	dominante	29/42/19	13/21/7	1.02 (0.46-2.26)	0.953	34/83/20	38/38/17	0.48 (0.27-0.84)	<b>0.010</b>
TFF1	rs9976977*	sobredominante	38/34/18	18/018/5	1.29 (0.61-2.73)	0.508	38/83/15	39/41/13	0.5 (0.29-0.86)	<b>0.011</b>

<sup>a</sup> El asterisco \* indica SNPs que fueron asociados con la evolución de las LPGs en el análisis general y el resto son nuevos. <sup>b</sup> Modelo genético donde fue obtenida la asociación más significativa. <sup>c,d</sup> Controles (progresión + estables) y casos (regresión) respectivamente. <sup>d</sup> OR (IC 95%) y p-valor asociado del modelo genético con la asociación más significativa.

## 2.6 Análisis bioinformáticos de efectos funcionales

El análisis de potenciales efectos funcionales usando el programa Variant Effect Predictor indicó que los SNPs asociados se localizaban en regiones que condicionan la expresión génica como UTR 5' (*CD14* y *CDX2*), UTR 3' (*MAPK3*, *PTGS2* y *TFF1*), regiones regulatorias (*CD14*, *NFKB1*, *PRKAA1*, *CDX2*, *IL1B* y *MUC2*), transcritos regulados por el mecanismo de degradación de ARNm aberrantes (*CD14*, *CDH1*, *MAPK3*, *PTGS2* y *NOD1*) y regiones 5' (*CD14*, *CDH1* y *IL1B*) (Tabla 25). El SNP rs10902073 de *MUC2*, único resultado significativo después de la corrección para comparaciones múltiples, tiene localización intergénica y se encuentra en una región regulatoria. Los SNPs asociados rs13047838 (*TFF1*) y rs2569190 (*CD14*), además de encontrarse en regiones conservadas también se localizan dentro de estructuras de triple hélice, las cuales regulan la expresión génica. Al compararse el número de SNPs asociados y no asociados respecto a sus localizaciones genéticas (Ej región 3') y efectos funcionales (Ej conservados, eQTLs) no existieron diferencias significativas en la prueba exacta de Fisher ( $p > 0.05$ ).

Como en el caso de la asociación con CG los SNPs de *CD14* asociados con la evolución de las LPGs, rs11167532, rs1583005, rs778588 y rs2569190 se asocian con niveles de expresión génica de ARNm y los mismos se encuentran en desequilibrio de ligamiento. Otros SNPs asociados con la evolución de las LPGs pertenecientes a genes *NFKB1*, *MAPK3* e *IL1B* fueron encontrados como eQTLs reguladores de la expresión génica. Mientras que SNPs no asociados con la evolución de LPGs también demostraron ser eQTLs (Tabla 29).

**Tabla 29. Potenciales eQTL entre los SNPs estudiados.**

SNPs asociados a LPGs	SNPs no asociados a LPGs	Gen	$r^2$ /p-valor o score <sup>a</sup>	Transcripto regulado	Tejido y referencia	Bases de datos
rs11167532	-----	<i>CD14</i>	0.196/10 <sup>-08</sup>	<i>WDR55</i>	-----	A
rs2569190	-----	<i>CD14</i>	0.236/10 <sup>-09</sup>	<i>WDR55</i>	-----	A
rs1583005	-----	<i>CD14</i>	0.363/10 <sup>-15</sup>	<i>WDR55</i>	-----	A
rs1583005	-----	<i>CD14</i>	30.498	<i>WDR55</i>	Diversos <sup>157</sup>	B
rs778588	-----	<i>CD14</i>	34.51/17.69	<i>SRA1/DFNA5</i>	Monocitos <sup>156</sup>	B
rs7674640	-----	<i>NFKB1</i>	47.042	<i>MANBA</i>	Diversos <sup>157</sup>	B
rs11865086	-----	<i>MAPK3</i>	0.944	<i>MAPK3</i>	Diversos <sup>157</sup>	B
rs11865086	-----	<i>MAPK3</i>	3.488	<i>TBC1D10B</i>	Linfoblastoide <sup>151</sup>	B
rs1143627	-----	<i>IL1B</i>	0.475	<i>IL1B</i>	Linfoblastoide <sup>151</sup>	B
-----	rs887343	<i>CDX1</i>	3.956	<i>TNIP1</i>	Linfoblastoide <sup>151</sup>	B
-----	rs17561	<i>IL1A</i>	129.51	<i>PVRL2</i>	Monocitos <sup>156</sup>	B
-----	rs178740	<i>TFF1</i>	4.587	<i>DYRK1A</i>	Linfoblastoide <sup>151</sup>	B
-----	rs2867316	<i>MAP3K14</i>	21.942	<i>MAPK8IP1</i>	Monocitos <sup>156</sup>	B
-----	rs7698	<i>MAPK3</i>	11.963	<i>THAP6</i>	Linfoblastoide <sup>155</sup>	B
-----	rs230547	<i>NFKB1</i>	62.605	<i>ORMDL3</i>	Diversos <sup>157</sup>	B
-----	rs696	<i>NFKBIA</i>	9.508	<i>FAM62B</i>	Hígado <sup>152</sup>	B
-----	rs17822304	<i>PTPN11</i>	12.426	<i>ALDH2</i>	Monocitos <sup>156</sup>	B
-----	rs11066322	<i>PTPN11</i>	14.441	<i>MAPKAPK5-AS1</i>	Monocitos <sup>156</sup>	B
-----	rs6094373	<i>SRC</i>	3.819	<i>SOGA1</i>	Linfoblastoide <sup>151</sup>	B

<sup>a</sup> Valor de  $r^2$  y p-valor asociado de la correlación entre la asociación del SNP con los niveles de expresión génica para la base de datos A y score para la B. <sup>b</sup> A: Base de datos eQTL browser (NCBI), B: Base de datos eQTL de Gilad/Pritchard.

### 3. Estudio de expresión génica por microarray en metaplasia intestinal que progresa a cáncer gástrico

#### 3.1. Medición de cantidad y calidad del ARN

Las 6 muestras analizadas para su uso en la prueba piloto de comparación de microarrays (Tabla 30, Figura 27), así como las empleadas para determinar mecanismos funcionales desregulados en la progresión de la MI a CG (N=45) (Tabla 31) resultaron poseer, de manera general, una cantidad y calidad adecuada de ARN (masa media 7.83 µg y RIN medio 2.01) para ser hibridadas en los microarrays Human Gene 2.0 ST y Almac-Xcel. A pesar de que los valores de RIN se encuentran alrededor de 2, lo cual indica un ARN muy degradado por la acción del formaldehído.

#### 3.2. Prueba piloto de comparación de microarrays Human Gene 2.0 ST versus Almac Xcel

Los análisis de calidad demostraron que ambos arrays eran adecuados para trabajar con tejido parafinado. Sin embargo, con el Almac-Xcel se obtuvo una mayor cantidad de genes diferencialmente expresados (N=266) respecto al Human Gene 2.0 ST (N=11). Entre los genes diferencialmente expresados obtenidos con el Almac-Xcel se encontraron múltiples genes codificantes para proteínas con funciones en la fisiología gastrointestinal que se conoce están desreguladas en la MI, Ej. *OLFM4*, *LYZ*, *CDH17*, etc.<sup>81</sup>. Los transcritos de relevancia en la MI obtenidos con el Human Gene 2.0 ST se restringían a *ATP4A/B* (ATPase, H<sup>+</sup>/K<sup>+</sup> exchanging), *GIF* (gastric intrinsic factor) y *PGA3* (pepsinogen 3), los cuales a su vez estaban comprendidos en los resultados del Almac-Xcel. Con base en su amplia cobertura y que el Almac Xcel es un array optimizado para tejido se decidió por el mismo para el análisis de muestras clínicas.

Tabla 30. Medición de cantidad y calidad de las muestras de ARN usadas en la prueba piloto de comparación de microarrays Human Gene 2.0 ST y Almac Xcel.											
Código	Tipo de MI al reclutamiento <sup>a</sup>	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Fin de seguimiento	Estatus <sup>c</sup>	Infección por <i>H. pylori</i>	Sexo <sup>d</sup>	c(ARN) (ng/μl)	A 260/280	RIN	c(ADN) (ng/μl) <sup>e</sup>
A1	No presenta	No presenta	Cuerpo	Sanos	Sanos	SI	H	593.35	2.1	2.1	7.5
A2	No presenta	No presenta	Cuerpo	Sanos	Sanos	NO	H	465.85	1.97	2.2	12
A3	No presenta	No presenta	Cuerpo	Sanos	Sanos	NO	M	487.25	2	2.2	9.8
A4	G1	E4	Antro	CG intestinal	MII-CG	SI	H	148.56	2.03	2.4	9.6
A5	G3	E4	Antro	CG intestinal	MII-CG	SI	M	304.0	2.02	2.4	7
A6	G3	E4	Antro	CG mixto	MII-CG	NO	H	918.74	2.06	2.4	10.3

<sup>a,b</sup> Según se indica en apartado 3.1 de Materiales y Métodos. <sup>c</sup> Sanos: Mucosa gástrica sana, MII-CG: Metaplasia intestinal incompleta que progresa a CG. <sup>d</sup> H: Hombre, M: Mujer. <sup>e</sup> Concentración de ADN residual en la muestra de ARN.

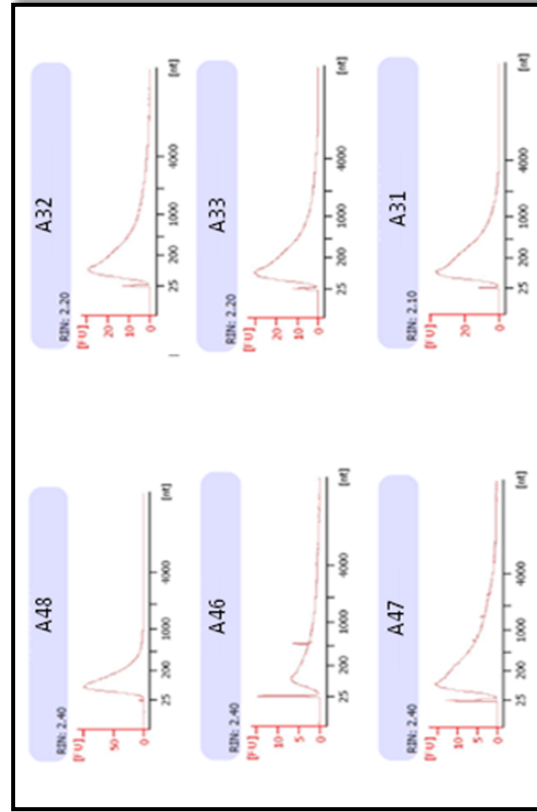


Figura 27. Electroforesis de ARN total obtenido de muestras parafinadas en Bioanalyzer 2100 (Agilent, EUA).

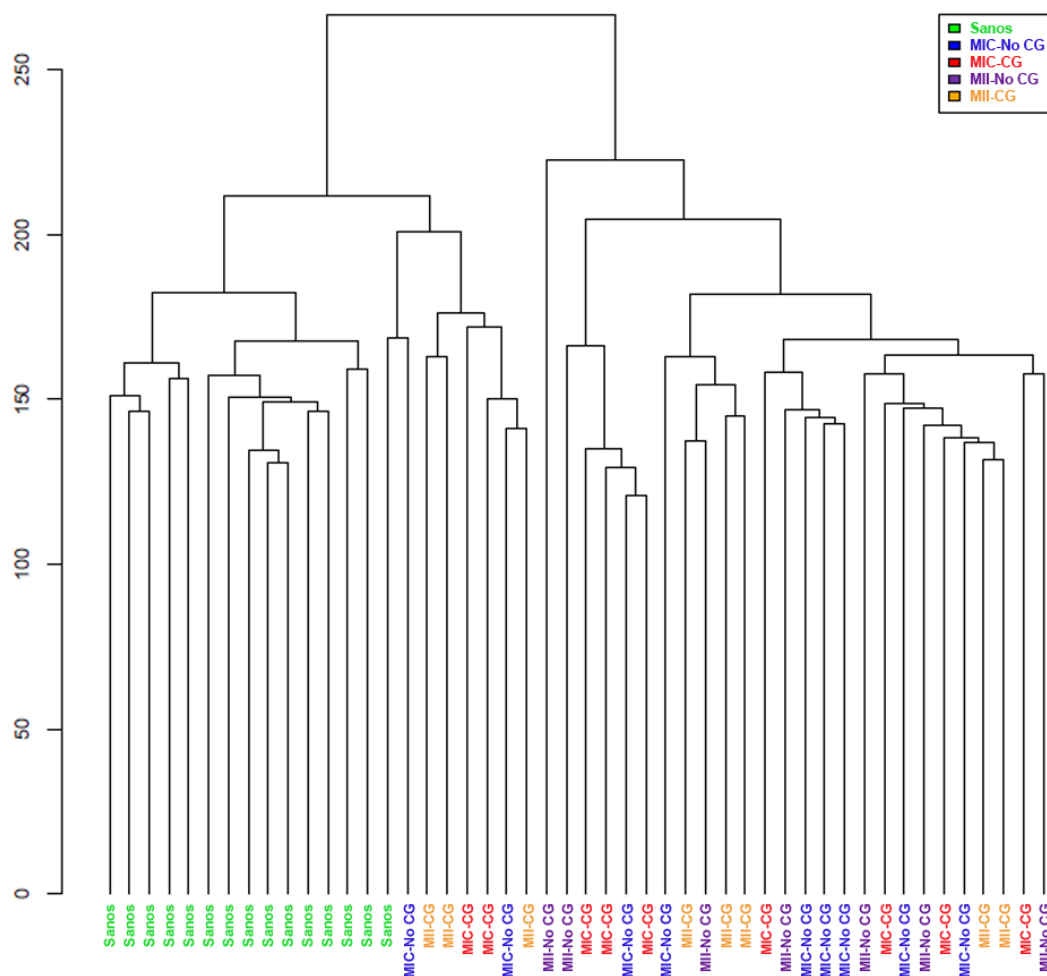
Tabla 31. Medición de cantidad y calidad de las muestras usadas en el microarray de expresión.											
Código	Diagnóstico al reclutamiento <sup>a</sup>	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Diagnóstico al final de seguimiento	Años de seguimiento <sup>c</sup>	Estatus <sup>d</sup>	Sexo <sup>e</sup>	Edad <sup>f</sup>	c(ARN) (ng/μl)	RIN	c(ADN) (ng/μl) <sup>g</sup>
A7	G4	E4	Antro	CG intestinal	10	MII-CG	M	52	198.7	2.4	3.5
A8	G4	E4	Antro	CG intestinal	8.4	MII-CG	H	69	199.5	2.5	6.8
A9	G4	E3	Antro	CG Difuso	7.7	MII-CG	H	45	254.8	2.4	7.9
A10	G4	E3	Antro	CG intestinal	2	MII-CG	H	73	188	2.4	5.6
A11	G4	E3	Antro	CG intestinal	1	MII-CG	H	60	303.9	2.1	10
A12	G4	E3	Antro	CG intestinal	11.8	MII-CG	M	66	157.8	2.6	8.6
A13	G3	E4	Antro	CG	13.4	MII-No CG	H	53	38.7	2.1	1.3
A14	G4	E4	Antro	MII	15	MII-No CG	H	64	53.6	2.3	4.8
A15	G3-G4	E3-E4	Antro	MII	11	MII-No CG	M	64	144	2.6	<2
A16	G3-G4	E4	Antro	MII	ND	MII-No CG	H	64	29.7	NA	1.8
A17	G3-G4	E3-E4	Antro	MII	14	MII-No CG	H	65	114.8	1.3	14
A18	G3-G4	E4	Antro	MII	ND	MII-No CG	M	74	114.6	1.5	<2
A19	G3-G4	E4	Antro	MII	ND	MII-No CG	H	71	136.1	2.1	8.7
A20	G1	E4	Antro	CG no cardias	2	MIC-CG	H	57	378	2.3	3.26
A21	G1	E4	Cuerpo	CG intestinal	8	MIC-CG	H	57	183	2.4	<2
A22	G1	E4	Antro	CG intestinal	1	MIC-CG	H	44	286	2.3	<2
A23	G1	E4	Antro	MIC	15.37	MIC-CG	H	56	69.3	2.4	6.4
A24	G2	E4	Antro	MIC	14	MIC-CG	M	59.74	62.9	2.2	8.14
A25	G1	E4	Antro	MIC	8.24	MIC-CG	M	57	215.5	2.1	9.65
A26	G1	E4	Antro	MIC	5.95	MIC-CG	M	78	119.6	2	12.3
A27	G1	E4	Antro	CG no cardias	6.2	MIC-CG	H	76	356.63	2.2	11.3
A28	G1	E4	Transicional	MIC	10.6	MIC-No CG	H	67	110.9	2.6	<2
A29	G2	E4	Antro	MIC	10.5	MIC-No CG	H	63	38.2	2.4	1.5
A30	G2	E4	Transicional	MIC	12.31	MIC-No CG	H	74	50.7	2.4	<2

Código	Diagnóstico al reclutamiento <sup>a</sup>	Extensión de MI al reclutamiento <sup>b</sup>	Localización anatómica	Diagnóstico al final de seguimiento	Años de seguimiento <sup>c</sup>	Estatus <sup>d</sup>	Sexo <sup>e</sup>	Edad <sup>f</sup>	c(ARN) (ng/μl)	RIN	c(ADN) (ng/μl) <sup>g</sup>
A31	G2	E4	Antro	MIC	13	MIC-No CG	H	56.98	104.9	2.4	<2
A32	G1	E4	Antro	MII	14	MIC-No CG	M	54	45.2	2.3	<2
A33	E4	G1	Antro	MII	15	MIC-No CG	H	57	98.6	2.5	12.5
A34	E4	G1	Antro	MIC	18	MIC-No CG	H	57	57.4	2.4	5.9
A35	E4	G1	Antro	MIC	14	MIC-No CG	H	62	67.7	2.2	6.3
A36	E4	G1	Antro	MIC	15	MIC-No CG	H	62	181	2.4	10.2
A37	-----	-----	Antro	-----	-----	Sanos	M	16	58	2.4	2.9
A38	-----	-----	Antro	-----	-----	Sanos	H	42	99.7	2.4	3.3
A39	-----	-----	Antro	-----	-----	Sanos	H	39	78.5	2	35.6
A40	-----	-----	Antro	-----	-----	Sanos	H	41	274	2.3	9.34
A41	-----	-----	Antro	-----	-----	Sanos	M	72	104	1.1	9.56
A42	-----	-----	Antro	-----	-----	Sanos	M	86	108	2.1	<2
A43	-----	-----	Antro	-----	-----	Sanos	M	18	77.4	2.2	<2
A44	-----	-----	Cuerpo	-----	-----	Sanos	H	27	139	2.3	3.79
A45	-----	-----	Antro	-----	-----	Sanos	H	68	111	1.6	3.25
A46	-----	-----	Antro	-----	-----	Sanos	M	31	244	2.2	12.3
A47	-----	-----	Antro	-----	-----	Sanos	M	19	476	2.4	5.47
A48	-----	-----	Cuerpo	-----	-----	Sanos	H	73	472	2.5	8.25
A49	-----	-----	Cuerpo	-----	-----	Sanos	M	58	409	2.4	3.96
A50	-----	-----	Antro	-----	-----	Sanos	M	11	286	2.3	2.9
A51	-----	-----	Cuerpo	-----	-----	Sanos	H	69	153	2.3	3.13

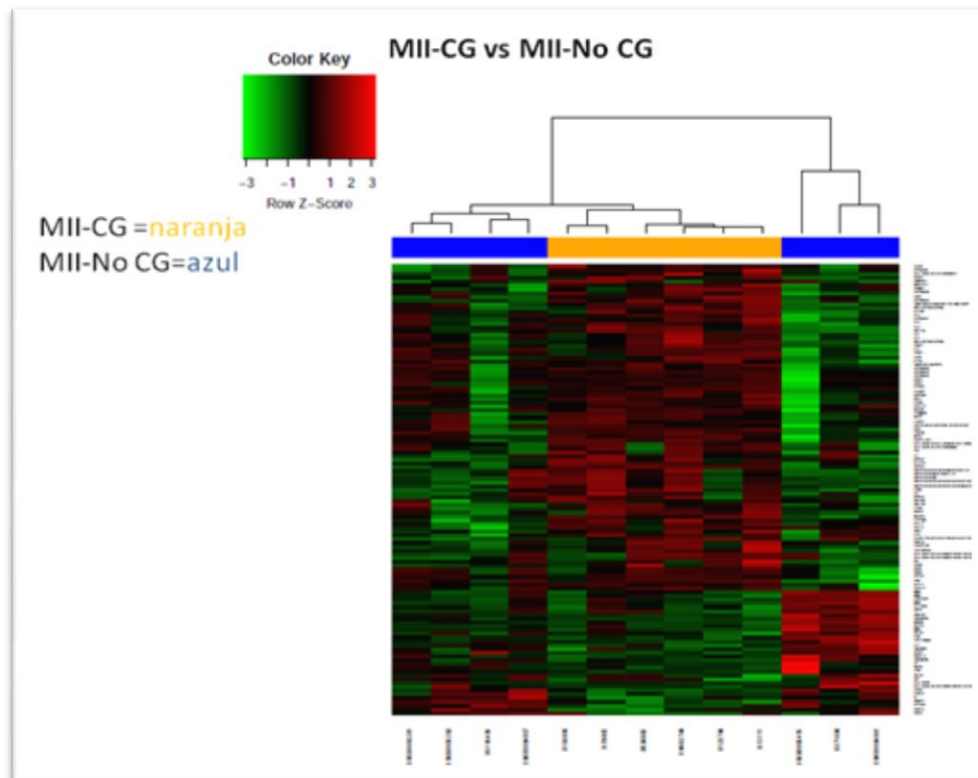
<sup>a</sup> Tipos histológicos de MI (G1-G4) descritos en apartado 3.1 de Materiales y Métodos. <sup>b</sup> E3>50%, E4>75%; Extensión de la MI en la muestra. En el caso de que existan dos diferentes en <sup>a</sup> ó <sup>b</sup> en la misma muestra se corresponden a los cortes inicial y final. <sup>c</sup> Se refiere al intervalo de tiempo entre el reclutamiento y el final de seguimiento. <sup>d</sup> MII/MIC-CG: Muestra de MII o de MIC al reclutamiento que progresa a CG al final del seguimiento. MII/MIC-No CG son controles de MII y MIC que no progresan a CG desde el reclutamiento hasta el final de seguimiento. <sup>e</sup> M: Mujer, H: Hombre. ND: No disponible. <sup>f</sup> Edad en el reclutamiento. <sup>g</sup> Concentración de ADN residual en la muestra de ARN.

### 3.3. Análisis estadísticos de los resultados del microarray

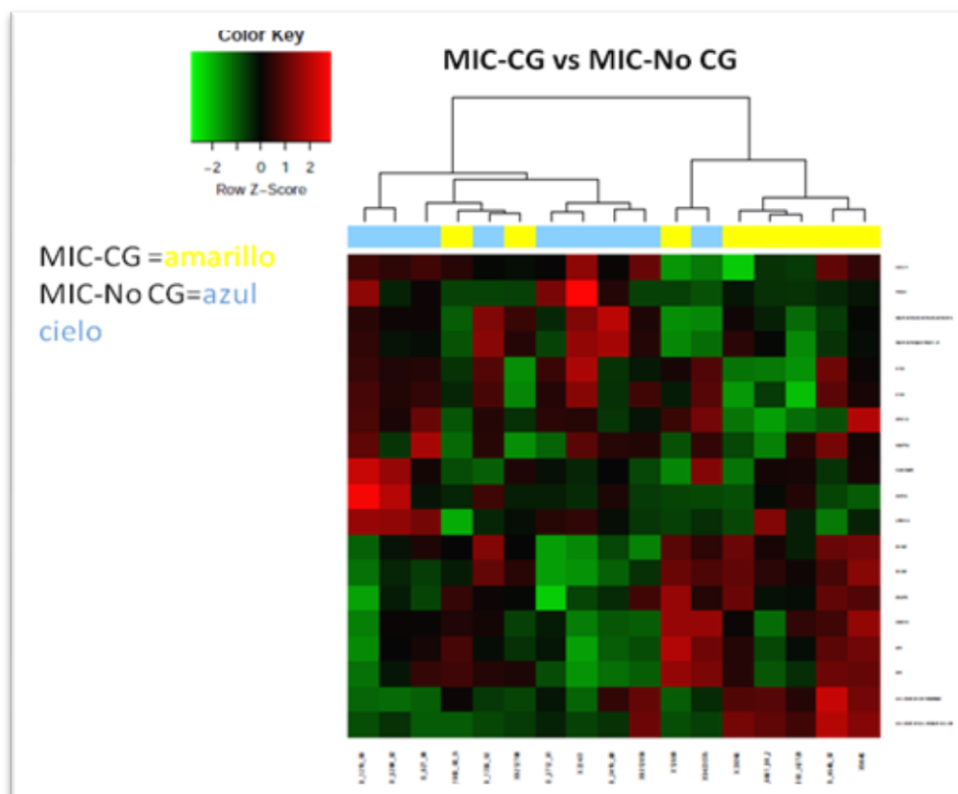
El dendrograma fue construido con los datos de expresión normalizados de cada muestra, a partir de los cuales se calculó y graficó la distancia euclídea en el eje de las ordenadas. El dendrograma muestra una evidente separación entre los controles de mucosa gástrica sana y las metaplasias intestinales, sin embargo no existe diferenciación entre las metaplasias intestinales que progresan a CG con respecto a las que no progresan, constituyendo un grupo indivisible (Figura 28). Un resultado similar se observa en los heatmaps de los grupos comparados (Figuras 29-31).



**Figura 28.** Dendrograma de todos los grupos analizados usando la distancia *Average* y método *Complete* en el programa informático *Combat*. Leyenda: Indica si la muestra progresa (MII-CG/MIC-CG) o no (MII-No CG/MIC-No CG) a CG, o si proviene de mucosa gástrica sana (Sanos).

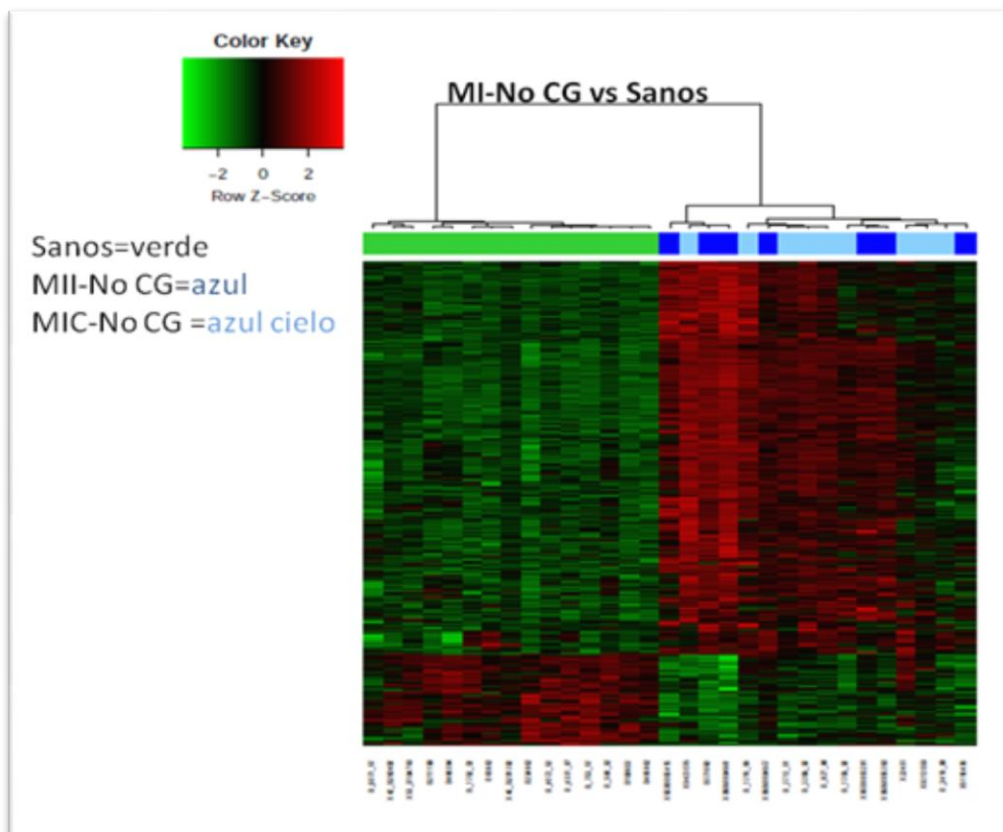


**Figura 29.** Heatmap de la comparación MII-CG vs MII-No CG.



**Figura 30.** Heatmap de la comparación MIC-CG vs MIC-No CG.





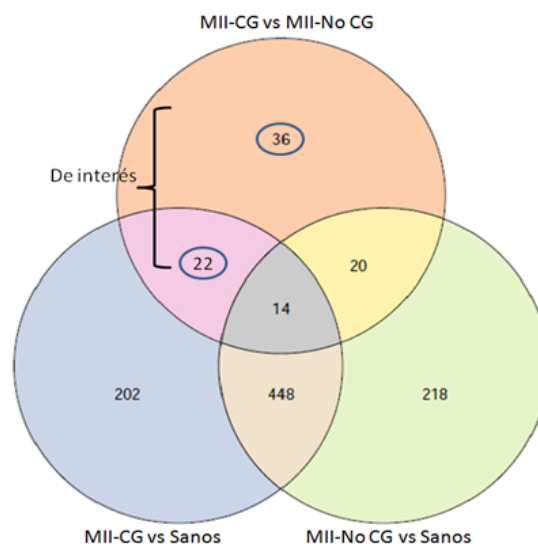
**Figura 31.** Heatmap de la comparación MI-No CG vs Sanos.

### 3.4. Genes diferencialmente expresados

Se analizaron los datos de expresión de 57 263 sondas con desviación estándar (DE)  $<0.34$  (30% de la DE global) de un total de 81 804, por lo que fueron eliminadas 24 541 sondas del análisis. El motivo de filtrar es eliminar las sondas que no se expresan suficientemente, produciendo “ruido” al hacer el análisis estadístico. El punto de corte escogido corresponde a las sondas que tienen una  $DE > 30\%$  del resto, y esto corresponde a una DE de 0.34.

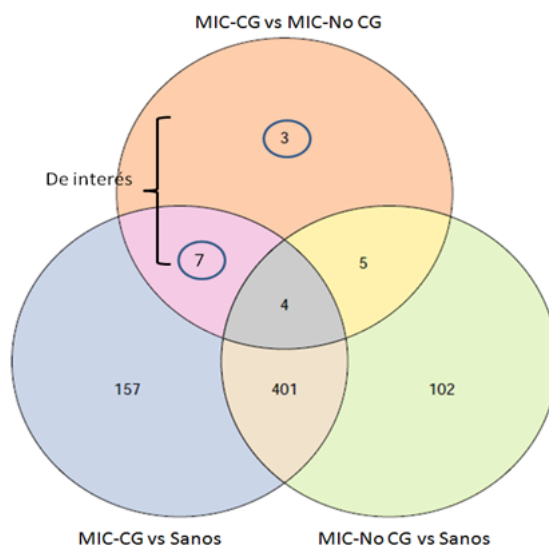
Los genes diferencialmente expresados entre los grupos analizados se reportan a continuación. En los casos en que el mismo gen fuera significativo más de una vez debido a sondas con la misma dirección de expresión (varias sobre-expresadas o sub-expresadas) solo fue tomada en cuenta la de mayor valor de *Fold Change*. En los casos de genes significativos más de una vez con un *Fold Change* de diferente sentido (*RHOA* y *KARS*) ambos se muestran y en los casos en que una misma sonda significativa hibrida en más de 1 transcrito (*FAM133B/FAM133CP/FAM133DP*), todos se tuvieron en cuenta. Se eliminó del análisis las sondas que no se encontraban anotadas a un gen específico o son marcos abiertos de lectura (Ej. *LOC100293977* y *C10orf116*). Los genes diferencialmente expresados se identificaron a partir de las siguientes comparaciones:

Metaplasia intestinal incompleta que progresa a cáncer gástrico (MII-CG) respecto a la que no progresa (MII-No CG): Se encontraron 106 genes con diferencias estadísticamente significativas por el p-valor nominal  $<0.05$ , no ajustado por FDR. De los cuales 82 (77.3%) están sobre-expresados ( $\text{Fold Change} > 2.0$ ) y 26 (22.7%) están sub-expresados ( $0.5 > \text{Fold Change} > 0.26$ ) en la MII-CG vs MII-No CG. El 54.7% de los genes desregulados ( $N=58$ ) también son significativos en la comparación MII-CG vs Sanos y no lo son en MII-No CG vs Sanos, siendo estos los candidatos de mayor interés (Figura 32, Tabla 32). Los genes *RHOA* y *KARS* presentan transcritos regulados tanto positiva como negativamente. Según el p-valor ajustado por FDR no se encuentran genes diferencialmente expresados. Los genes más sub-expresados en la MII-CG son *HLA-DRB1*, *RNU2-2*, *MME* y *MGAM*, mientras que los más sobre-expresados son *GKN1-2*, *CXCL17*, *HLA-C* y *PGC* (Tabla 32).



**Figura 32.** Diagrama de Venn que superpone los genes significativos en las comparaciones relativas a la MII.

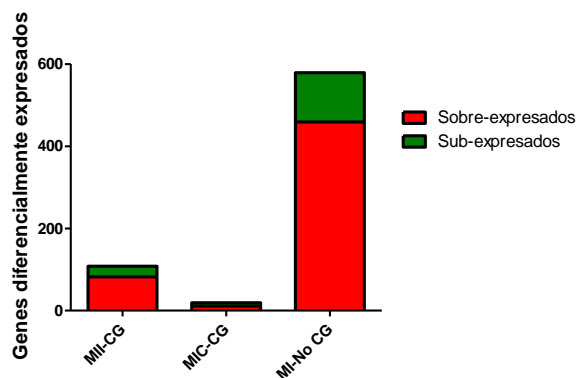
Metaplasia intestinal completa que progresa a cáncer gástrico (MIC-CG) respecto a la que no progresa (MIC-No CG): En esta comparación se encontraron 19 genes significativos ( $p < 0.05$ ), encontrándose la mayoría de los transcritos subexpresados ( $N=11$ ) en el grupo MIC-CG con respecto a MIC-No CG (Tabla 33). Los genes con mayor sub-expresión son *REG1B* y *GPR110*, mientras que los más sobre-expresados son *OLFM4*, *HLA-DRB1/HLA-DRB3/HLA-DRB5*. El 52.6% ( $N=10$ ) de los mismos también son significativos en la comparación MIC-CG vs Sanos y no en MIC-No CG vs Sanos, siendo los candidatos de mayor interés (Figura 33, Tabla 33). Los genes *IL1R2*, *CXCL17*, *IGHG1*, *IGHM* e *IGHV4-31* se encuentran sobre-expresados en la MII-CG mientras que están sub-expresados en la MIC-CG. Los genes *HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4* y *HLA-DRB5* se encuentran sobre-expresados en ambos grupos.



**Figura 33.** Diagrama de Venn que superpone los genes significativos en las comparaciones relativas a la MIC.

MI-CG vs MI-No CG: Al unir ambos subtipos histológicos de MI que progresan a CG y comparar con respecto a los que no progresan se observaron solamente 10 genes diferencialmente expresados, los cuales habían sido obtenidos en las comparaciones anteriores con igual dirección de expresión (Tabla 34).

MI-No CG vs Mucosa gástrica sana: En esta comparación se identificaron 579 genes con p-valor ajustado por FDR  $<0.05$ . Debido a la gran cantidad de resultados biológicamente relevantes solo mostramos los 188 transcritos con un *Fold Change*  $>3$  (sobre-expresados) o  $<0.33$  (sub-expresados) (Tabla anexa 2). Sin embargo es de destacar que existen 579 transcritos diferencialmente expresados en la MI-No CG, de los cuales 459 presentan un *Fold Change*  $\geq 1$  (sobre-expresados) y 120 un *Fold Change*  $<1$  (sub-expresados). Los genes con mayor sub-expresión en la MI-No CG son la gastrina, somatostatina, pepsinógeno y gastroquinas 1-2. Por su parte, los más sobre-expresados son *REG4*, *CLCA1*, *OLFM4*, *APOB*, *ANPEP*, *ALDOB*, *FABP1*, *PRDM10* y *DMBT1*. En todas las comparaciones realizadas predominan los genes sobre-expresados en los casos respecto a sus respectivos controles (Figura 34).



**Figura 34.** Número de genes sobre y sub-expresados en los grupos analizados.

Tabla 32. Genes diferencialmente expresados entre los grupos MII-CG vs MII-No CG.

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor <sup>c</sup>	Asociados con MII o CG <sup>d</sup>	Sign en MII-CG vs Sanos <sup>e</sup>	Sign en MII-No CG vs Sanos <sup>f</sup>	Candidatos de mayor interés <sup>g</sup>
<i>HLA-DRB1</i>	major histocompatibility complex, class II, DR beta 1	0.260	0.014	SI <sup>195</sup>	SI	NO	SI
<i>RNU2-2</i>	RNA, U2 small nuclear 2	0.297	0.008	Nueva molécula	SI	NO	SI
<i>MME</i>	membrane metallo-endopeptidase	0.363	0.007	Nueva molécula	NO	SI	NO
<i>MGAM</i>	maltase-glucoamylase (alpha-glucosidase)	0.374	0.000	SI <sup>196</sup>	NO	SI	NO
<i>IK</i>	IK cytokine, down-regulator of HLA II	0.384	0.002	SI <sup>192</sup>	SI	NO	SI
<i>TMEM25</i>	transmembrane protein 25	0.386	0.005	Nueva molécula	NO	SI	NO
<i>GSTA1</i>	glutathione S-transferase alpha 1	0.386	0.029	SI <sup>197</sup>	NO	SI	NO
<i>PPIA</i>	Peptidylprolyl isomerase A (cyclophilin A)	0.393	0.027	SI <sup>198</sup>	SI	NO	SI
<i>KARS</i>	lysyl-tRNA synthetase	0.396	0.032	Nueva molécula	NO	NO	SI
<i>RHOA</i>	ras homolog family member A	0.398	0.000	SI <sup>199</sup>	NO	SI	NO
<i>CDKL3</i>	cyclin-dependent kinase-like 3	0.403	0.000	Nueva molécula	SI	NO	SI
<i>TMEM236</i>	transmembrane protein 236	0.409	0.001	Nueva molécula	NO	SI	NO
<i>RBP2</i>	retinol binding protein 2, cellular	0.411	0.021	SI <sup>200</sup>	SI	SI	NO
<i>CP</i>	ceruloplasmin (ferroxidase)	0.412	0.040	SI <sup>201</sup>	SI	NO	SI
<i>GIP</i>	gastric inhibitory polypeptide	0.433	0.017	Nueva molécula	SI	SI	NO
<i>XPNPEP2</i>	X-prolyl aminopeptidase (aminopeptidase P) 2,	0.442	0.020	Nueva molécula	SI	SI	NO
<i>GNAS</i>	GNAS complex locus	0.445	0.009	SI <sup>202</sup>	NO	NO	SI
<i>SART3</i>	squamous cell carcinoma antigen recognized by T	0.448	0.000	SI <sup>203</sup>	NO	NO	SI
<i>RBBP7</i>	retinoblastoma binding protein 7	0.449	0.001	Nueva molécula	SI	NO	SI
<i>AQP10</i>	aquaporin 10	0.465	0.014	Nueva molécula	NO	SI	NO
<i>SLC13A2</i>	solute carrier family 13 (sodium-dependent)	0.470	0.007	Nueva molécula	SI	SI	NO
<i>DNAJB7</i>	DnaJ (Hsp40) homolog, subfamily B, member 7	0.482	0.001	Nueva molécula	NO	SI	NO
<i>XDH</i>	xanthine dehydrogenase	0.486	0.006	Nueva molécula	SI	SI	NO
<i>DTNB</i>	dystrobrevin, beta	0.487	0.001	Nueva molécula	NO	NO	SI
<i>LCT</i>	lactase	0.493	0.002	Nueva molécula	NO	NO	SI

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor <sup>c</sup>	Asociados con MI o CG <sup>d</sup>	Sign en MII-CG vs Sanos <sup>e</sup>	Sign en MII-No CG vs Sanos <sup>f</sup>	Candidatos de mayor interés <sup>g</sup>
<i>ATP1B3</i>	ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, beta 3 polypeptide	0.499	0.002	Nueva molécula	SI	SI	NO
<i>FAM133B</i> / <i>FAM133CP</i> / <i>FAM133</i>	family with sequence similarity 133, member B	2.001	0.004	Nueva molécula	NO	NO	SI
<i>MORF4L1</i>	mortality factor 4 like 1	2.006	0.016	Nueva molécula	NO	NO	SI
<i>CXCL14</i>	chemokine (C-X-C motif) ligand 14	2.007	0.028	SI <sup>204</sup>	NO	NO	SI
<i>LAPTM5</i>	lysosomal protein transmembrane 5	2.011	0.010	Nueva molécula	NO	NO	SI
<i>NUCB2</i>	nucleobindin 2	2.013	0.003	SI <sup>205</sup>	NO	SI	NO
<i>RGS16</i>	regulator of G-protein signaling 16	2.022	0.000	Nueva molécula	NO	NO	SI
<i>HIST2H3A</i> / <i>HIST2H3C</i> / <i>HIST2H3D</i>	histone cluster 2, H3a, H3c, H3d	2.024	0.048	Nueva molécula	NO	NO	SI
<i>MYOF</i>	myoferlin	2.035	0.006	Nueva molécula	SI	NO	SI
<i>C1R</i>	complement component 1, r subcomponent	2.048	0.005	Nueva molécula	NO	NO	SI
<i>CCT6A</i>	chaperonin containing TCP1, subunit 6A (zeta 1)	2.053	0.002	Nueva molécula	NO	NO	SI
<i>ABCC5</i>	ATP-binding cassette, sub-family C (CFTR/MRP),	2.056	0.013	SI <sup>206</sup>	NO	SI	NO
<i>HLA-A</i>	major histocompatibility complex, class I, A	2.059	0.029	SI <sup>207</sup>	SI	NO	SI
<i>RPS6KA2</i>	ribosomal protein S6 kinase, 90kDa, polypeptide 2	2.059	0.001	Nueva molécula	NO	NO	SI
<i>EIF5B</i>	eukaryotic translation initiation factor 5B	2.059	0.004	Nueva molécula	NO	NO	SI
<i>C1QBP</i>	complement component 1, q subcomponent binding	2.076	0.002	Nueva molécula	SI	NO	SI
<i>CTSB</i>	cathepsin B	2.078	0.001	SI <sup>208</sup>	NO	NO	SI
<i>IGKC</i>	immunoglobulin kappa constant	2.089	0.014	SI <sup>209</sup>	SI	SI	NO
<i>UHM1K1</i>	U2AF homology motif (UHM) kinase 1	2.092	0.012	Nueva molécula	NO	NO	SI
<i>NOL7</i>	nucleolar protein 7, 27kDa	2.095	0.001	Nueva molécula	NO	NO	SI
<i>MEGF6</i>	multiple EGF-like-domains 6	2.104	0.000	Nueva molécula	NO	SI	NO
<i>SMNDC1</i>	survival motor neuron domain containing 1	2.104	0.013	Nueva molécula	SI	NO	SI
<i>ZNF259</i>	zinc finger protein 259	2.110	0.022	Nueva molécula	SI	NO	SI
<i>CANX</i>	calnexin	2.117	0.027	Nueva molécula	SI	NO	SI

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor <sup>c</sup>	Asociados con MI o CG <sup>d</sup>	Sign en MII-CG vs Sanos <sup>e</sup>	Sign en MII-No CG vs Sanos <sup>f</sup>	Candidatos de mayor interés <sup>g</sup>
<i>TOMM22</i>	translocase of outer mitochondrial membrane 22	2.120	0.012	Nueva molécula	SI	NO	SI
<i>CAV1</i>	caveolin 1, caveolae protein, 22kDa	2.144	0.000	SI <sup>210</sup>	NO	NO	SI
<i>CCDC144A</i>	coiled-coil domain containing 144A-C	2.148	0.014	Nueva molécula	NO	NO	SI
<i>ANP32B</i>	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	2.155	0.006	Nueva molécula	NO	NO	SI
<i>RHOA</i>	ras homolog family member A	2.16	0.047	SI <sup>199</sup>	NO	SI	NO
<i>AFAP1-AS1</i>	AFAP1 antisense RNA 1 (non-protein coding)	2.161	0.001	Nueva molécula	SI	SI	NO
<i>HLA-DQB1</i>	major histocompatibility complex, class II, DQ beta 1	2.17	0.001	SI <sup>211</sup>	SI	NO	SI
<i>PI3</i>	peptidase inhibitor 3, skin-derived	2.183	0.012	Nueva molécula	SI	SI	NO
<i>CD53</i>	CD53 molecule	2.17	0.000	Nueva molécula	SI	NO	SI
<i>NHP2</i>	NHP2 ribonucleoprotein homolog (yeast)	2.187	0.002	Nueva molécula	NO	NO	SI
<i>C3</i>	complement component 3	2.193	0.020	Nueva molécula	NO	NO	SI
<i>CTBP2</i>	C-terminal binding protein 2	2.204	0.007	Nueva molécula	NO	NO	SI
<i>CRISPLD2</i>	cysteine-rich secretory protein LCCL domain	2.208	0.003	Nueva molécula	NO	NO	SI
<i>EIF3D</i>	eukaryotic translation initiation factor 3, subunit D	2.211	0.003	Nueva molécula	NO	NO	SI
<i>GNL3/SNORD19</i>	guanine nucleotide binding protein-like 3	2.218	0.000	Nueva molécula	NO	NO	SI
<i>CD24</i>	CD24 molecule	2.249	0.008	SI <sup>212</sup>	SI	NO	SI
<i>HLA-DRB4</i>	major histocompatibility complex, class II, DR beta 4	2.25	0.024	SI <sup>195</sup>	NO	NO	SI
<i>ZNF146</i>	zinc finger protein 146	2.264	0.000	Nueva molécula	SI	NO	SI
<i>PDCD5</i>	programmed cell death 5	2.266	0.001	SI <sup>213</sup>	NO	NO	SI
<i>VSIG1</i>	V-set and immunoglobulin domain containing 1	2.28	0.012	SI <sup>214</sup>	NO	SI	NO
<i>SLC1A4</i>	solute carrier family 1 (glutamate/neutral amino)	2.307	0.001	Nueva molécula	SI	NO	SI
<i>IL1R2</i>	interleukin 1 receptor, type II	2.31	0.019	SI <sup>103</sup>	NO	NO	SI
<i>LYZ</i>	lysozyme	2.32	0.042	SI <sup>81</sup>	NO	SI	NO
<i>FABP5 /FABP5P3</i>	fatty acid binding protein 5 (psoriasis-associated)	2.369	0.003	SI <sup>214</sup>	NO	SI	NO
<i>RAN</i>	RAN, member RAS oncogene family	2.41	0.004	SI <sup>215</sup>	NO	NO	SI

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor <sup>c</sup>	Asociados con MI o CG <sup>d</sup>	Sign en MII-CG vs Sanos <sup>e</sup>	Sign en MII-No CG vs Sanos <sup>f</sup>	Candidatos de mayor interés <sup>g</sup>
<i>HSP90AA1</i>	heat shock protein 90kDa alpha (cytosolic), class A	2.412	0.000	SI <sup>216</sup>	NO	NO	SI
<i>CLU</i>	clusterin	2.469	0.016	SI <sup>217</sup>	SI	SI	NO
<i>ATP5A1</i>	ATP synthase, H+ transporting, mitochondrial F1	1.304	0.016	Nueva molécula	SI	NO	SI
<i>HSP90AB1</i>	heat shock protein 90kDa alpha (cytosolic), class B	2.479	0.003	SI <sup>216</sup>	NO	NO	SI
<i>ATP6V0E1</i>	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit	2.488	0.005	Nueva molécula	NO	NO	SI
<i>GIF</i>	gastric intrinsic factor (vitamin B synthesis)	2.519	0.049	SI <sup>218</sup>	NO	SI	NO
<i>KARS</i>	lysyl-tRNA synthetase	2.542	0.019	Nueva molécula	NO	NO	SI
<i>IGHG1/IGHG2/IGHG3/IGHG4/IGHM/IGHV4-31</i>	immunoglobulin heavy constant gamma 1 in heavy constant gamma	2.670	0.011	SI <sup>209</sup>	SI	SI	NO
<i>BASP1</i>	brain abundant, membrane attached signal protein	2.704	0.005	Nueva molécula	NO	SI	NO
<i>BPIFB1</i>	BPI fold containing family B, member 1	2.866	0.009	Nueva molécula	NO	NO	SI
<i>DPCR1</i>	diffuse panbronchiolitis critical region 1	1.519	0.024	SI	NO	SI	NO
<i>HLA-DQA1</i>	major histocompatibility complex, class II, DQalpha1	3.182	0.024	SI <sup>219</sup>	SI	NO	SI
<i>HLA-DRB1/HLA-DRB3/HLA-DRB5</i>	major histocompatibility complex, class II, DR beta complex	3.368	0.048	SI <sup>195</sup>	SI	NO	SI
<i>LTF</i>	lactotransferrin	3.432	0.019	SI <sup>220</sup>	NO	SI	NO
<i>GKN2</i>	gastrokine 2	3.499	0.025	SI <sup>221</sup>	SI	SI	NO
<i>CXCL17</i>	chemokine (C-X-C motif) ligand 17	3.506	0.002	Nueva molécula	NO	SI	NO
<i>GKN1</i>	gastrokine 1	3.834	0.038	SI <sup>222</sup>	SI	SI	NO
<i>HLA-C</i>	major histocompatibility complex, class I, C	4.417	0.011	SI <sup>223</sup>	NO	NO	SI
<i>PGC</i>	progastricsin (pepsinogen C)	5.004	0.002	SI <sup>224</sup>	SI	SI	NO

<sup>a</sup> En el caso de la existencia de más de 1 gen significa que la sonda reconoce varios transcritos. <sup>b</sup> *Fold change* es el promedio de expresión en MII-CG dividido entre MII-No CG. Los genes están ordenados de menor a mayor por dicho valor. <sup>c</sup> p-valor resultante del t test moderado sin ser ajustado por comparaciones múltiples. <sup>d</sup> Genes asociados con MI o CG en estudios de expresión génica, proteica, de asociación genética, genómicos o funcionales. <sup>e</sup> Significación estadística en la comparación MII-CG vs Sanos. <sup>f</sup> Significación estadística en la comparación MII-No CG vs Sanos. <sup>g</sup> Se refiere a aquellos genes de interés del diagrama de Venn de la Figura 32.

Tabla 33. Genes diferencialmente expresados entre los grupos MIC-CG vs MIC-No CG.							
Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor <sup>c</sup>	Asociados con MI o CG <sup>d</sup>	Sign en MIC-CG vs Sanos <sup>e</sup>	Sign en MIC-No CG vs Sanos <sup>f</sup>	Candidatos de mayor interés <sup>g</sup>
REG1B	regenerating islet-derived 1 beta	0.421	0.011	SI <sup>214</sup>	NO	SI	NO
GPR110	G protein-coupled receptor 110	0.440	0.015	Nueva molécula	NO	SI	NO
CEACAM20	Carcinoembryonic antigen-related cell adhesion molecule 20	0.461	0.046	Nueva molécula similar <sup>225</sup>	SI	SI	NO
ANAPC5	anaphase promoting complex subunit 5	0.461	0.031	Nueva molécula	NO	NO	SI
EPB41L3	erythrocyte membrane protein band 4.1-like 3	0.474	0.014	SI <sup>226</sup>	NO	SI	NO
IL1R2	interleukin 1 receptor, type II	0.475	0.017	SI <sup>102</sup>	SI	NO	SI
CXCL17	chemokine (C-X-C motif) ligand 17	0.484	0.038	Nueva molécula	SI	SI	NO
IGHG1 / IGHM / IGHV4-31	immunoglobulin heavy constant gamma 1-4-31	0.493	0.014	SI <sup>209</sup>	SI	NO	SI
PRSS1	protease, serine, 1 (trypsin 1)	0.495	0.033	Nueva molécula	NO	NO	SI
HLA-DRB4	major histocompatibility complex, class II, DR	2.049	0.022	SI <sup>195</sup>	NO	NO	SI
GP2	glycoprotein 2 (zymogen granule membrane)	2.063	0.018	Nueva molécula	SI	NO	SI
HOXA13	homeobox A13	2.076	0.028	SI <sup>264</sup>	SI	NO	SI
IGFBP5	insulin-like growth factor binding protein 5	2.127	0.004	Nueva molécula	SI	NO	SI
OLFMD4	olfactomedin 4	3.332	0.0002	SI <sup>227</sup>	SI	NO	SI
HLA-DRB1 / HLA-DRB3 / HLA-DRB5	major histocompatibility complex, class II, DR beta 1-5	3.475	0.021	SI <sup>195</sup>	SI	NO	SI

<sup>a</sup> En el caso de la existencia de más de 1 gen significa que la sonda reconoce varios transcritos. <sup>b</sup> Fold change es el promedio de expresión en MIC-CG dividido entre MIC-No CG. Los genes están ordenados de menor a mayor por dicho valor. <sup>c</sup> p-valor resultante del t test moderado sin ser ajustado por comparaciones múltiples. <sup>d</sup> Genes asociados con MI o CG en estudios de expresión génica, proteica, de asociación genética, genómicos o funcionales <sup>e</sup> Significación estadística en la comparación MIC-CG vs Sanos. <sup>f</sup> Significación estadística en la comparación MIC-No CG vs Sanos. <sup>g</sup> Se refiere a aquellos genes de interés del diagrama de Venn de la Figura 33.



Tabla 34. Genes diferencialmente expresados entre los grupos MI-CG vs MI-No CG.				
Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor <sup>c</sup>	Asociados con MI o CG <sup>d</sup>
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	0.369	0.0052	SI <sup>195</sup>
CEACAM20	Carcinoembryonic antigen-related cell adhesion molecule 20	0.457	0.0071	Nueva molécula similar <sup>225</sup>
REG1B	regenerating islet-derived 1 beta	0.463	0.0026	SI <sup>214</sup>
RBP2	retinol binding protein 2, cellular	0.480	0.0038	SI <sup>200</sup>
HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1	2.028	0.0005	SI <sup>216</sup>
HLA-DRB4	major histocompatibility complex, class II, DR beta 4	1.09	0.0014	SI <sup>195</sup>
OLFM4	olfactomedin 4	1.10	0.002	SI <sup>227</sup>
HLA-DRB1/HLA-DRB3/HLA-DRB5	major histocompatibility complex, class II, DR beta 1-5	1.77	0.0025	SI <sup>195</sup>

<sup>a</sup> En el caso de la existencia de más de 1 gen significa que la sonda reconoce varios transcritos. <sup>b</sup> *Fold change* es el promedio de expresión en MI-CG dividido entre MI-No CG. Los genes están ordenados de menor a mayor por dicho valor. <sup>c</sup> p-valor resultante del t test moderado sin ser ajustado por comparaciones múltiples. <sup>d</sup> Genes asociados con MI o CG en estudios de expresión génica, proteica, de asociación genética, genómicos o funcionales.

### 3.5. Análisis de enriquecimiento funcional

#### 3.5.1. Gene Set Enrichment Analysis

A partir de los valores de expresión génica normalizados de todos los genes se obtuvieron los conjuntos génicos (*genesets*) diferencialmente expresados en los grupos que progresan a CG respecto a los que no progresan o a la mucosa sana, comparando con los catálogos c2All.v.5 y c3.tft.v.5 de MiSigDB (apartado 3.7, Materiales y Métodos). Solo se seleccionaron los conjuntos génicos con plausibilidad biológica en la MI o la carcinogénesis gástrica; lo cual significa que en algunos casos reflejan procesos moleculares previamente validados a nivel experimental, como el rol de los oncogenes en la carcinogénesis <sup>45</sup>.

Mediante la comparación con la base de datos c2All v5.0 se obtuvieron 144 conjuntos génicos significativamente ( $p$ -valor nominal  $<0.01$  y  $q$ -valor FDR  $<0.05$ ) enriquecidos en el grupo MII-CG con respecto al MII-No CG, los cuales poseen valores de *Enrichment Score* en el rango de 0.337-0.646 y están compuestos por entre 16 y 491 genes (Tabla anexa 3). Sin embargo, en el grupo MIC-No CG solo se encontraron dos conjuntos génicos significativamente sobre-expresados, los cuales carecen de interés en la patología estudiada, estos son: *reactome\_olfactory\_signaling\_pathway* (ES -0.497) y *kegg\_olfactory\_transduction* (ES -0.448).

En el grupo MII-CG la mayor cantidad de conjuntos génicos sobre-expresados están involucrados en procesos del ciclo y la proliferación celular (*PI3* y MAP quinasas, *WNT*, *TGFB1* y factores de transcripción *E2F*), oncogenes (*MYC*, *RAS*, *RHOA* y *MET*), supresores tumorales (*TP53*, *RB1* y *BRCA1/2*), procesos de invasión y metastasis, e inflamación. Mientras que otros procesos menos frecuentes son la adipogénesis, presentación y procesamiento antigénico, apoptosis, angiogénesis, fosforilación oxidativa, adhesión celular, degradación proteasomal, señalización insulínica, respuesta al daño genómico, endocitosis y fagocitosis, cáncer de esófago, degradación de ARNm por mecanismo NMD, respuesta a proteínas con incorrecto plegamiento y cáncer gástrico (Tabla anexa 3). Con el objetivo de acotar esta variedad de conjuntos génicos a los más relevantes se realizaron varios análisis. Primero se eligieron los que poseen los mayores valores de enriquecimiento (ES  $>0.5$ ), representados por genes regulados por insulina, degradación proteica por el proteasoma, oncogenes, adhesión celular, metabolismo lipídico, angiogénesis, supresores tumorales, ciclo y proliferación celular, respuesta al daño genómico, apoptosis, endocitosis o fagocitosis. Además, se eligieron los conjuntos con los valores más extremos (mayores o menores) del parámetro *Rank at max*, el cual representa la posición en la lista ranqueada de conjuntos en que se alcanza el valor máximo de ES (Tabla 35) y un análisis de genes *leading edge* fue realizado.

Cuando ambos tipos de metaplasias que no progresan a CG fueron agrupadas y realizada una comparación respecto a la mucosa sana (MI-No CG vs Sanos) se obtuvieron 103 conjuntos significativamente sobre-expresados en el grupo MI-No CG (p-valor nominal  $<0.05$  y q-valor FDR  $<0.25$ ). Los mismos presentan valores de ES entre 0.379 y 0.843, y se encuentran compuestos por entre 15 y 407 genes (Tabla anexa 4). Los conjuntos génicos sobre-expresados más frecuentemente observados representan a procesos fisiológicos alterados como un efecto Warburg no tumoral, metabolismo lipídico, la inflamación, diferenciación intestinal y presentación y procesamiento antigénico. Mientras que los menos frecuentes fueron la apoptosis, oncogenes, proliferación celular, respuesta al daño genómico, supresores tumorales, metabolismo de xenobióticos, invasión y metastasis, angiogénesis, cáncer gástrico, reflujo gastroesofágico, glicosilación proteica aberrante, infección por *H. pylori*, respuesta a proteínas con incorrecto plegamiento y fosforilación oxidativa. No existe ninguna vía sub-expresada en la MI-No CG en comparación con los sujetos sanos.

Los conjuntos génicos con mayor enriquecimiento (ES $>0.65$ ) en el grupo MI-No CG pertenecen al metabolismo lipídico (ES 0.843, 0.748, 0.737, 0.683, 0.665), procesamiento antigénico (0.789), diferenciación intestinal (ES 0.744, 0.738, 0.655), efecto Warburg no tumoral (ES 0.724, 0.666, 0.651), apoptosis (0.718, 0.682), glicosilación proteica aberrante (ES 0.716), respuesta ante el daño genómico (ES 0.682, 0.671), metabolismo de xenobióticos (ES 0.673), infección por *H. pylori* (ES 0.669), inflamación (ES 0.662) y respuesta a proteínas con plegamiento aberrante (ES 0.667) (Tabla anexa 4). Los conjuntos génicos con los valores más extremos de *Rank at max* se reportan en la tabla 36.

La inspección de los resultados de c3.tft.v5.0 en la MI-No CG muestra que existen 12 conjuntos génicos sobre-expresados de los factores de transcripción *HNF1/4* y *GATA1/6*, los cuales cooperan en la activación de la expresión de genes intestinales en la MI<sup>64</sup>.

Con el objetivo de validar nuestros resultados con el GSEA respecto a otros antes publicados, datos de expresión génica por microarray depositados en GEO (<http://www.ncbi.nlm.nih.gov/gds>, GSE47797), de metaplasia intestinal y mucosa gástrica sana fueron comparados<sup>186</sup>. En concordancia con nuestros resultados los principales procesos biológicos sobre-expresados en la metaplasia intestinal usando el catálogo c2All.v.5 fueron el efecto Warburg no tumoral (N=17), metabolismo lipídico (N=15), diferenciación intestinal (N=6), ciclo y proliferación celular (N=5), inflamación (N=5), supresores tumorales (N=4), cáncer gástrico (N=3), glicosilación proteica aberrante (N=2), apoptosis (N=2), metabolismo de xenobióticos (N=2) y respuesta al daño genómico (N=2). En el caso de C3All.v5.0 encontramos 6 conjuntos génicos sobre-expresados compuestos por factores

transcripcionales nucleares hepáticos *HNF1A* y *HNF4A*. En ningún caso (C2All.v5.0, C3All.v5.0) existían conjuntos génicos significativos en la mucosa sana.

### **Análisis de genes *leading edge* en la MII-CG y la MI-No CG**

Otro análisis realizado para obtener los conjuntos génicos más relevantes sobre-expresados en la MII-CG y MI-No CG fue la identificación de sus genes *leading edge* (responsables de la señal de enriquecimiento), y se inspeccionó si entre los mismos figuran al menos 3 genes sobre-expresados individualmente en la MII-CG o MI-No CG (Tablas 37 y 38).

Otro tipo de resultado que se puede obtener a partir del análisis *leading edge* es el de identificar cuales conjuntos génicos sobre-expresados en la MII-CG comparten genes reflejando una función biológica común. En este sentido encontramos dos núcleos (1 y 2) que comparten genes como evidencia la acumulación de color verde de la figura 35. A continuación, los conjuntos génicos de dichos núcleos fueron seleccionados y realizado nuevamente el mismo análisis *leading edge* (Figura 36) para visualizar mejor a los mismos e identificar un proceso fisiológico común y los genes *leading edge* responsables. Este análisis permitió conocer que el núcleo 1 está formado por conjuntos génicos con diversas funciones como regulación del ciclo celular (N=5), procesamiento antigénico (N=5), apoptosis (N=2), degradación proteasomal (N=3), fagocitosis (N=1), activación de la vía de señalización NFKB (N=1), respuesta al daño del ADN mediada por TP53 (N=2) y autodegradación de CDH1 (N=1) (Figura 36).

Los genes *leading edge* *PSMA1/2/3/4/7*, *PSMB2/5*, *PSMC3/6*, *PSMD8/11/14* comunes y sobre-expresados en todos los conjuntos codifican para componentes del proteasoma. Otros genes también del proteasoma como *PSMB1/3/4/7/9*, *PSMC1/4/5*, *PSMD1/2/6/10/13*, *PSME1/4* y *UBA52* son compartidos por entre el 90 y el 95% de todos los conjuntos (Figura 37). Se observa además un pequeño núcleo (2) formado por conjuntos con funciones implicadas en la fosforilación oxidativa (Figura 38), siendo los genes *leading edge* compartidos *COX5A/6C/7A2L/7B/7C*, *CYC1*, *NDUFA1/4-8*, *NDUFAB1*, *NDUFB1/2/4/6/8*, *NDUFC1*, *NDUFS3-6*, *NDUFV2*, *UQCR11* y *UQCRC1*. No se observó ningún núcleo con genes *leading edge* compartidos en la MI-No CG.

Tabla 35. Conjuntos génicos sobre-expresados en Mill-CG con los valores más extremos del parámetro Rank at max.						
Conjuntos génicos	Proceso funcional	ES	p-valor nominal <sup>a</sup>	q-valor FDR <sup>b</sup>	Rank at max <sup>c</sup>	
REACTOME_G_BETA_GAMMA_SIGNALING_THROUGH_PI3KGAMMA	Ciclo y Proliferación celular	0,593	0,007	0,024	895	
BERENIENO_TRANSFORMED_BY_RHOA_FOREVER_DN	Oncogenes	0,627	0,000	0,002	1908	
YU_MYC_TARGETS_UP	Oncogenes	0,512	0,003	0,016	2259	
SCHOEN_NFKB_SIGNALING	Inflamación	0,517	0,002	0,027	2552	
PEDERSEN_METASTASIS_BY_ERBB2(her2)_ISOFORM_1	Invasión y metástasis	0,489	0,005	0,027	2804	
CROONQUIST_NRAS_SIGNALING_UP	Oncogenes	0,589	0,000	0,003	2872	
SCIAN_CELL_CYCLE_TARGETS_OF_TP53_AND_TP73_DN	Supresores tumorales	0,598	0,004	0,018	2879	
SARTIPY_NORMAL_AT_INSULIN_RESISTANCE_UP	Genes regulados por insulina	0,646	0,000	0,001	2898	
ASTIER_INTEGRIN_SIGNALING	Ciclo y Proliferación celular	0,449	0,003	0,031	3343	
PEDERSEN_METASTASIS_BY_ERBB2_ISOFORM_4	Invasión y metástasis	0,419	0,000	0,023	3368	
REN_BOUND_BY_E2F	Ciclo y Proliferación celular	0,571	0,000	0,001	3374	
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_6HR	Respuesta a daño genómico	0,423	0,003	0,032	3429	
MARKEY_RB1_ACUTE_LOF_UP	Supresores tumorales	0,369	0,000	0,037	3592	
MARZEC_IL2_SIGNALING_UP	Inflamación	0,437	0,000	0,016	3796	
PID_INTEGRIN5_PATHWAY	Adhesión celular	0,631	0,003	0,018	3850	
LU_TUMOR_ANGIOGENESIS_UP	Angiogénesis	0,564	0,002	0,034	3855	
PHONG_TNF_RESPONSE_VIA_P38_PARTIAL	Inflamación	0,366	0,002	0,050	3878	
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_24HR	Respuesta a daño genómico	0,429	0,000	0,014	3976	
WANG_TUMOR_INVASIVENESS_UP	Invasión y metástasis	0,420	0,000	0,002	6767	
ST_INTEGRIN_SIGNALING_PATHWAY	Adhesión celular	0,398	0,006	0,048	6779	
GOTZMANN_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_DN	Invasión y metástasis	0,375	0,000	0,026	6789	
REACTOME_CELL_CYCLE_MITOTIC	Ciclo y Proliferación celular	0,382	0,000	0,011	6864	
KEGG_CELL_CYCLE	Ciclo y Proliferación celular	0,419	0,000	0,018	6872	

Conjuntos génicos	Proceso funcional	ES	p-valor nominal <sup>a</sup>	q-valor FDR <sup>b</sup>	Rank at max <sup>c</sup>
THEILGAARD_NEUTROPHIL_AT_SKIN_WOUND_DN	Inflamación	0,379	0,000	0,019	6872
WELCSH_BRCA1_TARGETS_UP	Supresores tumorales	0,392	0,000	0,016	6907
WELCSH_BRCA1_TARGETS_DN	Supresores tumorales	0,424	0,000	0,014	6946
CHANDRAN_METASTASIS_UP	Invasión y metástasis	0,374	0,000	0,027	6995
APRELIKOVA_BRCA1_TARGETS	Supresores tumorales	0,501	0,000	0,016	7075
KEGG_OXIDATIVE_PHOSPHORYLATION	Fosforilación oxidativa	0,476	0,000	0,003	7229
ZHOU_TNF_SIGNALING_30MIN	Inflamación	0,452	0,002	0,034	7334
KEGG_LYSOSOME	Endocitosis o Fagocitosis	0,404	0,000	0,022	7368
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EX	Degradación de ARNm	0,455	0,000	0,009	7378
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_	Fosforilación oxidativa	0,511	0,000	0,002	7589
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	Fosforilación oxidativa	0,473	0,000	0,003	7589
BURTON_ADIPOGENESIS_2	Metabolismo lipídico	0,440	0,000	0,023	7604
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	Fosforilación oxidativa	0,517	0,000	0,003	8038

<sup>a</sup> p-valor nominal de cada conjunto génico sin corregir por comparaciones múltiples. <sup>b</sup> p-valor resultante de la corrección para comparaciones múltiples FDR (tasa de falsos positivos). <sup>c</sup> Posición en la lista ranqueada en la cual se obtiene el mayor valor de ES, los conjuntos génicos más interesantes poseen Rank at max muy pequeño o muy grande.

Tabla 36. Conjuntos génicos sobre-expresados en MI-No CG con los valores más extremos del parámetro Rank at max.						
Conjuntos génicos	Proceso funcional	ES	p-valor nominal <sup>a</sup>	q-valor FDR <sup>b</sup>	Rank at max <sup>c</sup>	
REACTOME_CHYLOMICRON_MEDIATED_LIPID_TRANSPORT	Metabolismo lipídico	0,843	0,000	0,157	86	
REACTOME_LIPOPROTEIN_METABOLISM	Metabolismo lipídico	0,748	0,000	0,123	86	
REACTOME_ABCA_TRANSPORTERS_IN_LIPID_HOMEOSTASIS	Metabolismo lipídico	0,683	0,009	0,202	167	
REACTOME_TERMINATION_OF_O_GLYCAN_BIOSYNTHESIS	Glicosilación proteica aberrante	0,716	0,002	0,144	210	
WANG_NEOPLASTIC_TRANSFORMATION_BY_CCND1_MYC	Oncogenes	0,570	0,002	0,191	284	
DAZARD_UV_RESPONSE_CLUSTER_G24	Respuesta a daño genómico	0,671	0,000	0,152	312	
LIU_CDX2_TARGETS_UP	Diferenciación intestinal	0,738	0,000	0,193	385	
REACTOME_LIPID_DIGESTION_MOBILIZATION_AND_TRANSPORT	Metabolismo lipídico	0,665	0,000	0,157	396	
WOTTON_RUNX_TARGETS_DN	Supresores tumorales	0,624	0,008	0,201	437	
MOOTHA_GLUCEOGENESIS	Efecto Warburg no tumoral	0,651	0,000	0,168	527	
MCMURRAY_TP53_HRAS_COOPERATION_RESPONSE_UP	Supresores tumorales	0,533	0,022	0,241	633	
GENTILE_UV_HIGH_DOSE_UP	Respuesta a daño genómico	0,605	0,021	0,170	661	
WANG_BARRETTES_ESOPHAGUS_UP	Diferenciación intestinal	0,744	0,000	0,144	829	
PID_IL8CXCR1_PATHWAY	Inflamación	0,662	0,000	0,140	860	
BAUS_TFF2_TARGETS_UP	Procesamiento antigénico	0,789	0,000	0,167	952	
PID_AMB2_NEUTROPHILS_PATHWAY	Inflamación	0,498	0,014	0,241	1044	
MURATA_VIRULENCE_OF_H_PILORI	H.pylori	0,669	0,002	0,137	1081	
KANNAN_TP53_TARGETS_UP	Supresores tumorales	0,439	0,019	0,242	1165	
KEGG_ETHER_LIPID_METABOLISM	Metabolismo lipídico	0,532	0,029	0,246	1240	
BROWN_MYELOID_CELL_DEVELOPMENT_UP	Inflamación	0,422	0,014	0,240	4270	
REACTOME_PHOSPHOLIPID_METABOLISM	Metabolismo lipídico	0,425	0,002	0,161	4273	
GENTILE_UV_LOW_DOSE_UP	Respuesta a daño genómico	0,682	0,000	0,138	4283	
AMIT_EGF_RESPONSE_120_HELA	Proliferación celular	0,425	0,023	0,241	4300	
ZUCCHI_METASTASIS_DN	Invasión y metastasis	0,505	0,039	0,236	4466	
REACTOME_RIP_MEDIATED_NFKB_ACTIVATION_VIA_DAI	Inflamación	0,621	0,032	0,242	4586	
REACTOME_TRAF6_MEDIATED_NFKB_ACTIVATION	Inflamación	0,569	0,016	0,241	4586	

Conjuntos génicos	Proceso funcional	ES	p-valor nominal <sup>a</sup>	q-valor FDR <sup>b</sup>	Rank at max <sup>c</sup>
KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	Metabolismo lipídico	0,630	0,019	0,200	4792
BIOCARTA_MITOCHONDRIA_PATHWAY	Apoptosis	0,682	0,002	0,167	4809
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	Procesamiento antigénico	0,515	0,004	0,169	4873
NEMETH_INFLAMMATORY_RESPONSE_LPS_UP	Inflamación	0,513	0,032	0,204	4873
KEGG_APOPTOSIS	Apoptosis	0,492	0,000	0,138	4882
PID_P53DOWNSTREAMPATHWAY	Supresores tumorales	0,429	0,004	0,158	4936
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	Procesamiento antigénico	0,436	0,004	0,159	5026
REACTOME_INTRINSIC_PATHWAY_FOR_APOPTOSIS	Apoptosis	0,613	0,006	0,172	5120
ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_DN	Oncogenes	0,395	0,020	0,199	5253
PID_IL12_2PATHWAY	Inflamación	0,479	0,036	0,239	5464
PID_PDGFBRPATHWAY	Angiogénesis	0,418	0,030	0,235	5748
REACTOME_TRAF6_MEDIATED_IRF7_ACTIVATION	Inflamación	0,543	0,009	0,203	6196
DANG_MYC_TARGETS_DN	Oncogenes	0,519	0,019	0,204	7011

<sup>a</sup> p-valor nominal de cada conjunto génico sin corregir por comparaciones múltiples. <sup>b</sup> p-valor resultante de la corrección por comparaciones múltiples FDR (tasa de falsos positivos). <sup>c</sup> Posición en la lista ranqueada en la cual se obtiene el mayor valor de ES, los conjuntos génicos más interesantes poseén Rank at max muy pequeño o muy grande.



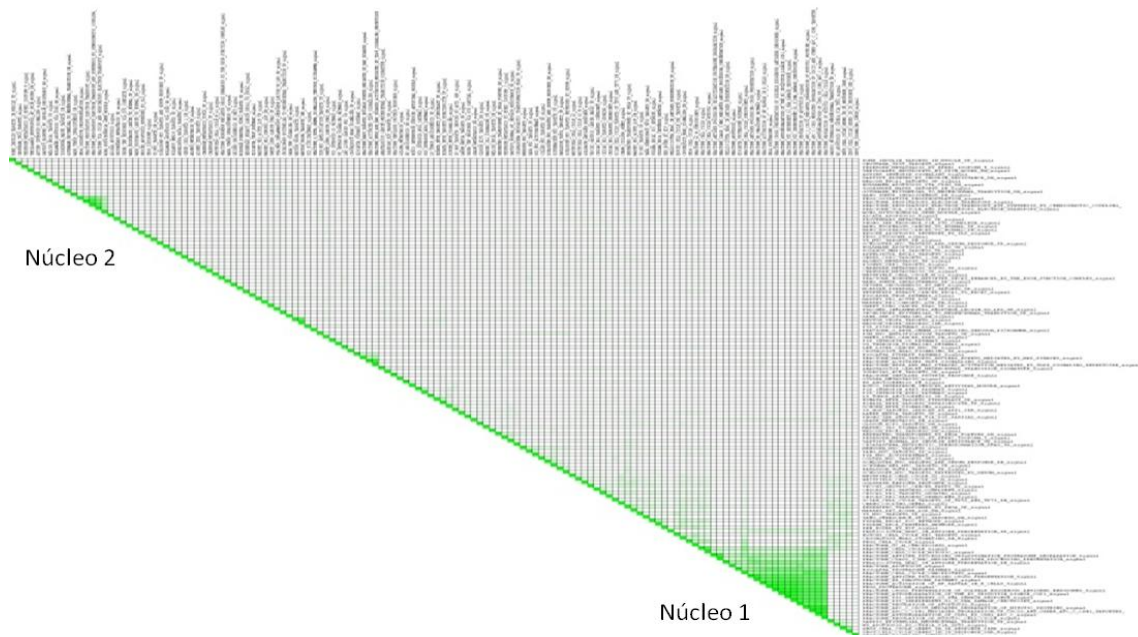
Tabla 37. Conjuntos génicos compuestos por al menos 3 genes <i>leading edge</i> sobre-expresados en el grupo MII-CG.						
Conjuntos génicos <sup>a</sup>	Proceso funcional <sup>b</sup>	ES <sup>c</sup>	N <sup>d</sup>	Genes <i>leading edge</i> sobre-expresados en la MII-CG <sup>e</sup>	Rank at max <sup>f</sup>	
IGLESIAS_E2F_TARGETS_UP	Ciclo y proliferación celular	0,418	76	CD53, HSP90AA1, MORF4L1, C3, RBBP7, LAPTM5	5813	
CHICAS_RB1_TARGETS_CONFLUENT	Supresores tumorales	0,361	183	CLU, ATP6V0E1, CAV1, EIF5B, MYOF, C1R	4907	
GOLDRATH_ANTIGEN_RESPONSE	Presentación y procesamiento antigénico	0,42	171	HLA-A, C3, CD24, RAN	6476	
LABBE_WNT3A_TARGETS_UP	Ciclo y proliferación celular	0,39	53	RBBP7, CAV1, FABP5, CD24	6333	
CHICAS_RB1_TARGETS_GROWING	Supresores tumorales		95	CLU, C1R, CAV1, FABP5	4697	
ROME_INSULIN_TARGETS_INMUSCLE_UP	Genes regulados por insulina	0,417	208	HLA-C, KARS, CCT6A	6051	
BERENIENO_TRANSFORMED_BY_RHOA_DN	Oncogenes	0,428	157	CAV, CD24, C1R	4980	
REACTOME_CELL_CYCLE	Ciclo y proliferación celular	0,355	149	NHP2, HSP90AA1, RBBP7	5615	
SANA_TNF_SIGNALING_DN	Inflamación	0,417	32	ATP5A1, CLU, MYOF	5458	
MENSEN_MYC_TARGETS	Oncogenes	0,615	30	HSP90AB1, C1QBP, HSP90AA1	5289	
HEDENFALK_BREAST_CANCERBRCA1_VS_BRCA2	Supresores tumorales	0,424	67	PDCD5, CCT6A, FABP5	5652	
ALONSO_METASTASIS_UP	Invasión y metástasis	0,477	88	NHP2, CAV1, RAN	5250	
WANG_ESOPHAGUS_CANCER_VS_NORMAL	Cáncer de esófago	0,477	56	HLA-A, C1R, LAPTM5	5578	

<sup>a</sup> Conjuntos génicos sobre-expresados en la MII-CG ordenados decrecientemente por el número de genes *leading edge* sobre-expresados en la MII-CG. <sup>b</sup> Proceso funcional al que pertenecen los conjuntos génicos. <sup>c</sup> *Enrichment score*. <sup>d</sup> Número de genes *leading edge* que componen el conjunto génico. <sup>e</sup> Genes *leading edge* que a su vez están sobre-expresados en la MII-CG. <sup>f</sup> Posición en la lista ranqueada en la cual se obtiene el mayor valor de ES, los conjuntos génicos más interesantes poseen Rank at max muy pequeño o muy grande.

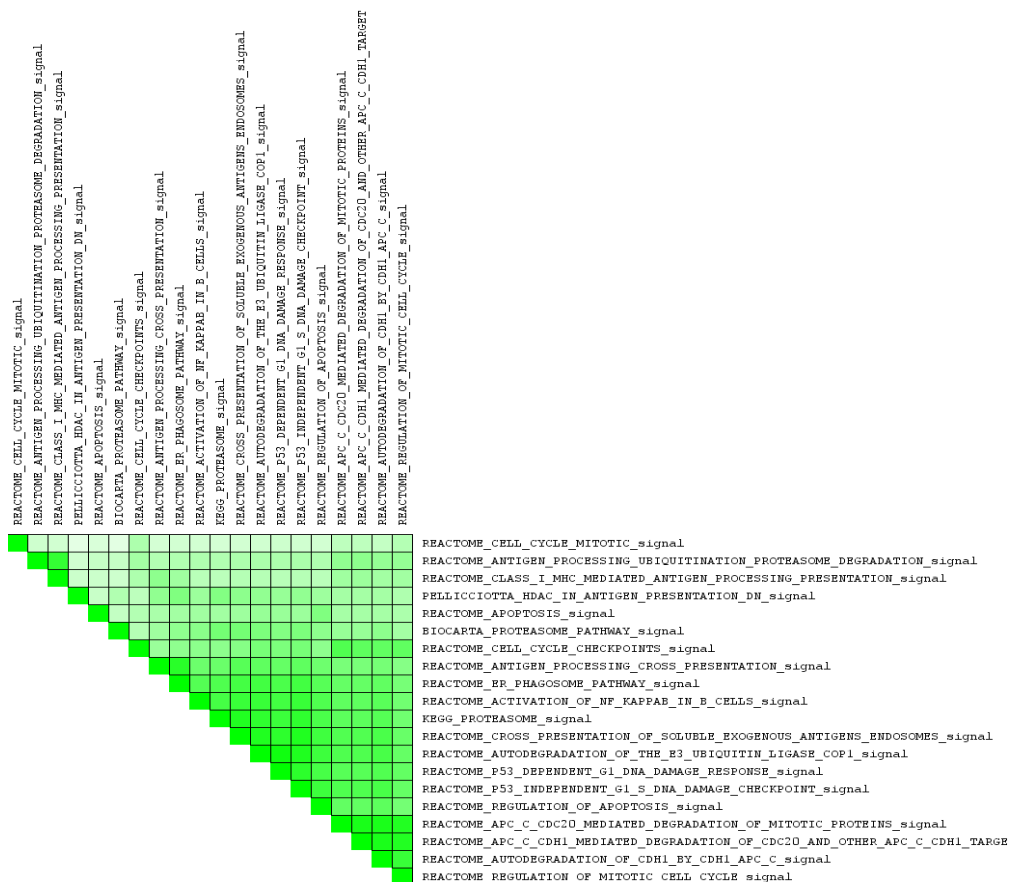
Tabla 38. Conjuntos génicos compuestos por al menos 3 genes <i>leading edge</i> sobre-expresados en el grupo MI-No CG.					
Conjuntos génicos <sup>a</sup>	Proceso funcional <sup>b</sup>	ES <sup>c</sup>	N <sup>d</sup>	Genes <i>leading edge</i> sobre-expresados en la MI-CG <sup>e</sup>	Rank at max <sup>f</sup>
BILD_HRAS_ONCOGENIC_SIGNATURE	Oncogenes	0,392	54	TRIB1, ELOVL7, AGPAT9, ACOT7, FGFBP1, PI3, CXCL1, SLC25A37, SERPINB5, TBX3, LYN, CD55, NT5E, TMEM45B, GK	2504
VECCHI_GASTRIC_CANCER_EARLY_UP	Cáncer Gástrico	0,508	133	ACAA2, LAPTM4B, RAD18, CDH3, EPPK1, CASC5, MACC1, OVOL1, TRIM31, CLDN3, CAMK2N1,	4205
VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_DN	Cáncer Gástrico	0,614	59	ATP10B, FAM3D, CLDN15, BTNL8, CFTR, REEP6, CLRN3, DHRS11, TM4SF20, GPA33, REG3A	2735
KEGG_DRUG_METABOLISM_OTHER_ENZYMES	Metabolismo de xenobióticos	0,673	12	UGT2B7, NAT2, TPMT, UGT2A3, UGT1A6, UGT1A7, CDA, CES2, CYP3A4	1348
OHGUCHI_LIVER_HNF4A_TARGETS_DN	Diferenciación intestinal	0,605	36	CFI, NLRP6, SULT1B1, SLC30A10, SLC3A1, S100A10, ETHE1, HSD17B2, MTPP	1356
BAUS_TFF2_TARGETS_UP	Procesamiento antigénico	0,789	11	IGHM, RBP2, DGKA, TFF3, ITLN1, ABCG2, ZG16, APOA4	952
BROWN_MYELOID_CELL_DEVELOPMENT_UP	Inflamación	0,422	52	UGT2B17, HEBP1, CLDN1, CCND2, CEACAM1, XDH, GLRX, ANPEP	4270
LABBE_TARGETS_OF_TGFB1_AND_WNT3A_DN	Proliferación celular	0,450	30	TPK1, TFRC, LGALS4, SLC30A4, HNF4A, VIL1, VNN1, SERPINA1	2919
ALCALA_APOPTOSIS	Apoptosis	0,524	30	GPI, MTH2, CASP1, HADHA, QPRT, ATP1A1	3103
KEGG_ARGININE_AND_PROLINE_METABOLISM	Efecto Warburg no tumoral	0,605	16	NAGS, MAOB, PRODH, AGMAT, OAT, ABPI	1415
SERVITJA_ISLET_HNF1A_TARGETS_DN	Diferenciación intestinal	0,515	35	NR1H4, CDHR2, RNF186, MTMR11, TM4SF4, ACE2	1707
REACTOME_AMINO_ACID_AND_OLIGOPEPTIDE_SLC_TRANSPORTERS	Efecto Warburg no tumoral	0,618	13	SLC15A2, SLC6A20, SLC1A1, SLC15A1, SLC7A9, SLC6A19	1480

Conjuntos génicos <sup>a</sup>	Proceso funcional <sup>b</sup>	ES <sup>c</sup>	N <sup>d</sup>	Genes <i>leading edge</i> sobre-expresados en la MII-CG <sup>e</sup>	Rank at max <sup>f</sup>
KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	Efecto Warburg no tumoral	0,563	8	GPT, ASS1, GLS, ACY3, CPS1	2144
REACTOME_METABOLISM_OF_CARBOHYDRATES	Efecto Warburg no tumoral	0,379	59	CHST5, B3GNT7, CHST6, SLC2A5, SLC5A1	3526
REACTOME_O_LINKED_GLYCOSYLATION_OF_MUCINS	Glicosilación proteica aberrante	0,565	17	GCNT3, MUC4, MUC13, MUC12, MUC17	2594
ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_DN	Oncogenes	0,395	44	SLCO2B1, ARL4A, IRF4, MPP1	5253
AUNG_GASTRIC_CANCER	Cáncer Gástrico	0,566	12	DEFA6, DEFA5, REG4, OLFM4	2619
KEGG_CHEMOKINE_SIGNALING_PATHWAY	Inflamación	0,389	46	PLCB3, CCL24, CCL15, CCL25	3524
KEGG_CITRATE_CYCLE_TCA_CYCLE	Efecto Warburg no tumoral	0,643	16	SUCLG1, IDH3A, PCK1, PCK2	2846
KEGG_GLYCOLYSIS_GLUONEOGENESIS	Efecto Warburg no tumoral	0,511	12	GALM, ADH6, LDHA, ALDOB	1249
LIU_CDX2_TARGETS_UP	Diferenciación intestinal	0,738	9	KRT20, CDX1, HEPH, MUC2	385
REACTOME_PHASE_II_CONJUGATION	Metabolismo de xenobióticos	0,647	20	SULT1A1, NAT1, GSTA1, SULT1E1	2050
TRANSPORT_OF_GLUCOSE_AND_OTHER_SUGARS_BILE_SALTS_AND_ORGANIC_ACIDS_METAL_IONS_AND_AMINE_COMPOUNDS	Efecto Warburg no tumoral	0,492	18	SLC22A18, SLC39A4, SLC39A5, SLC13A2	1446
KEGG_ETHER_LIPID_METABOLISM	Metabolismo lipídico	0,532	4	PLA2G12B, PLD1, PLA2G2A	1240
KEGG_GLYCEROLIPID_METABOLISM	Metabolismo lipídico	0,543	11	DGAT1, DGKQ, DAK	1926
LUCAS_HNF4A_TARGETS_UP	Diferenciación intestinal	0,601	17	ATP7B, CALML4, CIDEB	1688

<sup>a</sup> Conjuntos génicos sobre-expresados en la MII-CG ordenados decrecientemente por el número de genes *leading edge* sobre-expresados en la MII-CG. <sup>b</sup> Proceso funcional al que pertenecen los conjuntos génicos. <sup>c</sup> *Enrichment score*. <sup>d</sup> Número de genes *leading edge* que componen el conjunto génico. <sup>e</sup> Genes *leading edge* que a su vez están sobre-expresados en la MII-CG. <sup>f</sup> Posición en la lista ranqueada en la cual se obtiene el mayor valor de ES, los conjuntos génicos más interesantes poseen *Rank at max* muy pequeño o muy grande.



**Figura 35.** Análisis de genes *leading edge* realizado con todos los conjuntos génicos sobre-expresados en MII-CG. La acumulación de color verde indica que los conjuntos génicos comparten genes *leading edge*.



**Figura 36.** Subanálisis de genes *leading edge* del núcleo 1 (degradación proteasomal).

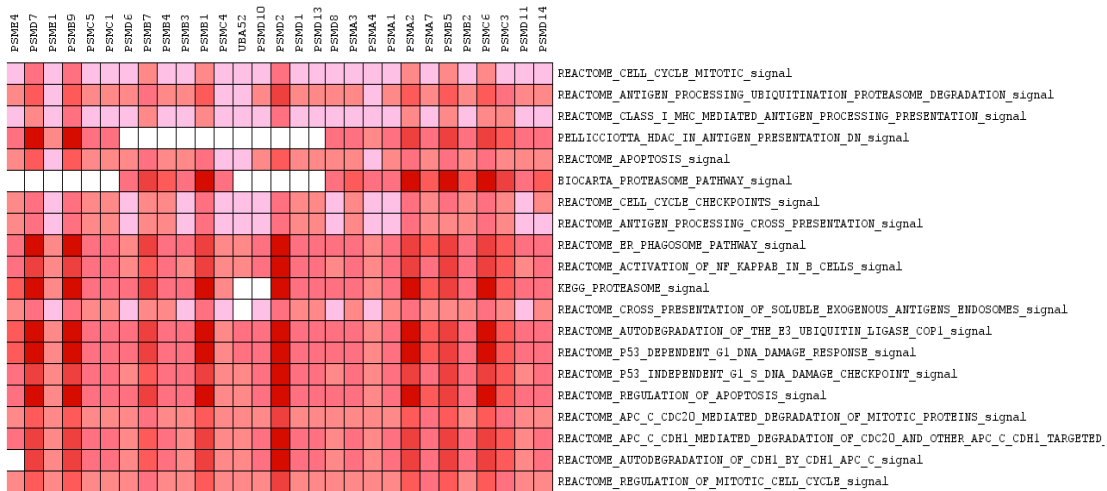


Figura 37. Heat map de los genes *leading edge* del núcleo 1 de la figura 35. Los colores rosado y rojo indican sobre-expresión génica.



Figura 38. Subanálisis de genes *leading edge* del núcleo 2 (fosforilación oxidativa) de la figura 35.

### 3.5.2. Ingenuity Pathway Analysis

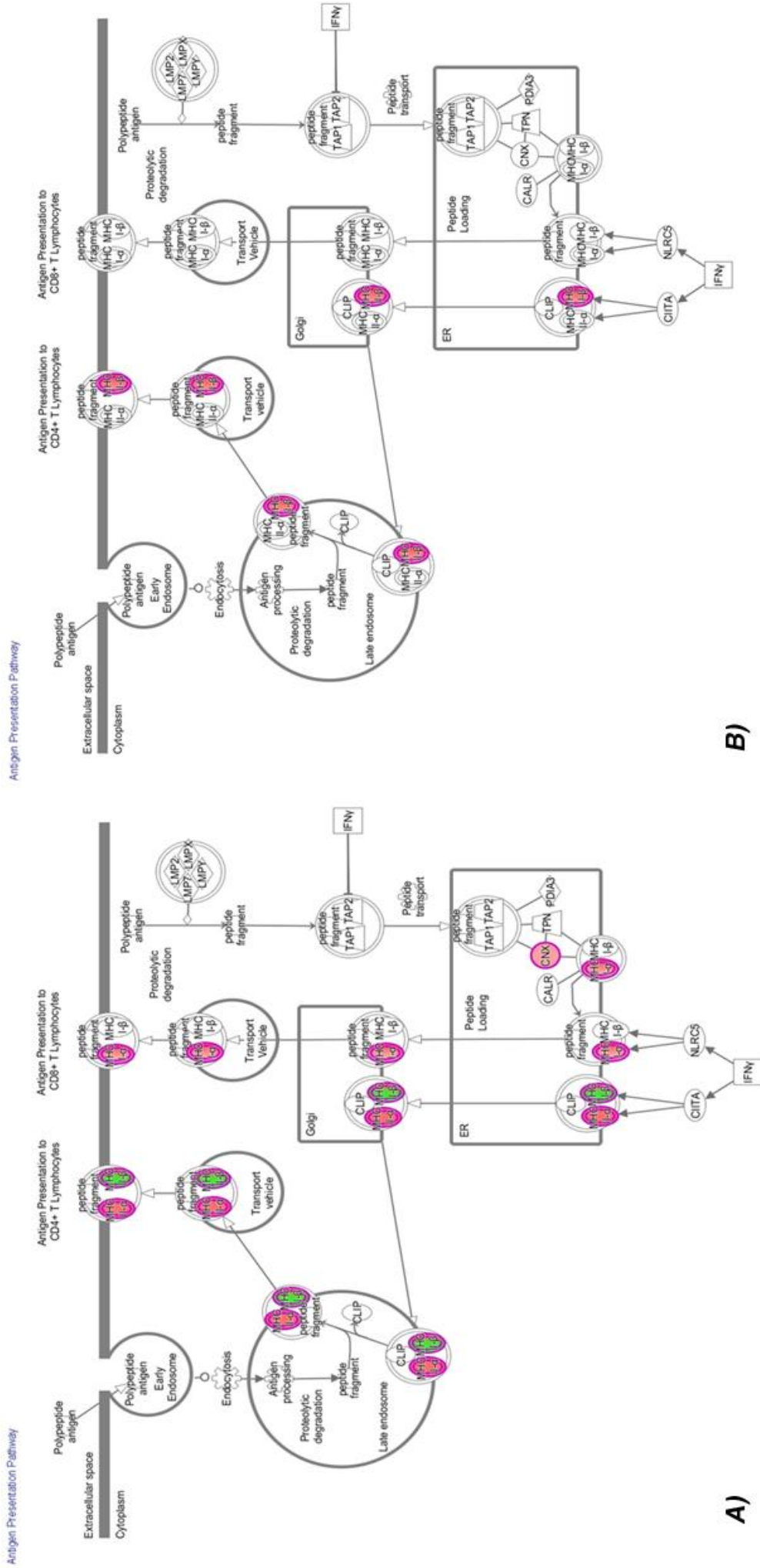
Con el objetivo de identificar las vías canónicas de señalización celular y redes moleculares a partir solamente de los genes diferencialmente expresados en los grupos MII-CG, MIC-CG y MI-No CG se empleó el programa IPA (Tabla 39). De entre todos los resultados significativos se seleccionaron solamente las vías relacionadas con la MI o la carcinogénesis. Todas contienen a moléculas HLA clase I y II sobre-expresadas, facilitando el procesamiento antigénico (Figura 39). A diferencia del análisis GSEA, el IPA muestra varias vías sobre-expresadas que son comunes a la MII-CG y a la MIC-CG como presentación antigénica, señalización de TNFRSF4 (OX40), comunicación entre células innatas y adaptativas, enfermedad tiroidea autoinmune y maduración del fagosoma (Tabla 39). La presentación antigénica, maduración del fagosoma (fagocitosis en MII-CG), maduración de células dendríticas y metabolismo de xenobióticos (MI-No CG) son vías que también se obtuvieron en el análisis con el GSEA.

Debido a la pequeña cantidad de genes diferencialmente expresados en la MIC-CG, las vías sobre-expresadas en este grupo están compuestas por un número reducido de moléculas (Figura 39B) y una menor significación estadística en comparación con la MII-CG (Tabla 39), sugiriendo indirectamente como en el análisis GSEA que a nivel molecular la MII confiere un mayor riesgo de progresión a CG respecto a la MIC<sup>93</sup>. Las vías más frecuentes en MI-No CG implican al metabolismo de xenobióticos y sus componentes receptores FXR, PXR, LXR, RXR.

El IPA también genera información de funciones moleculares, en el grupo MII-CG se evidencia la proliferación celular, lo cual está en concordancia con los resultados obtenidos con el GSEA, en los que se observa una gran cantidad de vías formadas por oncogenes y moléculas señalizadoras del ciclo y la proliferación celular. En todos los grupos predominan las enfermedades inflamatorias y gastrointestinales, lo cual concuerda con el hecho de que el principal factor de riesgo de la MI es la infección por *H. pylori*.

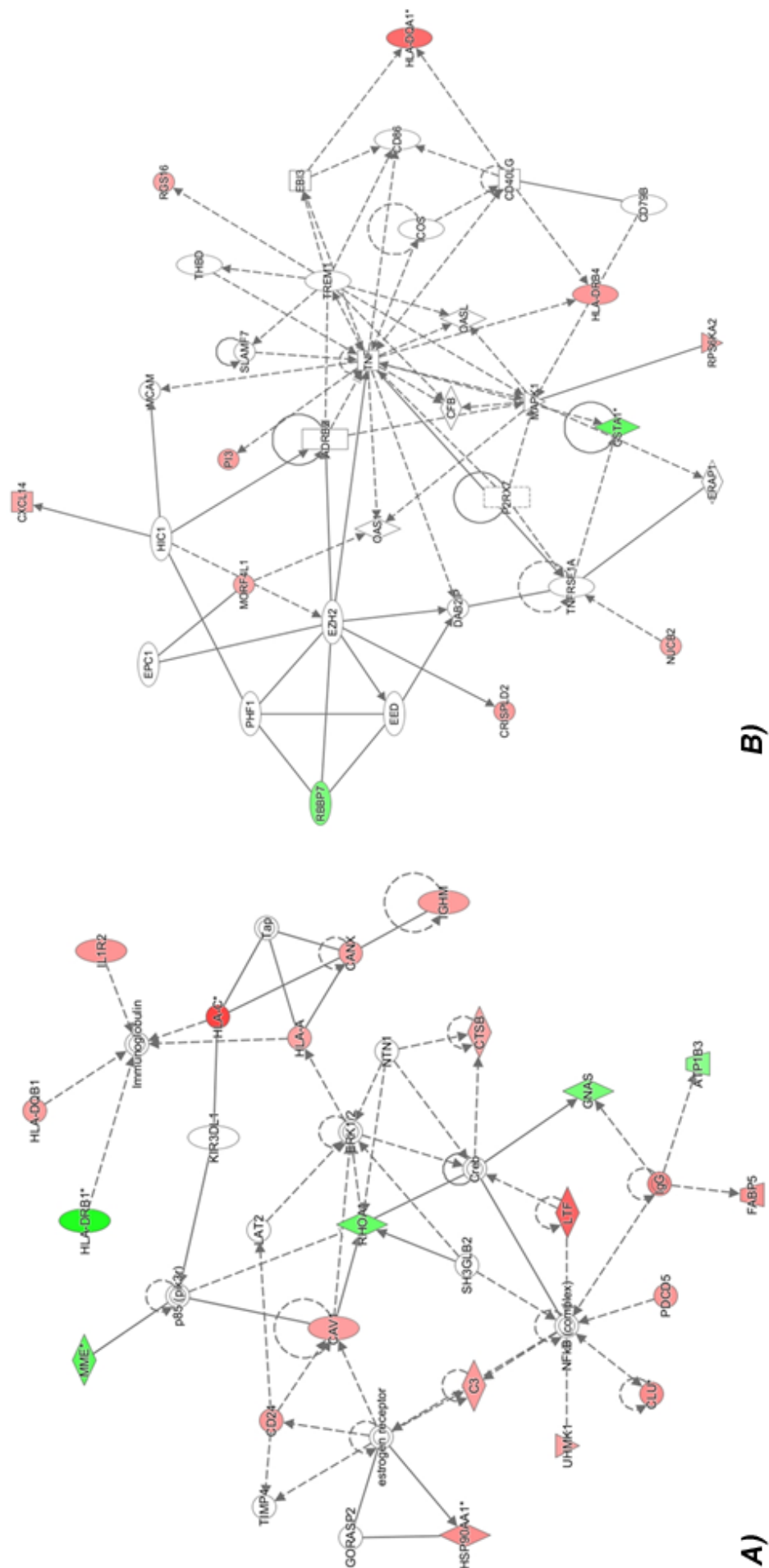
El IPA construyó redes moleculares de alto *score* a partir de los genes diferencialmente expresados. En la figura 40 se muestran las que poseen los mayores *scores*. La red molecular con mayor *score* en la MII-CG está compuesta por efectores de la respuesta inmune como moléculas HLA clase I (*HLA-A*, *HLA-C*) y II (*DQB1*, *DRB1*), el receptor de citoquina *IL1R2* y el componente de inmunoglobulinas *IGHM* (Figura 40A). Debido a la pequeña cantidad de genes significativos en la MIC-CG, en esta última no se obtiene ninguna red de alta puntuación.

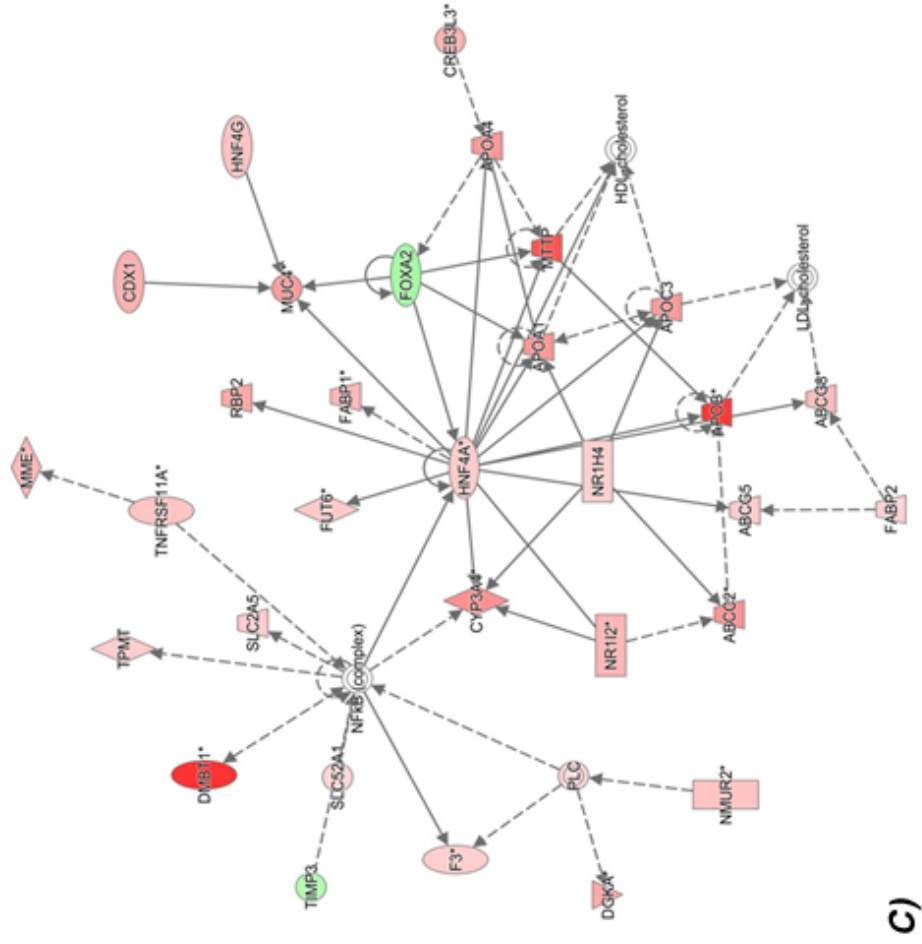
Tabla 39. Vías canónicas desreguladas e información funcional obtenida en el IPA.					
Comparación	Vías canónicas	p-valor	Reguladores <i>upstream</i>	Enfermedades	Funciones moleculares y celulares
MII-CG vs MII-No CG	Presentación antigénica	3,15E-10	HOXC11	-Inmunológicas -Del tejido conectivo -Inflamatorias -Musculares y esqueléticas -Del desarrollo	-Proliferación celular -Muerte y supervivencia celular -Movimiento celular -Desarrollo celular -Señalización e interacción célula-célula
MII-CG vs MII-No CG	Señalización de TNFRSF4 (OX40)	3,27E-09	RAD21		
MII-CG vs MII-No CG	Enfermedad tiroidea autoinmune	1,66E-08	EBI3 NLR5		
MII-CG vs MII-No CG	Desarrollo de células B	1,51E-07	SMC3		
MII-CG vs MII-No CG	Maduración del fagosoma	6,03E-02			
MII-CG vs MII-No CG	Comunicación entre células innatas y adaptativas	8,11E-02			
MII-CG vs MII-No CG	Comunicación entre células innatas y adaptativas	3,14E-05	LGALS3		
MIC-CG vs MIC-No CG	Maduración de células dendríticas	3,22E-04	LHCGR MYBL2	-Endocrinas -Gastrointestinales -Inmunológicas -Inflamatorias -Del desarrollo	-Señalización e interacción célula-célula -Movimiento celular -Muerte y supervivencia celular -Ciclo celular -Ensamblaje y organización celular
MIC-CG vs MIC-No CG	Presentación antigénica	3,98E-04	mir-296		
MIC-CG vs MIC-No CG	Señalización de TNFRSF4 (OX40)	3,13E00	CHI3L1		
MIC-CG vs MIC-No CG	Interacción e/ células dendríticas y NK	2,3E-02			
MIC-CG vs MIC-No CG	Maduración del fagosoma	1,72E-02			
MIC-CG vs MIC-No CG	Señalización de enfermedad tiroidea autoinmune	1,53E00			
MIC-CG vs MIC-No CG	Maduración de células dendríticas	3,22E-04			
MI-No CG vs Sanos	Activación de FXR/RXRG	1,43E-08	HNF4A	-Endocrinas -Metabólicas -Gastrointestinales -Hepáticas -Cardiovasculares	- Movimiento celular - Muerte y supervivencia celular -Metabolismo lipídico -Transporte molecular -Bioquímica de moléculas pequeñas, drogas y aminoácidos
MI-No CG vs Sanos	Activación de PXR/RXRG	1,34E-07	HNRNPA2B1		
MI-No CG vs Sanos	Inhibición por LPS/IL1 de RXRG	1,43E-08	HNF1A		
MI-No CG vs Sanos	Desarrollo de células B	3,4E-04	CTNBB1		
MI-No CG vs Sanos	Degradación de melatonina	2,09E-01	SREBF1		
MI-No CG vs Sanos	Degradación de melatonina	2,08E-01			
MI-No CG vs Sanos	Metabolismo de hormona tiroidea	1,28E-07			
MI-No CG vs Sanos	Hematopoesis desde células madres	7,29E-02			
MI-No CG vs Sanos	Activación de LXR/RXR	9,32E-02			



**Figura 39.** Esquemas de las vías canónicas presentación y procesamiento antigénico sobre-expresada en los grupos MII-CG (A) y MIC-CG (B) obtenidas con el programa *Ingenuity Pathway Analysis*. Las moléculas en colores rojo y verde se encuentran sobre y sub-expresadas, respectivamente.







**Figura 40.** Redes de alto score IPA de las comparaciones MII-CG vs MII-No CG (A-B) y MI-No CG vs Sanos (C). La intensidad del color en los nodos indica el grado de expresión: verde sub-expresado, rojo sobre-expresado y sin color no tiene diferencia de expresión. La longitud de las líneas discontinuas refleja la evidencia en la literatura en la relación entre moléculas. Las líneas no discontinuas indican interacción proteína-proteína. Las flechas indican la dirección de asociación.

La red molecular con mayor score (44) en el grupo MI-No CG contiene a varias moléculas del metabolismo lipídico como APOA4, APOB, APOC3, APOA1, FABP1/2, MTP y HNF4A, sobre-expresadas en la MI-No CG y que son componentes de los conjuntos génicos significativos del GSEA (Tabla 40, Figura 40C). El factor de transcripción de hepatocitos HNF4A, con un rol en el desarrollo intestinal, está sobre-expresado en los datos e interactúa con varias lipoproteínas, siendo un nodo de esta red molecular y quizás responsable del incremento del metabolismo lipídico anteriormente mencionado.

### **3.5.3. Análisis de ontología génica**

El análisis de enriquecimiento de ontología génica de los genes diferencialmente expresados en las categorías de procesos biológicos, funciones moleculares y componentes celulares con validación experimental nos permitió corroborar diversos resultados previos de IPA y GSEA de procesos funcionales en cada uno de los grupos estudiados (Tabla 41). Por ejemplo, el rol del metabolismo lipídico en el grupo MI-No CG se confirma ya que es una de las funciones moleculares obtenidas.

Tabla 40. Redes moleculares obtenidas en el IPA para las comparaciones MII-CG vs MII-No CG y MI-No CG vs Sanos.				
Redes moleculares	Comparación	Score	Moléculas centrales (Fold Change) <sup>a</sup>	Comúnes con GSEA <sup>b</sup>
Enfermedades del tejido conectivo, enfermedades inflamatorias, esqueléticas y musculares	MII-CG vs MII-No CG	39	Sobre-expresadas: HLA-DQB1, IL1R2, HLA-C, HLA-A, CANX, IGHM, CD24, CAV1, HSP90AA1, C3, CTSB, UHMK1, CLU, PDCD5, LTF, FABP, HLA-DRB1, RHOA Sub-expresadas: MIME, HLA-DRB1, RHOA, GNAS, ATP1B3	HLA-A, CD24, C3, LTF
Señalización e interacción celular, desarrollo del sistema hematológico, respuesta inmune humoral	MII-CG vs MII-No CG	16	Sobre-expresadas: CXCL14, MORF4L1, PI3, RGS16, HLA-DQA1, HLA-DRB4, CRISPLD2, NUCB2, Sub-expresadas: RBBP7, GSTA1	NUCB2
Muerte y sobrevivencia celular, enfermedad inmunológica y metabólica	MII-CG vs MII-No CG	16	Sobre-expresadas: CCDC144B, C1R, HLA-DRB3, IGHG1, ANP32B, GNL3, LYZ Sub-expresadas: RBP2, DNAJB7, CP, GIP	C1R, ANP32B
Cáncer, daño orgánico, muerte y sobrevivencia	MIC-CG vs MIC-No CG	5	No son redes de alto score (>15)	
Respuesta anti-microbiana, inflamatoria, señalización e interacción celular	MIC-CG vs MIC-No CG	3	No son redes de alto score (>15)	
Respuesta inflamatoria, desarrollo tisular y morfología celular	MIC-CG vs MIC-No CG	3	No son redes de alto score (>15)	

Redes	Comparación	Score	Moléculas centrales <sup>a</sup>	Comúnes con GSEA <sup>b</sup>
Metabolismo lipídico, Transporte molecular, Bioquímica de moléculas pequeñas	MI-No CG vs Sanos	44	Sobre-expresadas: ABCC2, ABCG5, ABCG8, APOA1, APOA4, APOB, APOC3, CDX1, CREB3L3, CYP3A4, DGKA, DMBT1, F3, FABP1, FABP2, FUT6, HNF4A, HNF4G, MME, MTPP, MUC4, NMUR2, NR1H4, NR1I2, RBP2, SLC2A5, SLC52A1, TNFRSF11A, TPMT  Sub-expresadas: FOXA2, TIMP3	ABCC2, APOA4, CDX1, FABP1, FABP2, HNF4A, MUC4, CASPI, CD55, CEACAM1, TRIB1, CXCL1, EPHX2, PRODH,
Muerte y supervivencia celular, Desarrollo y función del sistema digestivo, Morfología de órganos	MI-No CG vs Sanos	41	Sobre-expresadas: ACHE, ACSL5, ANXA2, CASP1, CD55, CEACAM1, CIDEB, CIDEA, CYCS, CYP4F3, EPHX2, GPRC5A, HMOX1, IL32, IL2RG, PRODH, MDK, PFKFB2, PIGR, S100A10, SERPINB5, TFF3, TRIB1, VDR  Sub-expresadas: CCKAR, CSTA, MAL, CXCL1, DGKD, GAST	CD55, CEACAM1, TRIB1, CXCL1, EPHX2, PRODH,
Cáncer, daño y anomalías a nivel de organismo, Ciclo celular	MI-No CG vs Sanos	37	Sobre-expresadas: PSMD1, ARL4A, ATP1B3, ATP7B, CCND2, CDKN1A, CLDN1, COL17A1, DEFA5, EPCAM, GCG, HSD17B2, IgG e IgM, LGALS3, LYN, MSLN, NT5E, SERPINA1, SLAMF7, SLC39A4, TFRC, TOP1  Sub-expresadas: ALDH3A1, CLU, FABP5, GPR64, MUC6, SLC29A1	-----

<sup>a</sup> Moléculas sobre y sub-expresadas que interactúan con otras y sus niveles de expresión (*Fold Change*). <sup>b</sup> Moléculas centrales de las redes construidas por el IPA que a su vez son componentes de los conjuntos génicos obtenidos con el GSEA.

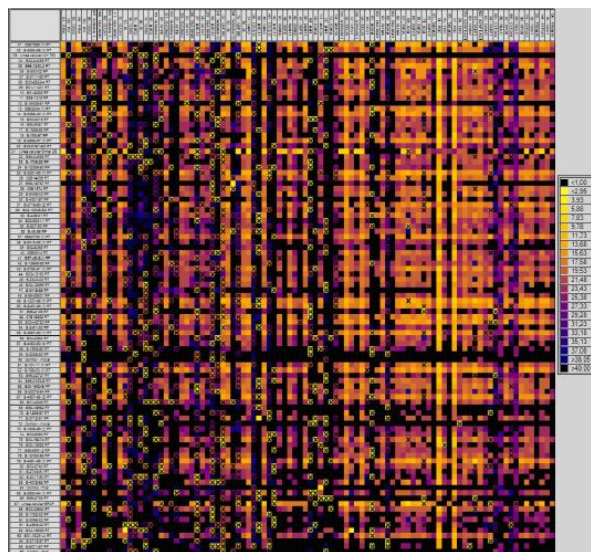
Tabla 41. Resultados del análisis de ontología génica en las comparaciones de estudio.

Comparación	Procesos Biológicos	Funciones Moleculares	Componentes Celulares
MII-CG vs MII-No CG	Presentación y procesamiento antigénico de péptidos o polisacáridos vía MHC clase II (p=6.59E-06). Respuesta inmune humoral (p-valor=2.02E-04).	Unión de antígenos (p=9.77E-04)	Complejo MHC (p=2.35E-08), orgánulos y vesículas extracelulares (p=1.55E-04), exosoma extracelular (p=5.56E-13), membrana del endosoma tardío (p=4.77E-03), vesículas unidas a membrana (p=4.56E-12)
MIC-CG vs MIC-No CG	Presentación y procesamiento antigénico de péptidos o polisacáridos vía MHC clase II (p=7.36E-06). Respuesta de defensa celular (p=9.06E-03)	No clasificado	Complejo MHC (p=9.88E-08), membranas del endosoma tardío (p=4.73E-06), lisosoma (p=5.88E-03), vacuola (p=7.38E-03)
MI-No CG vs Sanos	Metabolismo de: flavonoides (p=4.82E-07), glucuronidación (p=4.35E-09), ácido urónico (p=4.35E-09), glucuronato (p=4.35E-09), xenobióticos (p=5.17E-08), monosacáridos (p=2.88E-07), carbohidratos (p=2.98E-06), moléculas pequeñas (p=1.55E-04). Homeostasis del colesterol (p=2.03E-02), esterol (p=2.03E-02) y lípidos (p=2.59E-02). Metabolismo de lípidos (p=6.83E-08) y ácidos grasos (p=1.55E-02)	UDP-glucuronosil transferasa (p=1.32E-07), unión a ácidos nucleicos (p=4.64E-05), UDP-glicosil transferasa (p=6.97E-04), unión de iones (p=1.24E-06), transferencia de grupos glicosilo (p=7.41E-03), actividad sulfo transferasa (p=1.17E-02), unión de ácidos carboxílicos (p=2.01E-03), transportadores de colesterol y esterol (p=2.01E-03), unión a lípidos (p=1.13E-02)	Complejo MHC (p=2.35E-08), región extracelular (p=7.77E-04), complejos proteicos (p=6.38E-03) y macromoleculares (p=6.38E-03), membranas del retículo endoplásmico y nuclear externa (p=1.82E-03), complejos proteicos (p=6.38E-03) y macromoleculares (p=3.82E-03), matriz extracelular (p=4.88E-02)

### 3.6. Validación de la expresión génica por qRT-PCR

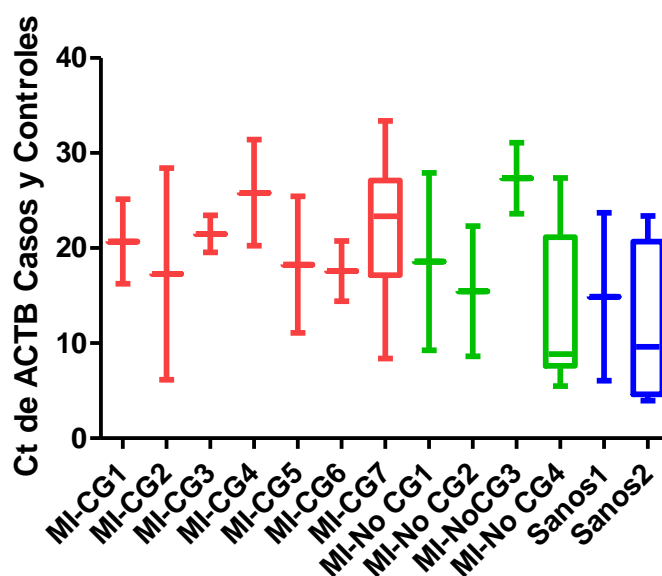
El método de la qRT-PCR se escogió como una técnica alternativa para intentar validar las diferencias significativas entre casos y controles de 37 genes diferencialmente expresados en el microarray de expresión, así como de 3 genes de la familia *Schlafen* entre los que se encuentra *Schlafen5*. Además, se escogieron 5 genes de referencia para normalizar los resultados de expresión basados en reportes específicos de tejidos gástricos<sup>187</sup>. Específicamente, fueron seleccionados 19 genes diferencialmente expresados en la MII-CG, 3 en la MIC-CG y 18 en la MI-No CG, entre los cuales existen 3 genes compartidos.

El análisis del array de expresión por qRT-PCR en la plataforma Biomark HD 96x96 (Fluidigm, EUA) indicó que solamente el 50.98% de las reacciones (Figura 41) tenían calidad suficiente para ser analizadas estadísticamente; según el criterio *Quality Threshold* >0.3 del programa Fluidigm Real-Time PCR Analysis v4.0.1 que busca un equilibrio entre la fiabilidad de los resultados y el número de datos disponibles, teniendo en cuenta que valores mayores de *Quality Threshold* son más restrictivos y muchas muestras son perdidas del análisis. Se atribuye el alto fallo técnico al hecho de que el ARN proviene de muestras parafinadas viejas (media de 16.4 años de antigüedad) con un elevado grado de degradación (RIN≈2). De todos los genes de referencia escogidos, *ACTB* no evidenció diferencias significativas en los valores de Ct entre los grupos MI-CG vs MI-No CG vs Sanos por lo que fue seleccionado como el mejor (Tabla 42 y Figura 42). El gen *GAPDH* tampoco mostraba diferencias estadísticas entre MI-CG vs MI-No CG; sin embargo, no se disponía de los datos del mismo en controles sanos. Los experimentos con *G6PD* fallaron en la mayoría de las muestras del grupo MI-No CG.



**Figura 41.** Heatmap de la qRT-PCR en la plataforma Biomark HD 96x96 (Fluidigm, EUA).

Tabla 42. Resultados de prueba ANOVA de los genes de referencia en los grupos MI-CG, MI-No CG y controles sanos.			
Gen de referencia	Comparación	ANOVA	Observaciones
<i>RPL29</i>	MI-CG vs MI-No CG	< 0.0001	
<i>RPL29</i>	MI-CG vs MI-No CG vs Sanos	< 0.0001	
<i>B2M</i>	MI-CG vs MI-No CG	< 0.0001	
<i>B2M</i>	MI-CG vs MI-No CG vs Sanos	< 0.0001	
<i>G6PD</i>	Casos MI-CG vs Sanos	0.2023	Falló en MI-No CG
<i>GAPDH</i>	Casos MI-CG vs MI-No CG	0.6617	Falló en controles sanos
<i>ACTB</i>	MI-CG vs MI-No CG	0.4478	Gen escogido
<i>ACTB</i>	MI-CG vs MI-No CG vs Sanos	0.3946	Gen escogido



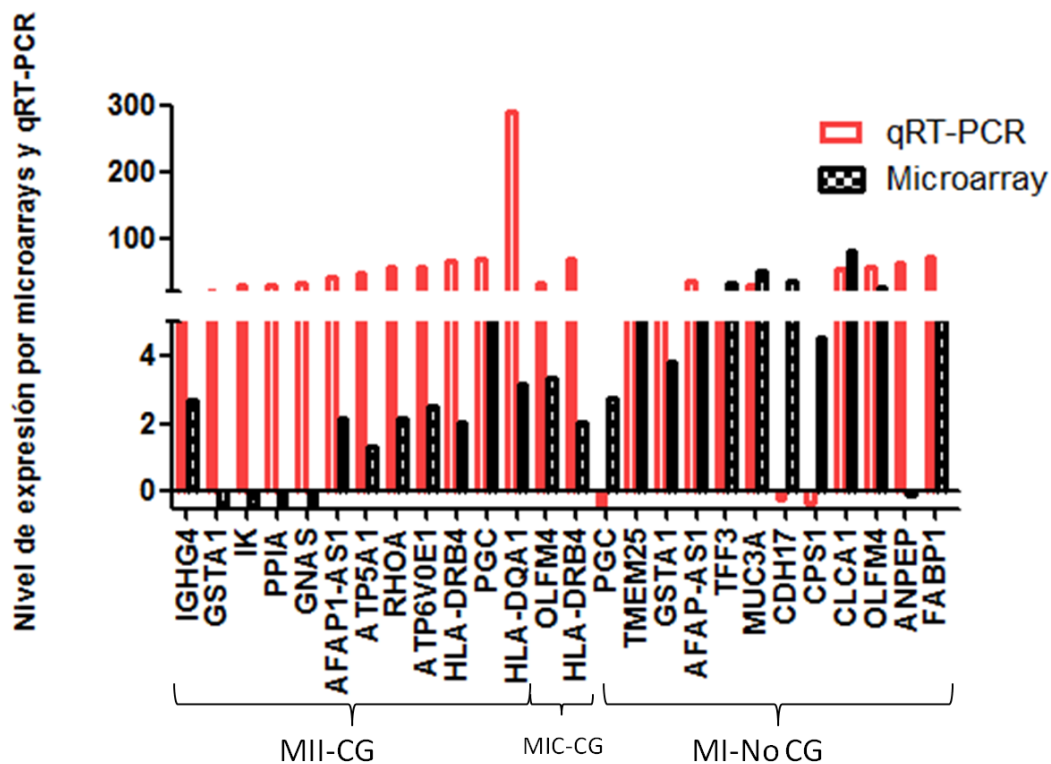
**Figura 42.** Gráfico de cajas de los valores de Ct del gen *ACTB* en las muestras analizadas de los grupos MI-CG, MI-No CG y Sanos.

A continuación, se eliminaron los valores extremos de Ct, se calculó la media de Ct (debido a que el chip duplica las muestras) y dCt de los genes de interés respecto al gen de referencia *ACTB* ( $dCt = Ct \text{ GenX} - Ct \text{ ACTB}$ ), se eliminaron los valores extremos de dCt y se aplicó la fórmula  $2^{-\text{Media } dCt}$ . Las diferencias estadísticamente significativas de expresión génica fueron determinadas a partir de los valores de la relación  $2^{-dCt(\text{Casos})} / 2^{-dCt(\text{Controles})}$  introducidos en el programa Bootstratio (<http://regstattools.net/br>). Casos son los grupos MII-CG, MIC-CG y MI-No CG, mientras que controles son MII-No CG, MIC-No CG y Sanos respectivamente.

El resultado del test de permutaciones realizado para determinar la significación estadística muestra que 27 genes de los 37 evaluados por qRT-PCR están diferencialmente expresados en los grupos MII-CG vs MII-No CG y MI-No CG vs



Sanos en los resultados de qRT-PCR (Tabla 43 y Figura 43). Sin embargo en solo 21 de ellos fue validada el mismo sentido de la diferencia de expresión, para un 56.75% de efectividad. Los genes *GSTA1*, *IK*, *PPIA*, *GNAS* del grupo MII-CG se encuentran sub-expresados en el microarray mientras que están sobre-expresados en la qRT-PCR. Los genes *IGHG4*, *AFAP1-AS1*, *ATP5A1*, *RHOA*, *ATP6V0E1*, *HLA-DRB4*, *PGC*, *HLA-DQA1*, *OLFM4*, *HLA-DRB4* están sobre-expresados en MII-CG y MIC-CG mediante ambas tecnologías; pero tuvieron un mayor *Fold Change* en la qRT-PCR que en el microarray. Respecto al grupo MI-No CG la mayoría de los genes (*TMEM25*, *GSTA1*, *AFAP-AS1*, *MUC3A*, *CLCA1*, *OLFM4*, *ANPEP*, *FABP1*) se encuentran sobre-expresados mediante ambas tecnologías. Las excepciones son *CDH17* y *CPS1*, sobre-expresados en el microarray y sub-expresados en la qRT-PCR.



**Figura 43.** Gráfica del nivel de expresión del microarray y la qRT-PCR de los genes significativos por ambas metodologías.

<b>Tabla 43. Resultados del test de permutaciones del programa Bootstratio empleado para estimar diferencias significativas en grupos MII-CG, MIC-CG y MI-No CG.</b>						
Genes significativos del microarray	Tamaño de muestra <sup>a</sup>	Fold Change en qRT-PCR <sup>b</sup>	Fold Change en microarray <sup>c</sup>	Grupo donde es significativo <sup>d</sup>	p-valor en qRT-PCR <sup>e</sup>	Validado
<i>IGHG4</i>	11/7	12.82	2.67	MII-CG	p<0.0005	SI
<i>GSTA1</i>	13/7	20.18	0.39	MII-CG	0.0015	NO
<i>IK</i>	13/6	28.12	0.38	MII-CG	0.006	NO
<i>PPIA</i>	12/7	30.54	0.39	MII-CG	p<0.0005	NO
<i>GNAS</i>	11/7	31.56	0.45	MII-CG	p<0.0005	NO
<i>SLFN5</i>	6/4	33.46	No sign	MII-CG	5.00E-04	SI
<i>AFAP1-AS1</i>	12/7	41.53	2.16	MII-CG	p<0.0005	SI
<i>ATP5A1</i>	12/7	46.39	1.3	MII-CG	5.00E-04	SI
<i>RHOA</i>	13/7	55.91	2.16/0.4	MII-CG	0.001	SI
<i>ATP6VOE1</i>	10/6	57.18	2.49	MII-CG	0.0075	SI
<i>HLA-DRB4</i>	10/5	67.13	2.25	MII-CG	0.031	SI
<i>PGC</i>	13/8	67.59	5.01	MII-CG	0.003	SI
<i>HLA-DQA1</i>	50/3	288.90	3.18	MII-CG	0.0345	SI
<i>OLFM4</i>	22/29	33,30	3.33	MIC-CG	5.00E-04	SI
<i>HLA-DRB4</i>	14/14	68,81	2.05	MIC-CG	0.0335	SI
<i>PGC</i>	39/9	0.65	0.13	MI-No CG	0.0075	SI
<i>TMEM25</i>	6/2	15.8	2.78	MI-No CG	p<0.0005	SI
<i>GSTA1</i>	26/4	17.35	3.81	MI-No CG	p<0.0005	SI
<i>AFAP1-AS1</i>	33/7	35.67	4.52	MI-No CG	p<0.0005	SI
<i>TFF3</i>	38/8	7.18	9.08	MI-No CG	0.009	SI
<i>MUC3A</i>	27/6	30.68	15.9	MI-No CG	p<0.0005	SI
<i>CDH17</i>	30/3	0.25	18.02	MI-No CG	p<0.0005	NO
<i>CPS1</i>	32/6	0.37	26.25	MI-No CG	p<0.0005	NO
<i>CLCA1</i>	34/2	55.26	33.77	MI-No CG	p<0.0005	SI
<i>OLFM4</i>	36/3	56.36	37.04	MI-No CG	p<0.0005	SI
<i>ANPEP</i>	31/4	62.3	49.98	MI-No CG	p<0.0005	SI
<i>FABP1</i>	25/2	70.85	80.58	MI-No CG	p<0.0005	SI

<sup>a</sup> Número de muestras analizadas en los grupos de casos (MI-CG)/controles (MI-No CG).  
<sup>b</sup> Media de valores de expresión de casos (MI-No CG) dividido entre controles (Sanos) en la qRT-PCR. <sup>c</sup> Media de expresión de casos dividido entre controles en el microarray. <sup>d</sup> Se refiere a el microarray. <sup>e</sup> p-valor del test de permutaciones que evalúa la significación de la diferencia de expresión entre casos y controles en la qRT-PCR.

#### 4. Estudio de expresión génica e inmuno-histoquímica de Schlafen5 en lesiones precursoras y cáncer gástrico

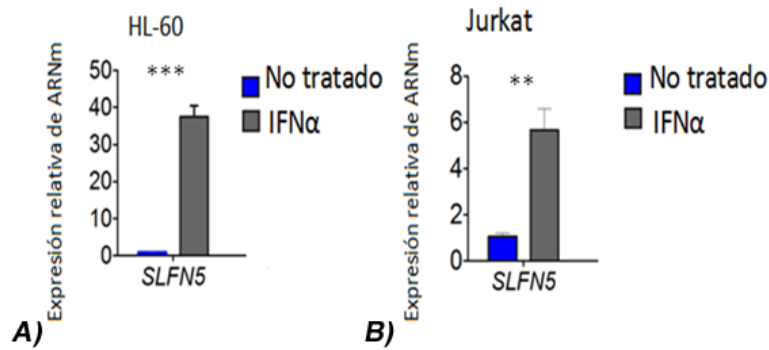
##### 4.1. Inducción de la expresión de Schlafen5 por IFN $\alpha$ y *Helicobacter pylori*

Al incubar líneas celulares humanas de monocitos (HL-60) y de células T (Jurkat) con IFN2A se observó un aumento significativo en la expresión del ARNm de Schlafen5 en ambas líneas celulares, aunque dicho incremento es mayor en la HL-60 (Tabla 44 y Figura 44).

Cuando fueron incubadas estas líneas con *H. pylori* solo se observó un incremento en la expresión de Schlafen5 en la HL-60, mientras que en las Jurkat se observó un decremento de la expresión, aunque ambos p-valores son nominales.

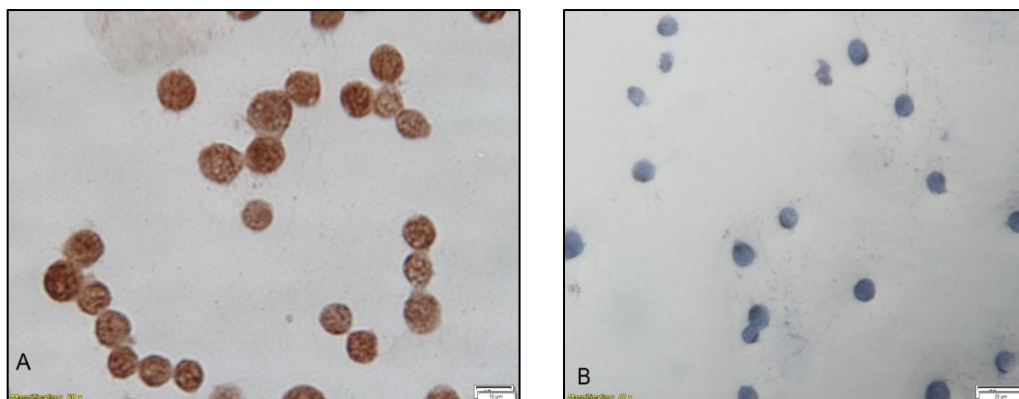
<b>Tabla 44. Expresión génica de Schlafen5 en líneas celulares expuestas a diversas condiciones.</b>				
<b>Línea celular</b>	<b>Expresión normalizada<sup>a</sup></b>		<b>Razón de medias <math>\pm</math> DE<sup>b</sup></b>	<b>Significación estadística<sup>c</sup></b>
	<b>Tratadas con IFN2A</b>	<b>No tratadas con IFN2A</b>		
HL-60	34.655	1	37.455 $\pm$ 13.57	5.00E-04
	50.914	1		
	47.34	1		
	47.9	0.791		
	27.039	0.836		
	31.167	1.373		
	35.52	0.712		
	29.153	1.021		
	33.372	1.266		
Jurkat	2.7	1	5.3604 $\pm$ 2.631	0.0035
	3.498	1		
	3.539	0.663		
	3.605	1		
	8.037	1.45		
	5.376	2.004		
	5.075	1		
	9.646	0.51		
	9.491	0.881		
	<b>Tratadas con <i>H. pylori</i></b>	<b>No tratadas con <i>H. pylori</i></b>		
HL-60	6.505	1.096	3.8103 $\pm$ 2.365	0.039
	0.875	1.486		
	4.901	0.558		
	5.709	0.888		
	1.24	1.02		
Jurkat	0.694	1.346	0.6369 $\pm$ 0.257	0.044
	0.724	0.725		
	0.629	0.681		
	0.558	0.999		
	0.711	1.456		

<sup>a</sup> Valores de expresión de SLFN5 normalizados respecto al ARNr 18S (Hs 99999901\_s1). <sup>b</sup> Valor promedio de la expresión de las líneas celulares tratadas/no tratadas con IFN2 $\alpha$  o *H. pylori* y su desviación estándar. <sup>c</sup> p-valor resultante del test de permutaciones del programa Bootstratio.



**Figura 44.** Inducción de Schlafen5 por el IFN $\alpha$  en las líneas celulares HL-60 y Jurkat.

Con el objetivo de confirmar la inducción de Schlafen5 por el IFN $\alpha$  mediante una técnica inmunohistoquímica alternativa, se realizó un cytopspin de células T (Jurkat) incubadas con IFN $2\alpha$  después de haber sido transfectadas o no (control positivo) con el ARN de interferencia específico para Schlafen5. En la figura 45 puede observarse la expresión de esta proteína, debido al color pardo de la tinción con diaminobenzidina en las células donde no está inhibido su ARNm (Figura 45A), mientras que en las transfectadas con el ARN de interferencia de Schlafen5 no se observa dicho color (Figura 45B).



**Figura 45.** Análisis inmunohistoquímico de Schlafen5 después de su inducción por incubación con IFN $\alpha$  de la línea celular Jurkat. A) Línea celular sin tratamiento. B) Línea celular tratada con ARN de interferencia (siRNA) para *SLFN5*.

#### 4.2. Expresión de Schlafen5 en tissue arrays comerciales de tejido gástrico

Al realizarse la tinción in-muno-histoquímica de Schlafen5 en los *tissue arrays* comerciales de tejido gástrico, la tinción estromal de Schlafen5 se vió incrementada con respecto a la severidad de las LPGs según la cascada de Correa (GNA, GCA, MI) y disminuyó en el CG. Sin embargo, la intensidad de la tinción fue baja en la mayoría de los casos quizás debido a que las muestras provienen de gastrectomías en lugar de gastroscopias, en muchas el tejido está dañado o no posee mucosa gástrica para ser evaluada. Considerando, además, que no se posee información clínica de estas

muestras, se decidió analizar también muestras del estudio de seguimiento de LPGs de Soria y considerar exclusivamente los resultados estadísticos de estas últimas.

### **4.3. Controles de la tinción inmuno-histoquímica de Schlafen5**

La mucosa gástrica normal mostró una tinción completamente negativa en la mayoría de las muestras (Figura 46A). Los controles negativos (sin anticuerpo primario) también mostraron una ausencia de tinción (Figura 46B). Se observó una tinción falsa positiva en el citoplasma de las glándulas gástricas que no se tuvo en cuenta en los análisis debido a que esta proteína solo se expresa en núcleo.

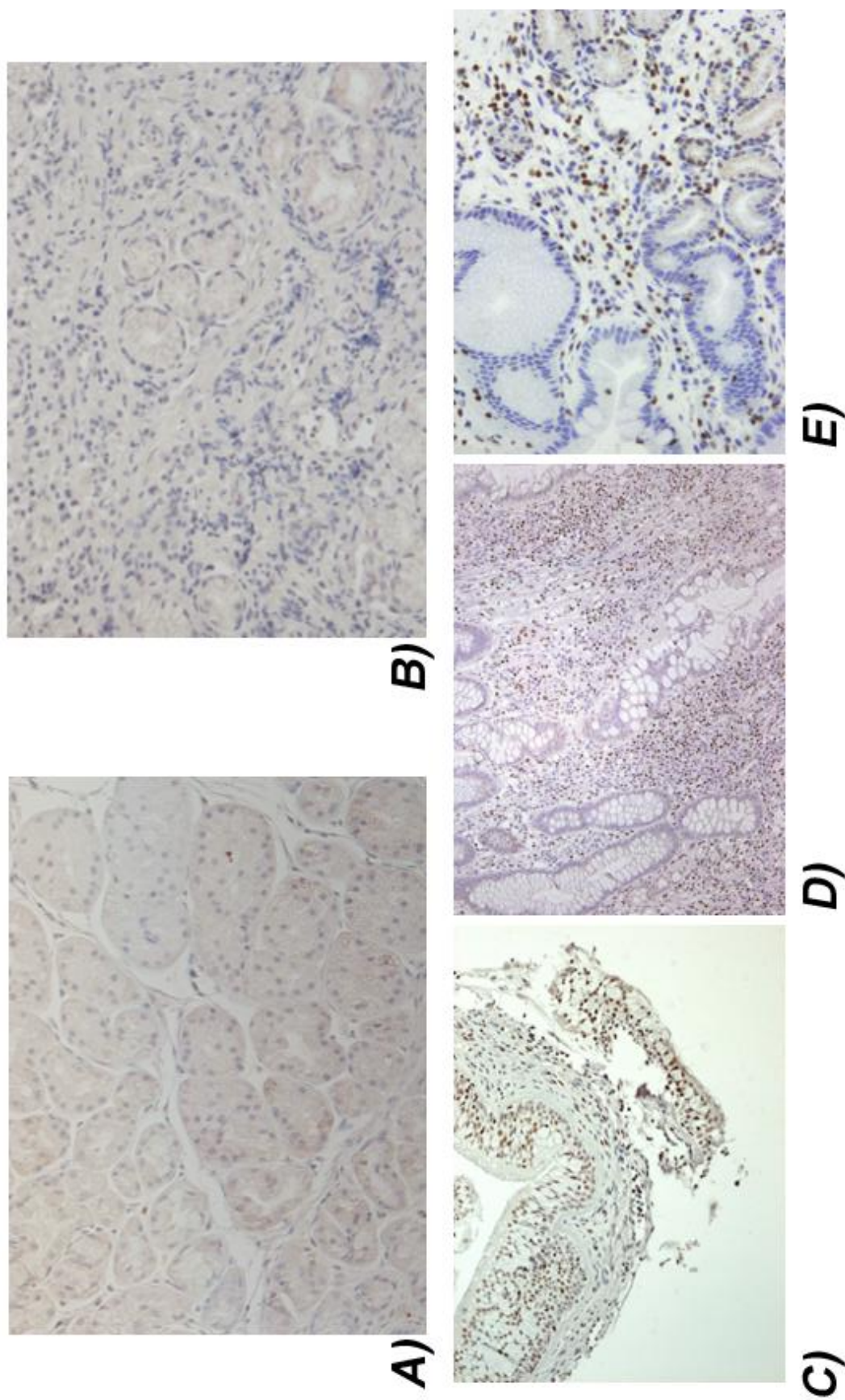
En los controles positivos de nasofaringe e íleon con enfermedad de Crohn (Figura 46C-D) se tiñen los núcleos de células del estroma y en menor medida epiteliales. Los resultados de la tinción de Schlafen5 con el anticuerpo ab121537 (Abcam, Reino Unido) se corroboraron con el anticuerpo HPA 017760 (Sigma-Aldrich, EUA) <sup>88</sup> en algunas muestras de LPGs, demostrando que reconoce a esta proteína de la misma manera y validando los resultados del primero (Figura 46E).

### **4.4. Resultados cualitativos de la tinción inmuno-histoquímica de Schlafen5**

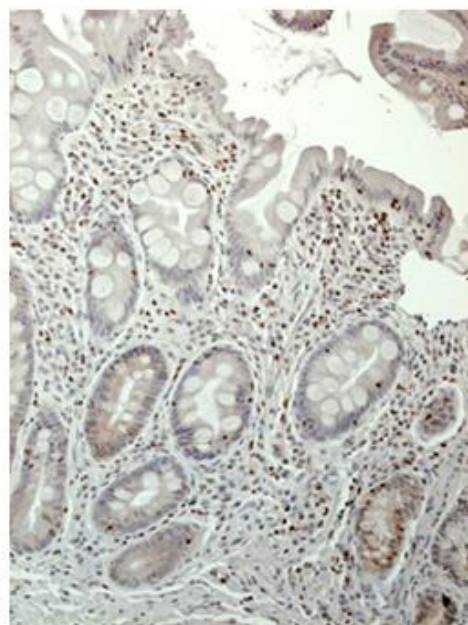
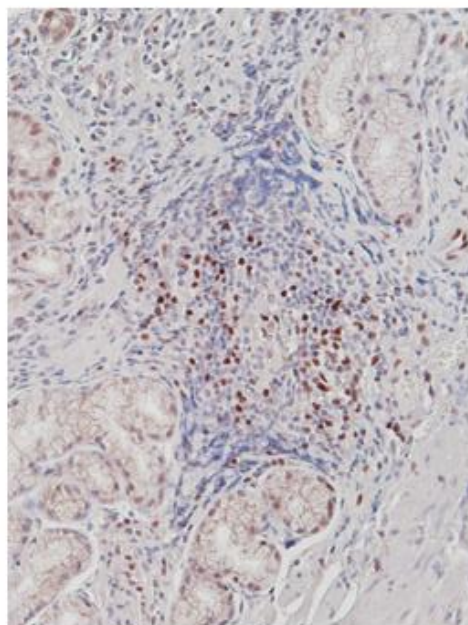
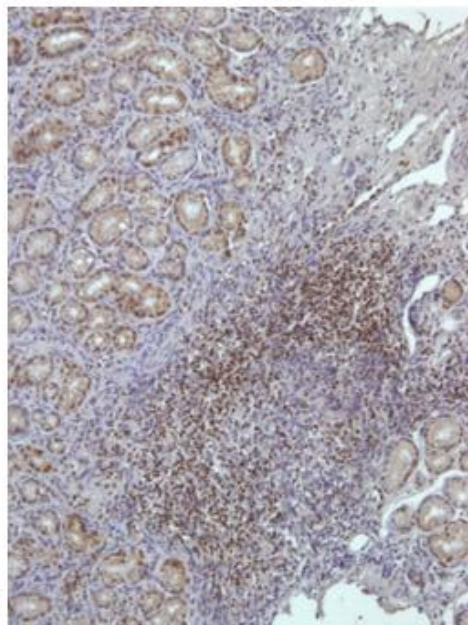
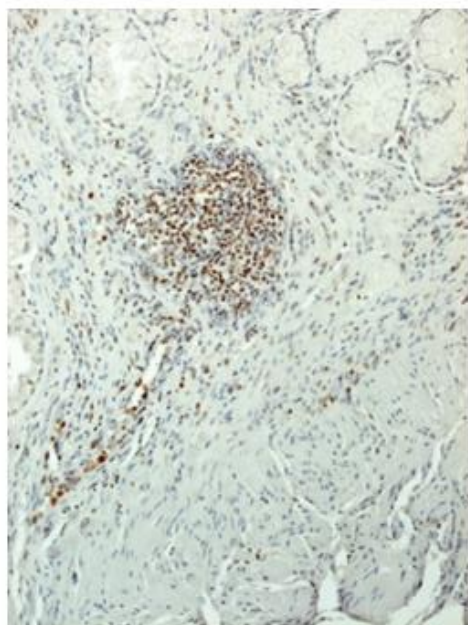
La expresión de Schlafen5 fue nuclear, y mayor en el estroma que en las glándulas gástricas. La figura 47 muestra la tinción de Schlafen5 en los diferentes tipos de muestras estudiadas: GNA, GCA, MIC-No CG, MIC-CG, MII-No CG, MII-CG, CG tipo intestinal y CG tipo difuso. Las tinciones más altas se observan en el estroma de las muestras de metaplasia intestinal (MIC y MII) que progresan a cáncer gástrico (Figura 47 D y F), sobre todo en el estroma periférico de centros germinales constituidos por grandes cantidades de células T y mieloides que expresan Schlafen5.

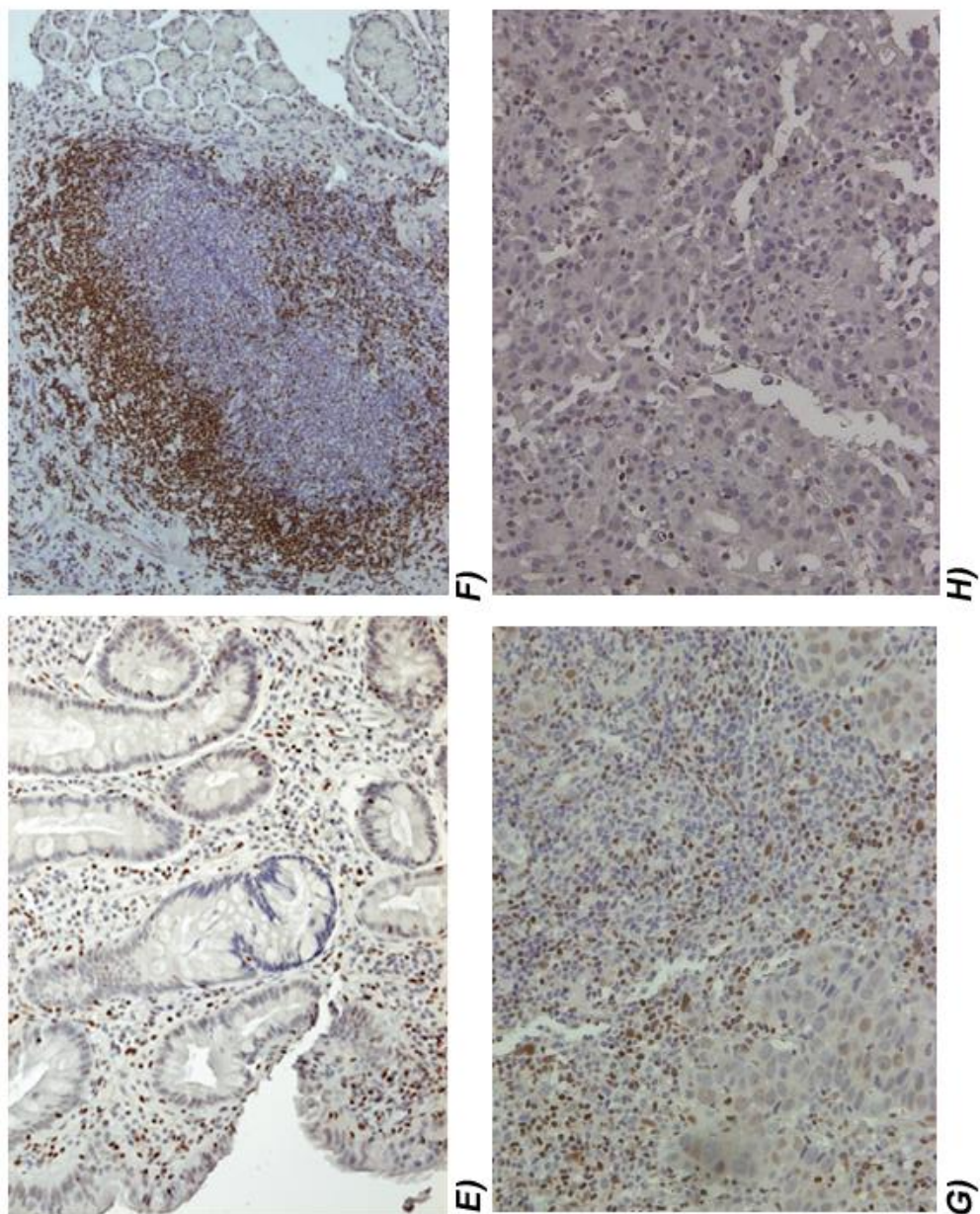
### **4.5. Resultados estadísticos de la tinción inmuno-histoquímica de Schlafen5**

Hay una tendencia general de aumento de la expresión de Schlafen5 en el estroma gástrico de acuerdo con la severidad de las lesiones precursoras y su posterior disminución en el cáncer gástrico (Figura 48 y Tabla 45). Sin embargo, la expresión de Schlafen5 en glándulas gástricas es, en general, más baja que en el estroma y no se observó el patrón típico de incremento de expresión respecto a la cascada de Correa observado en el estroma (Tabla 45). Teniendo en cuenta lo anterior y que las células epiteliales no expresan naturalmente esta proteína, decidimos no analizar estadísticamente dicha localización anatómica.



**Figura 46.** Inmuno-histoquímica de Schlafen5. A) Mucosa gástrica normal. B) Control negativo. C) Control positivo (nasofaringe). D) Control positivo (íleon) con enfermedad de Crohn. E) Validación usando el anticuerpo HPA 017760 (Sigma-Aldrich, EUA).  
Imágenes a 200X.



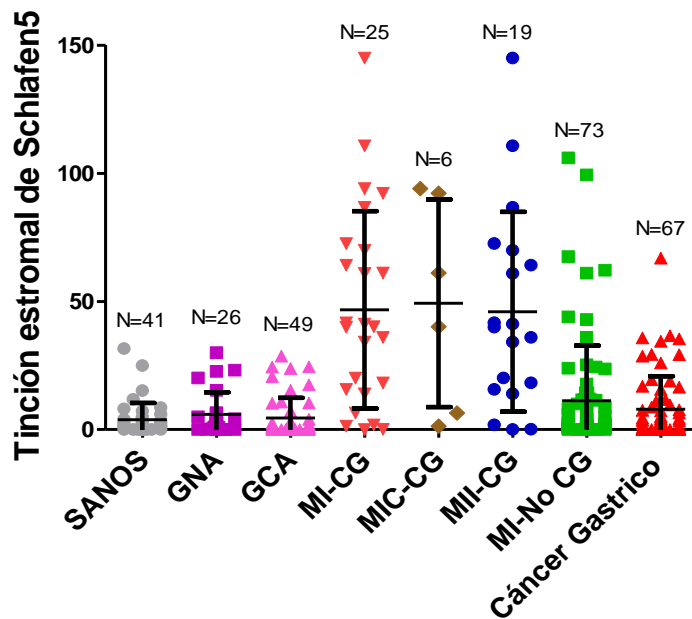


**Figura 47.** Inmuno-histoquímica de la proteína Schlafen5. A) Gastritis no atrófica. B) Gastritis crónica atrófica. C) MIC que no progresa a CG. D) MIC que progresa a CG. E) MII que no progresa a CG. F) MII que progresa a CG. G) CG tipo intestinal. H) CG tipo difuso. Imágenes a 200X.



Tabla 45. Cuantificación de la tinción estromal y glandular de Schlafen5 mediante el score H en los grupos analizados.										
Diagnóstico Histológico <sup>a</sup>	N <sup>b</sup>	Edad <sup>c</sup> (% ± DE)	Sexo (% hombres)	Infección por <i>H. pylori</i> (% +)	Score H estromal de Schlafen5 <sup>d</sup> (media ± mediana)	Score H glandular de Schlafen5 (media ± DE)	Score H estromal de Schlafen5 categorizado			
						Negativo			Débil +	Moderado +
Normal	41	42.5 (18.8)	51.2	0	3.87 (1.25)		37 (90.24)	4 (9.76)	0 (0)	0.89 (0.00)
GNA	26	44.2 (12.3)	46.1	88.5	5.84 (2.53)		21 (80.77)	5 (19.23)	0 (0)	8.54 (0.42)
GCA	49	48.5 (12.6)	38.8	49.0	4.56 (0.84)		40 (81.63)	9 (18.37)	0 (0)	5.15 (0.00)
MIC	44	53.1 (10.5)	45.4	63.6	19.73 (5.73)		25 (56.82)	18 (40.91)	1 (2.27)	14.78 (5.82)
MIC-CG	6	62.0 (6.6)	50	33.3	49.33 (50.68)		2 (33.33)	4 (66.67)	0 (0)	23.23 (14.82)
MIC- No CG	38	51.7 (10.4)	44.7	68.4	15.05 (2.53)		23 (60.53)	14 (36.84)	1 (2.63)	13.41 (5.82)
MII	54	58.3 (12.8)	50	51.8	20.95 (5.11)		32 (59.26)	20 (37.04)	2 (3.7)	7.69 (0.00)
MII-CG	19	66.5 (13.6)	63.2	42.1	45.94 (40.00)		3 (15.79)	14 (73.68)	2 (10.53)	7.99 (1.00)
MII-No CG	35	53.8 (9.9)	42.9	57.1	7.39 (1.14)		29 (82.86)	6 (17.14)	0 (0)	7.52 (0.00)
MI	98	55.9 (12.1)	48	57.1	20.34 (31.8)		57 (58.04)	38 (38.97)	3 (2.98)	10.7 (1.7)
MI-CG	25	65.4 (12.4)	60	40	47.63 (40.18)		5 (24.56)	18 (70.17)	2 (5.26)	11.6 (1.48)
MI-No CG	73	52.7 (10.1)	43.8	63.0	11.22 (2.00)		52 (71.69)	20 (26.99)	1 (1.31)	10.4 (1.00)
CG	67	72.6 (11.0)	61.2	25.4	7.9 (1.38)		51 (76.12)	16 (23.88)	0 (0)	3.22 (0.00)

<sup>a</sup> Diagnóstico histológico al reclutamiento. <sup>b</sup> Tamaño de muestra. <sup>c</sup> Edad del paciente en el diagnóstico al reclutamiento. <sup>d</sup> Score H calculado según %Células+ con intensidad Baja x 1 + %Células+ con intensidad Media x 2 + %Células+ con intensidad Alta x 3 ± mediana. <sup>e</sup> Categorización del score H según 10>scoreH≥0 (Negativo), 100≥scoreH≥10 (Débil Positivo), 200≥scoreH≥100 (Moderado Positivo) y 300≥scoreH≥200 (Fuerte Positivo) <sup>191</sup>.



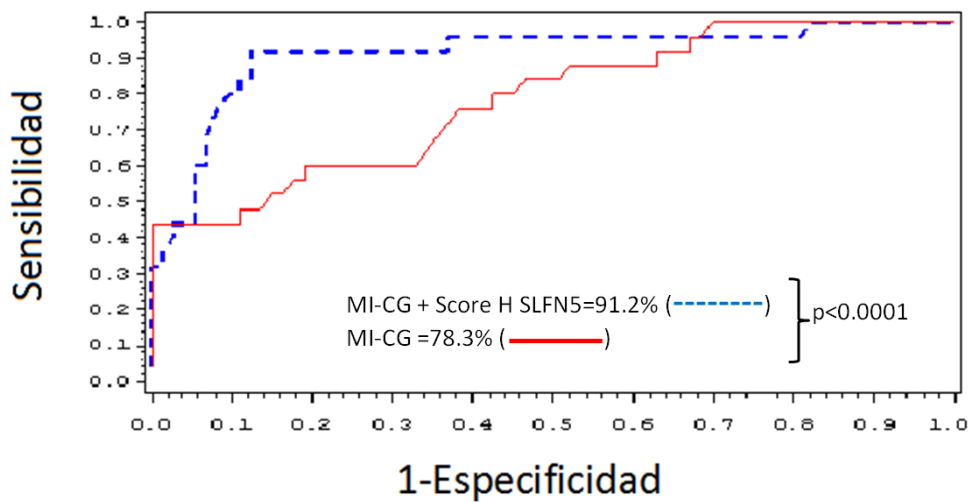
**Figura 48.** Representación gráfica de la tinción estromal de Schlafen5. Corresponde a los resultados de la tabla 45.

La prueba de Kruskal-Wallis muestra diferencias estadísticamente significativas en la expresión estromal de Schlafen5 (score H) entre los distintos diagnósticos histológicos (Normal, GNA, GCA, MIC, MII y CG) con  $X^2=17.75$  y  $p=0.0033$ . Además, la prueba de Mann-Whitney muestra que la expresión de Schlafen5 es significativamente mayor en la metaplasia intestinal que progresa a CG (MI-CG) respecto a la que no progresa (MI-No CG) con  $U=322.5$  y  $p<0.0001$  (Tabla 45). También se obtiene un resultado significativo al comparar los subtipos de MI que progresan (MIC/MII-CG) respecto a los que no progresan (MIC/MII-No CG) ( $p<0.0001$ ). A continuación, categorizamos el score H estromal de Schlafen5 según  $10 > \text{scoreH} \geq 0$  (Negativo),  $100 \geq \text{scoreH} \geq 10$  (Débil Positivo),  $200 \geq \text{scoreH} \geq 100$  (Moderadamente Positivo) y  $300 \geq \text{scoreH} \geq 200$  (Fuerte Positivo)<sup>191</sup>. Observamos que el grupo MII-CG posee más muestras y mayores valores de score H en las categorías de Débil+ (73.68) y Moderado+ (10.53) en comparación con el grupo MIC-CG (Débil+ 66.67 y Moderado+ 0) (Tabla 45).

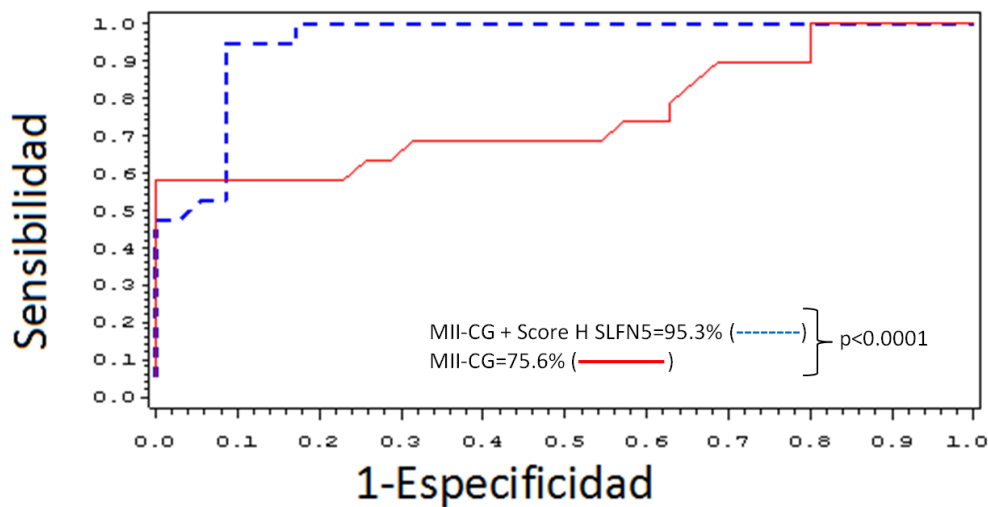
La conversión del score H estromal de Schlafen5 en una variable categórica<sup>191</sup> permitió construir dos modelos de regresión logística en el grupo de MI-CG (MIC-CG + MII-CG), incluyendo o no el score H categórico a las variables sexo, edad y diagnóstico histológico al reclutamiento. Las variables que predicen significativamente la progresión desde la MI al CG son el score H estromal (OR 18.1, IC95% 4.14-79.14,  $p=0.0001$ ), edad (OR 1.13, IC95% 1.05-1.21,  $p=0.0010$ ) y diagnóstico histológico (OR 4.65, IC95% 1.15-18.83,  $p=0.031$ ). El área bajo la curva del grupo MI-CG sin añadir la variable score H estromal de Schlafen5 es  $AUC=0.783$ ; sin embargo, al añadir la misma incrementa el  $AUC=0.912$ , siendo estadísticamente

significativa la diferencia entre ambos modelos logísticos por el test de razón de verosimilitud ( $p < 0.0001$ ) (Figura 49).

Al realizar el mismo análisis solamente en el grupo MII-CG debido al mayor riesgo de progresión de la MII respecto a la MIC, encontramos que el score H también predice significativamente la progresión (OR 71.35, IC95% 7.14-712.88,  $p = 0.0003$ ). En este caso el área bajo la curva del grupo MII-CG sin añadir la tinción de Schlafen5 es  $AUC = 0.756$ , y al añadir la misma aumenta a  $AUC = 0.953$ , siendo estadísticamente significativa la diferencia entre ambos modelos por el test de razón de verosimilitud ( $p < 0.0001$ ) (Figura 50). Este análisis no fue realizado en el grupo MIC-CG debido a su bajo tamaño de muestra ( $N = 6$ ).



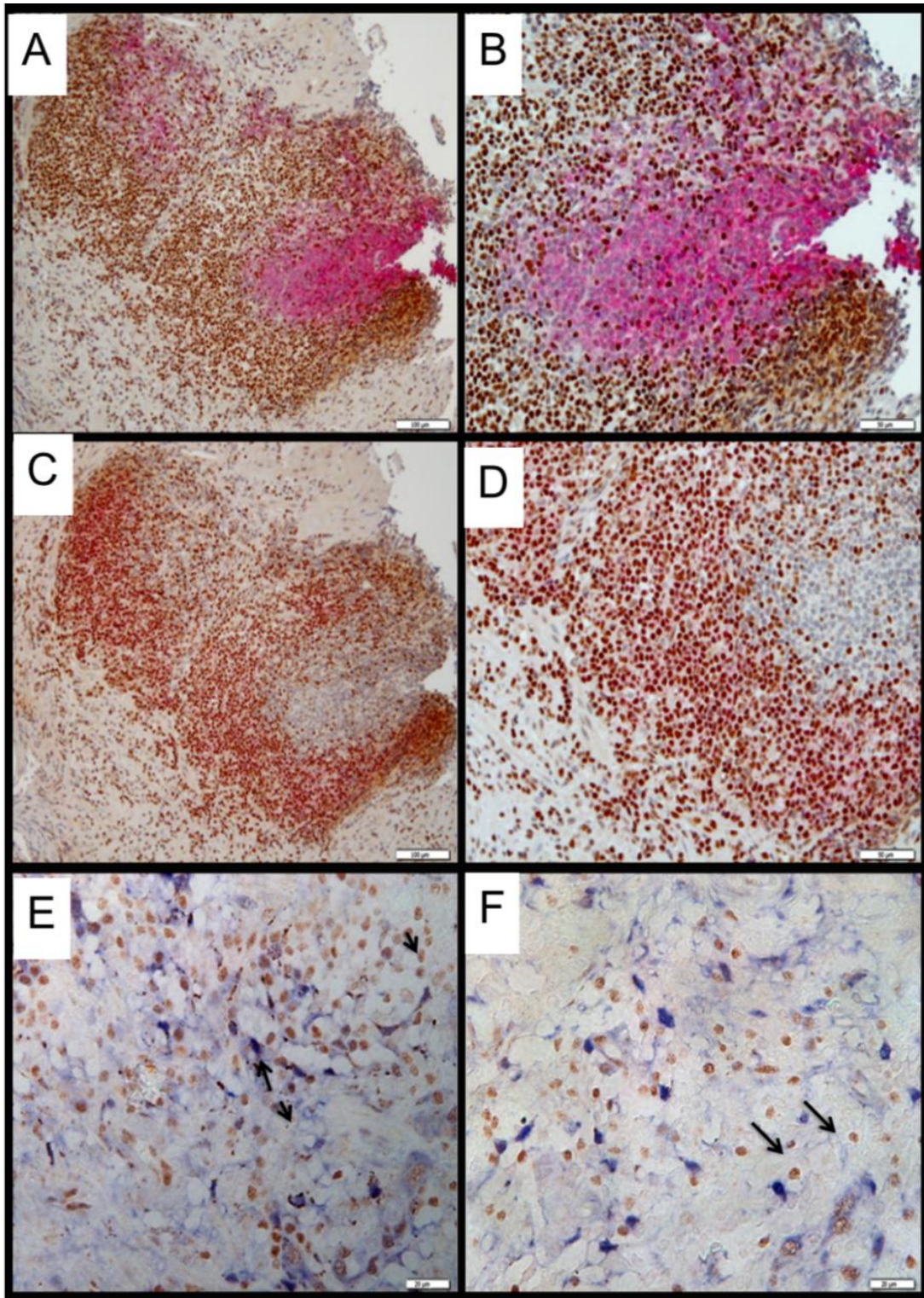
**Figura 49.** Curvas de sobrevivencia ROC para los modelos MI-CG con adición de Score H estromal de Schlafen5,  $AUC = 91.2\%$  (-----) y MI-CG sin adición de Score H estromal de Schlafen5,  $AUC = 78.3\%$  (—————).



**Figura 50.** Curvas de sobrevivencia ROC para los modelos MII-CG con adición de Score H estromal de Schlafen5, AUC=95.3% (-----) y MII-CG sin adición de Score H estromal de Schlafen5, AUC=75.6% (—).

#### 4.6. Doble marcaje de Schlafen 5 y CD2, CD20 o Mac2

Teniendo en cuenta que los miembros de la familia de genes Schlafen se expresan de forma natural en células T murinas, pero no se sabía con precisión qué tipo de células del estroma produce Schlafen5 en la mucosa gástrica, se realizó un doble marcaje con Schlafen5 y CD2/CD20/MAC2, marcadores de superficie de células T, células B y mieloides, respectivamente (Figura 51). Se encontró que Schlafen5 no se expresa en las células B (Figura 51 A-B), sino que se produce principalmente por las células T (Figura 51 C-D) y en menor medida por macrófagos u otras células mieloides (Figura 51 E-F).



**Figura 51.** Doble marcaje de Schlafen5 con marcadores de superficie. A-B) CD20 células B. C-D) CD2 células T. E-F) MAC2 células mieloides. Schlafen5 marca los núcleos con color pardo, CD20 y CD2 con rosado y Mac2 con azul. A la izquierda magnificación 200X y a la derecha 400X.

## **DISCUSIÓN**



## 1. Estudio de asociación de genes de señalización de *Helicobacter pylori* y riesgo de cáncer gástrico

Una importante limitación de la mayoría de estudios de asociación genética en CG es que no analizan la asociación con los subtipos anatómicos, cardias y no cardias, ni con los subtipos histológicos intestinal y difuso, sino que utilizan la entidad global de cáncer gástrico, la cual es menos informativa desde el punto de vista clínico y biológico debido a diferencias moleculares, clínicas y epidemiológicas entre los distintos subtipos<sup>46 10 1</sup>. Adicionalmente, las muestras provienen de estudios caso-control hospitalarios y, en algunos estudios, los polimorfismos no son escogidos teniendo en cuenta el patrón de desequilibrio de ligamiento, por lo que no pueden detectar asociaciones a nivel haplotípico.

Teniendo en cuenta las anteriores limitantes, el objetivo de este estudio ha sido evaluar si la variabilidad en los genes *CD14*, *NOD2*, *TLR4* y *NFKB1*, pertenecientes al mecanismo de infección de *H. pylori*, se asocia estadísticamente con subtipos histológicos y anatómicos de CG. Este estudio es tipo caso-control anidado en una cohorte prospectiva de base principalmente poblacional (EPIC-Eurgast). Se han analizado subtipos anatómicos e histológicos del CG, y la selección de tagSNPs en genes humanos ha permitido aumentar la cobertura de la variabilidad genética de la región analizada y realizar el análisis de la asociación con los principales haplotipos definidos por los distintos bloques de desequilibrio de ligamiento de la región. Todos los SNPs analizados se encuentran en equilibrio de Hardy-Weinberg, lo cual indica que no existen errores de genotipado o presencia de fuerzas evolutivas actuando sobre los genes estudiados.

### 1.1. *NOD2*. Asociación genética y plausibilidad biológica

Las asociaciones más significativas del estudio se observan con los SNPs de *NOD2* rs7202124 y rs2111235, los cuales manifiestan asociaciones inversas con el riesgo de CG no cardias aún después de aplicar la corrección para comparaciones múltiples. La variante rs7202124 es un eQTL en *cis* que controla el nivel de expresión de *NOD2* en monocitos por lo que es posible que dicha asociación genética refleje este hecho. El mapeo de eQTLs es una técnica estadística que correlaciona los genotipos de polimorfismos genéticos con la expresión de ARNm con el objetivo de localizar variantes genéticas que regulan la expresión de cada transcrito<sup>228</sup>.

El análisis de haplotipos corrobora los resultados a nivel de SNPs individuales. De este modo, la combinación GTTCGC formada por los alelos con asociación inversa con el CG no cardias (G del rs7202124, T del rs2111235) muestra también una asociación de este tipo con el CG no cardias. Dicho haplotipo presenta una frecuencia >10% (0.135) en los casos y un p-valor significativo (p=0.005), aspectos



que permiten confiar en la validez de dicha asociación. A pesar de que estos mismos SNPs de *NOD2* no fueron asociados con el subtipo intestinal después de aplicar el test de corrección para comparaciones múltiples y el de heterogeneidad, el mismo haplotipo GTCGC (frecuencia 0.141) también se asocia negativa y nominalmente ( $p=0.031$ ) con el CG intestinal. Lo cual indica la tendencia a la asociación de estos SNPs y el análisis haplotípico con mayor potencia es capaz de detectar la asociación.

NOD2 es miembro de la familia NLRs, los cuales son receptores de reconocimiento de patrones moleculares (PRRs, del inglés *pattern recognition receptors*). La mayoría de los PRRs son receptores de membrana con un dominio N-terminal de reclutamiento de caspasas (CARD), un dominio central de unión a nucleótido y repeticiones ricas en leucina en el extremo C-terminal (LRRs)<sup>229</sup>. NOD2 se expresa en monocitos/macrófagos, células dendríticas<sup>229</sup> y se sobre-expresa en la mucosa gástrica debido a la infección por *Helicobacter pylori*<sup>133</sup>.

NOD2 participa en el reconocimiento del dipéptido muramilo (MDP), un producto de degradación del peptidoglicano, componente de la pared celular de *Helicobacter pylori*. A continuación, sufre un cambio conformacional que provoca su oligomerización y la activación de MAPK quinasas ERK-1, ERK-2 y JNK, lo cual conlleva a la sobre expresión de citoquinas pro-inflamatorias e inducción de la respuesta Th2. Como consecuencia de la infección bacteriana NOD2 también induce autofagia, consistente en que porciones del citoplasma son encapsuladas por membrana para formar el autofagosoma, el que se fusiona con lisosomas para degradar a los microorganismos<sup>229</sup>.

Debido a las razones anteriores, NOD2 participa en la defensa contra patógenos en el tracto gastrointestinal. Se considera que esta función es sinérgica respecto a los TLRs y que NOD2 juega un rol crítico cuando la señalización mediada por TLRs es reducida como en el intestino, por la continua exposición a bacterias comensales. La señalización de NOD2 en células dendríticas promueve diferenciación de células T CD4+ en patrones Th2 y Th17, y la expresión de MHC clase II y anticuerpos de clase IgG<sup>229</sup>. Estas funciones de NOD2, como sensor de la infección por *H. pylori* e inductor de respuesta Th2 (asociada con gastritis más leve<sup>33</sup>) y humoral está en consonancia con la asociación inversa observada de este gen con el cáncer gástrico no cardias. El cual, a diferencia del cardias se encuentra asociado con la infección por *H. pylori*<sup>230</sup>.

No se han encontrado estudios previos de asociación de los tagSNPs de *NOD2* rs7202124 y rs2111235 con el CG, quizás debido a que los trabajos previos no se basaron en una estrategia de selección de tagSNPs. Sin embargo, otros SNPs de *NOD2* han sido asociados con CG como la variante de riesgo rs2066842 (P268S) en población polaca<sup>231</sup>. No fue posible analizarla debido a la mala separación de los

clusters genotípicos. Asimismo, la variante rs3135500 incrementa el riesgo de CG en la población china<sup>232</sup>, un hallazgo no confirmado en el presente estudio. Sin embargo, según el análisis de Haploview, el haplotipo asociado inversamente con el CG en este estudio está formado por las variantes comunes de estos dos polimorfismos. En una población italiana, los SNPs de *NOD2* rs2066844 y rs5743293 incrementan la susceptibilidad a GC<sup>233</sup>, pero estos SNPs no se incluyen en la versión de HapMap usada en este trabajo para el análisis en Haploview.

Cuando se estratificó el CG no cardias atendiendo a la presencia del factor de virulencia *cagA*, los SNPs de *NOD2* (rs7202124 y rs2111235) permanecieron significativamente asociados con CG no cardias en los individuos *cagA+* y apareció una nueva asociación, el rs5743289 en el grupo *cagA-*. Sin embargo, el número de casos en este grupo es demasiado bajo para sacar alguna conclusión.

Los SNPs genotipados fueron seleccionados por ser tagSNPs, es decir SNPs marcadores de bloques de haplotipos en desequilibrio de ligamiento con otros SNPs que pudieran tener un efecto funcional, por lo que es posible que dichos tagSNPs no tengan un efecto funcional en sí mismo. Algunas maneras de saberlo son mediante estudios de mapeo fino o funcionales, que no fueron realizados en el presente trabajo.

Teniendo en cuenta la anterior limitante, se ha empleado el programa Variant Effect Predictor para la predicción de efectos funcionales encontrando que los SNPs de *NOD2* asociados a CG rs7202124, rs2111235 y rs5743289 están localizados en regiones intrónicas correspondientes a transcritos (ENST00000568993, ENST00000527070, ENST00000524712, ENST00000527052, ENST00000529633, ENST00000534067) sujetos al mecanismo de degradación del ARNm NMD (*Nonsense Mediated Decay*). Un mecanismo de vigilancia que detecta diversos rasgos anómalos en el ARNm como codones prematuros de parada, *exon skipping*, retención de intrones, anomalías en el *splicing* alternativo etc., con el fin de promover la degradación de transcritos codificantes para proteínas aberrantes para la protección celular. Es posible que el gen *NOD2* presente un codón prematuro de parada de la traducción a más de 50 nucleótidos upstream del exón final que active el mecanismo NMD. Este proceso incrementa la severidad del cáncer cuando degrada proteínas truncadas que retienen parte de su actividad biológica nativa<sup>234</sup>.

En este sentido es relevante llamar la atención a que existe un incremento significativo de todos los SNPs asociados a CG sujetos al mecanismo NMD en comparación con aquellos no asociados a CG (test exacto de Fisher  $p=0.0026$ ) y que la mayoría se localizan en *NOD2*. Sin embargo, se ha reportado previamente la existencia de una isoforma corta de *NOD2* (*NOD2-S*) debida a la presencia de un codón de parada en el exón 4 pero no esta sometida a degradación por el

mecanismo NMD<sup>235</sup>. El programa Pupasuite no reconoció los alelos de dichos SNPs como que muestren un reconocimiento diferencial por *enhancers* (ESE) o silenciadores (ESS) del *splicing* alternativo, fenómeno que puede crear un exón en el interior de un intrón propiciando la degradación NMD, como se ha reportado en diabetes mellitus<sup>236</sup>. Basado en lo cual, no se puede concluir una función de estos SNPs relacionada con el mecanismo NMD.

## 1.2. CD14. Asociación genética y plausibilidad biológica

Los SNPs de *CD14* rs778588 y rs1583005 se asocian positivamente con el CG tipo cardias, mientras que rs11167532 y rs2569190 lo hacen inversamente; a su vez, rs778588 se asocia con un mayor riesgo de CG intestinal. Dichas asociaciones permanecen significativas después del test de corrección para comparaciones múltiples. Coincidentemente, el haplotipo GCCCGT formado por los alelos de riesgo del CG tipo cardias (C del rs778588 y T de rs1583005) así como por los alelos salvajes de rs11167532 y rs2569190 se asocia positivamente con el CG tipo cardias e intestinal, con un p-valor más significativo (p=0.0004) para el tipo cardias respecto a las significaciones obtenidas con los SNPs individuales en este grupo. Al igual que en *NOD2* este haplotipo presenta una frecuencia >10% (0.193-0.212) en los casos, p-valores bien significativos (p=0.006 y 0.0004) e interroga el 81% de la variabilidad genética de la región, varios aspectos que permiten confiar en dicha asociación.

CD14, al igual que *NOD2*, pertenece a la familia de los receptores de reconocimiento de patrones y es un co-receptor de los TLRs (*toll-like receptors*). El mismo existe como receptor de membrana (CD14m) y proteína soluble, el primero se expresa principalmente en monocitos y en macrófagos que se acumulan en la mucosa gástrica de los pacientes infectados por *H. pylori*<sup>237</sup>. CD14m unido a TLR4 y MD2 interactúa con el LPS de la pared celular de *H. pylori*. Posteriormente, CD14 presenta y transfiere el LPS al complejo TLR4-MD2 para el inicio de la transducción de señales que activa a NFκB1 e induce la transcripción de citoquinas pro-inflamatorias<sup>238</sup>. CD14 es también un promotor de la transición epitelio-mesénquima mediada por TNF en CG<sup>239</sup>.

Los estudios publicados de *CD14* se han centrado en el SNP rs2569190 (-260 C/T), que se encuentra localizado colindante al sitio de unión al factor de transcripción Sp1, el cual influye en la expresión de *CD14*<sup>240</sup>. A diferencia de estos resultados, estudios en poblaciones de Polonia, EUA<sup>105</sup> y Taiwán<sup>241</sup> no encontraron asociación de esta variante con el CG no cardias, un estudio en Japón encontró una asociación inversa con el subtipo intestinal<sup>242</sup> y un meta-análisis solo encuentra asociación para el SNP rs2569190 (-260C/T) en pacientes infectados por *H. pylori*<sup>106</sup>. Por lo que sabemos, no se ha realizado ningún estudio con respecto a la asociación de *CD14* con GC cardias.

Muchas de las variantes genéticas asociadas a enfermedades complejas encontradas en estudios de asociación resultan ser eQTLs, los cuales producen pequeños cambios en los niveles de expresión de ARNm y por ende proteicos que pueden contribuir a una enfermedad <sup>157</sup>. Teniendo esto en cuenta, tagSNPs asociados al CG que sean eQTLs pudieran ser SNPs que influyeran en la asociación, aunque no es posible asegurarlo ya que son predicciones obtenidas en tejidos no gástricos y, además, estos eQTLs regulan la expresión de otros genes diferentes de los estudiados. Por lo que estas predicciones son preliminares y no permiten determinar que dichos SNPs sean funcionales.

En este sentido, la búsqueda de eQTLs entre todos los SNPs encontró que las variantes asociadas al CG de la región de *CD14* rs11167532, rs1583005, rs2569190, y rs778588 han sido encontradas previamente como reguladores de la expresión génica de los genes *SRA1* <sup>156</sup> y *WDR55* <sup>157</sup> en monocitos. Este tipo celular se acumula en el estroma gástrico a consecuencia de la inflamación generada por *Helicobacter pylori*, transformándose a macrófagos que incrementan la inflamación por un aumento de la generación de radicales libres <sup>27</sup>. Respecto a la variante rs2569190 (-260T) se ha descrito experimentalmente que regula los niveles de CD14 presentes en la membrana de monocitos sanguíneos, estando asociado los homocigotos del alelo de riesgo con el incremento de la expresión de CD14 <sup>245</sup>, lo cual está en concordancia con que sea identificado como un eQTL en estudios de este tipo. Existen múltiples casos en los que tagSNPs de estudios GWAS asociados con cáncer son eQTLs como en cáncer de próstata <sup>244</sup> y colorectal <sup>245</sup>. Por lo que se puede sugerir que la asociación identificada en *CD14* (rs2569190) sea mediada por un control de la expresión de CD14 en monocitos ejercido por este SNP, para el resto de las variantes es necesario hacer análisis o experimentos *in vitro* adicionales que permitan aclarar dicho efecto.

Debido a que la infección por *H. pylori* no se asocia con CG del cardias <sup>230</sup> y, sin embargo, CD14 es receptor del lipopolisacárido de bacterias gram-negativas como *H. pylori*, es posible que otros mecanismos puedan explicar la asociación de *CD14* con el CG del cardias. En este sentido, CD14 podría actuar como receptor para el virus de Epstein-Barr, una infección que sí está asociada con el CG del cardias <sup>246</sup>, ya que se ha planteado la posible existencia de un nuevo receptor diferente a CR2, el receptor del virus de Epstein-Barr <sup>247</sup>. CD14 es un candidato por ser un receptor poli-específico que interactúa con el dipéptido muramilo del LPS, peptidoglicano, lipoarabinomannan y fosfolípidos <sup>248</sup>. Lamentablemente, no se dispone de datos sobre la infección por el virus de Epstein-Barr en la población estudiada. Una evidencia que pudiera apoyar la anterior aseveración es el hecho que CD14 no se asocia en nuestra población con CG no cardias, localización anatómica fuertemente asociada con la infección por *H. pylori*.

### 1.3. *TLR4* y *NFKB1*. Asociación genética y plausibilidad biológica

A pesar de que no alcanzaron significación estadística los SNPs de *TLR4* y *NFKB1* después de aplicar el tests de corrección para comparaciones múltiples, el análisis haplotípico sugiere que estos genes contienen combinaciones que incrementan el riesgo para el CG del cardias y no cardias, respectivamente. Sus frecuencias haplotípicas en los casos son cercanas al 10% en ambos genes (*TLR4*: 0.112, *NFKB1*: 0.095) y los valores de significación no son nominales (*TLR4*:  $p=0.016$ , *NFKB1*:  $p=0.009$ ), por lo que añaden información de asociación respecto a la detectada por el análisis de SNPs individuales. En el caso de *TLR4* el haplotipo CGAGACG contiene al alelo común del único SNP asociado de este gen, rs10491851; en este caso inversamente con CG global.

La mayoría de los estudios previos sobre *TLR4* y CG se centran en los SNPs no sinónimos rs4986790 y rs4986791, cuyas variantes alélicas manifiestan una menor respuesta inmune frente al LPS y en consecuencia aumentan el riesgo de desarrollar LPGs y CG<sup>109</sup>. El rs4986790 ha sido asociado con CG no cardias en poblaciones de Polonia y EUA<sup>109</sup>, pero no se ha confirmado en mexicanos<sup>249</sup>; mientras que rs4986791 también se ha asociado positivamente con el CG intestinal en Italia<sup>250</sup>. Sin embargo, las frecuencias de los alelos de riesgo eran muy pequeñas (0.057 y 0.061 en los controles) en la población estudiada, y sólo se pudo estimar la asociación bajo el modelo co-dominante, sin resultados significativos. Dicha conclusión es similar a la de los estudios realizados en Méjico<sup>250</sup> y Japón<sup>142</sup>. A diferencia de estos resultados, un meta-análisis concluye que rs4986790 y rs4986791 incrementan el riesgo de CG en poblaciones europeas<sup>111</sup>.

*TLR4* también pertenece a la familia de los PRRs y junto a CD14 y MD-2 participa en el reconocimiento del LPS de *H. pylori*. Después de su activación las proteínas adaptadoras MyD88, TIRAP, TRAM y TRIF se unen a los dominios TIR (*toll-interleukin (IL)-1 receptor*) del *TLR4* y activan las quinasas IRAK, MAPK y TRAF induciendo la activación de los factores transcripcionales *NFKB1*, AP-1 e IRF. En relación con la falta de asociación a nivel de SNPs individuales de *TLR4* obtenida en este estudio es interesante apuntar que la infección de la línea celular HEK293 con *H. pylori* es capaz de inducir *NFKB* en células transfectadas con *TLR2* o *TLR5* pero no por *TLR4*<sup>251</sup> y que un estudio afirma que el reconocimiento de *H. pylori* por el epitelio gástrico es independiente de *TLR4*<sup>252</sup>.

Respecto a *NFKB1* existió una tendencia a la asociación ya que los SNPs rs765789, así como, rs228611 y rs1585214 se encontraron positiva y negativamente asociados con CG no cardias y con el subtipo difuso; aunque ninguna de estas asociaciones fue significativa después del test de comparaciones múltiples. Sin embargo, el haplotipo TGAGGTGACCAG se encuentra positivamente asociado con el CG no cardias. Esta combinación está formada por el alelo de riesgo A de rs765789 y los alelos salvajes

G de rs228611 y rs1585214 asociados negativamente con CG difuso, siendo coherente la asociación a nivel de SNPs individuales y combinación haplotípica.

NFKB1 es un factor transcripcional de genes de la respuesta inmune, adhesión, diferenciación, proliferación celular, angiogénesis y apoptosis, que se encuentra sobre-expresado en las enfermedades inflamatorias y tumorales. Su función es inhibida por unión a su inhibidor IκB y degradación proteasomal, a su vez la activación de este es regulada por las quinasas de IκB (IKK) activada por interacción con *H. pylori* pero también por muchos otros estímulos como TNF, IL1B, factores de crecimiento, activación de células B o T, radicales libres del O<sub>2</sub> y el N<sub>2</sub>, luz UV, infecciones virales y microbianas. Teniendo en cuenta la diversidad de eventos activadores de NFKB es posible que la falta de asociación con el CG esté relacionada con el hecho de no tener una relación específica con la carcinogénesis gástrica, sino general con muchas enfermedades y ser activada por múltiples estímulos. Lo cual es confirmado por el hecho de que NFKB promueve o inhibe la carcinogénesis en dependencia del tipo celular<sup>253</sup>.

A manera de resumen, se pueden mencionar las ventajas de este estudio de asociación en comparación con otros anteriores. Evalúa la asociación de tagSNPs con ambas localizaciones anatómicas y subtipos histológicos del CG, un aspecto no siempre considerado en estudios previos. También, la cohorte utilizada (EPIC-Eurgast) está constituida por un relativamente elevado tamaño de muestra de sujetos inicialmente sanos que han sido sometidos a un seguimiento prospectivo de muchos años para detectar la aparición de diferentes cánceres<sup>144</sup>. Dichos sujetos están caracterizados para diversos factores de riesgo como son la dieta, tabaquismo e infección por *H. pylori*. También, es necesario señalar sus limitaciones que incluyen el que el número de casos de CG es relativamente pequeño cuando se dividen por subtipos histológicos (intestinal/difuso) y localizaciones anatómicas (cardias/no cardias), con la consiguiente disminución del poder estadístico para detectar un efecto. Además, no se genotiparon SNPs en otros genes (*MD2*, *TLR2/5*, etc.) que pueden interactuar con el *H. pylori* durante la infección, y la cobertura genética de *TLR4*, *NOD2* y *NFKB1* se redujo debido a la imposibilidad técnica de genotipar algunos de los tagSNPs seleccionados.

Basado en los resultados a nivel de SNPs individuales y haplotipos es posible concluir que *NOD2* se asocia inversamente con el CG no cardias e intestinal. Mientras que *CD14* se asocia con un mayor riesgo de los CG del cardias e intestinal. Solamente a nivel haplotípico se obtienen asociaciones de *TLR4* y *NFKB1* con CG del cardias y no cardias respectivamente. A su vez tagSNPs asociados de *CD14* y *NOD2* son eQTLs reguladores de la expresión génica, sin embargo no existen experimentos funcionales en tejido gástrico que lo confirmen y otras variantes en LD con las mismas pudieran ser las funcionales.

## 2. Estudio de asociación genética en el seguimiento de lesiones precursoras gástricas

Este estudio ha explorado si la variación en 141 SNPs correspondientes a 28 genes candidatos de carcinogénesis gástrica se asocia con la evolución (progresión, regresión) de las LPGs. La población (N=559) proviene de un estudio longitudinal de seguimiento de LPGs durante una media de 12 años, con pacientes de 9 hospitales de 5 provincias de España (Madrid, Barcelona, País Vasco, Valladolid y Zaragoza). En dicho estudio el 3.7% del total de los pacientes desarrolló un adenocarcinoma gástrico, siendo la incidencia global 3.09 (2.07-4.6)/1000 personas-año (artículo pendiente de aceptación). Estas tasas son similares a las reportadas en un estudio de seguimiento (media 12.8 años) realizado exclusivamente en la provincia de Soria, donde el 4.8% de todos los pacientes (N=478) evolucionó a CG, con una incidencia de 3.77 (2.51-5.67)/1000<sup>57</sup>. Lo que indica que España es un país de riesgo moderado-bajo de CG en el cual las estrategias de prevención a amplia escala de cáncer gástrico<sup>6</sup> son impracticables por la baja incidencia de la enfermedad. En su lugar, la búsqueda de genes y variantes condicionantes de la evolución pudiera conllevar a identificar pacientes susceptibles de progresar a CG a lo largo del tiempo.

Solo el 17.2% de los pacientes progresan a una LPG de mayor severidad entre el reclutamiento y el final de seguimiento, lo cual explicaría que la mayoría de las asociaciones genéticas significativas fueran obtenidas al comparar el grupo de la regresión con el resto.

Al comparar las variables analizadas al reclutamiento (sexo, edad, consumo de AINES, historia familiar de CG, tabaquismo, infección por *H. pylori* y diagnóstico histológico) entre los grupos (progresión, regresión, estables) la única que mostró diferencias fue el diagnóstico histológico al reclutamiento (Tabla 24). Sin embargo, se reporta en la literatura un mayor riesgo de CG en función de la infección por *H. pylori*, mayor edad y el sexo masculino<sup>1</sup>. Pudiera esperarse que la variable con mayor poder predictivo de la progresión sea la infección por *H. pylori*, pero se conoce que la hipoclorhidria provocada por la misma dificulta su sobrevivencia, por lo que se observa una disminución de la infección a partir de la gastritis atrófica lo cual disminuye las diferencias entre los grupos analizados. En correspondencia con los presentes resultados, en el estudio de asociación de LPGs de Soria se obtuvieron diferencias significativas según el diagnóstico histológico al reclutamiento<sup>140</sup>.

La cascada de Correa plantea que en poblaciones de alta incidencia de CG la tasa de progresión de las LPGs es mayor que la tasa de regresión condicionando un lento movimiento hacia la progresión (Tabla 1)<sup>14</sup>. En poblaciones de bajo riesgo como la nuestra se observó que la mayoría de los pacientes se mantuvieron estables en la

misma lesión (46.3%), seguidos de los que regresaron (36.5%) en comparación con sólo un 17.2% de progresión.

Las principales ventajas de este estudio son que los pacientes pertenecen a distintas regiones españolas, su largo tiempo de seguimiento (media 12 años) y el relativo elevado tamaño de muestra. Su principal limitante es que la toma de biopsias fue más rigurosa al final del seguimiento que al reclutamiento, debido al cumplimiento del protocolo de Sydney de muestreo sólo al final del seguimiento. Sin embargo, esto fue compensado solicitando que cada biopsia al reclutamiento tuviera al menos 3 fragmentos con representación de cuerpo y antro. Adicionalmente, algunos tagSNPs no pudieron ser genotipados reduciéndose la cobertura de la variabilidad de los genes donde fueron seleccionados.

### **2.1. Asociación genética con la evolución de las lesiones precursoras gástricas**

Mediante este estudio se ha pretendido por una parte replicar SNPs previamente asociados con la evolución de las LPGs en el estudio de Soria y, por otra, estudiar si la variabilidad en nuevos genes candidatos se asocia estadísticamente con la evolución en el tiempo de las lesiones precursoras gástricas. Los genes estudiados realizan funciones claves en la fisiología gastrointestinal como se evidencia en los modelos murinos de carcinogénesis gástrica, que se caracterizan por el hecho de que al cabo de un tiempo aparece alguna o varias de las LPGs seguidas del desarrollo de CG.

También, se han genotipado genes de la vía de señalización de *H. pylori*, cuyas proteínas interactúan con esta bacteria durante su infección, influyendo en la severidad de la reacción inflamatoria. La identificación de polimorfismos genéticos asociados con la progresión y regresión es relevante desde el punto de vista clínico, debido a que esta estrategia exploratoria permite identificar y/o confirmar genes que pudieran condicionar la evolución desde las LPGs hacia el CG. De esta manera, mediante posteriores estudios inmunohistoquímicos y funcionales se pudieran identificar biomarcadores de riesgo o protección de la progresión tumoral.

Como resultado de este estudio se asociaron SNPs con un incremento del riesgo de progresión en los genes *CD14*, *DNAH11*, *MAPK3*, *PRKAA1*, *RUNX3* y *TFF1*. Otras variantes localizadas en los genes *CD14*, *CDH1*, *PTGS2*, *TFF1* y *TFF2* se asociaron inversamente con la progresión de las lesiones (Tabla 25 A). El análisis de la regresión mostró que SNPs en los genes *CD14*, *IL1B*, *IL1RN*, *NFKB1*, *TFF1* y *TFF2* se asociaban con una mayor regresión, mientras que aquellos en *CDH1*, *MUC1*, *MUC2* y *NFKB1* se asociaron con una menor regresión (Tabla 25 B). El SNP más significativo y el único que se mantiene así después de aplicar la corrección para comparaciones múltiples es rs10902073 de *MUC2*, el cual es un tagSNP localizado en una región



regulatoria que controla la expresión génica; su asociación inversa con la regresión de las LPGs ya se había observado en el estudio previo de Soria <sup>140</sup>.

Respecto a otros SNPs previamente genotipados en el estudio de Soria, fueron replicadas las asociaciones entre variantes de *CD14* (rs2569190), *TFF2* (rs1079380) y *CD14* (rs5744455) con una menor y mayor probabilidad de progresión respectivamente; *MUC2* (rs10902073, rs10794281) y *TFF1* (rs424694) con menor y mayor probabilidad de regresión de las LPGs respectivamente. Por el contrario, no se confirmaron las asociaciones con *CDH1* (rs1125557), *TFF1* (rs9976977, rs13047838, rs225358), *MUC2* (rs7944723, rs10794293, rs3924453, rs4077759), *IL1RN* (rs2637988) e *IL1A* (rs17561) (Tabla 26). Respecto a los SNPs que habían sido previamente asociados con CG en estudios GWAS, se observó la asociación de rs2285947 de *DNAH11* y rs13361707 de *PRKAA1*, como factores de riesgo de la progresión, lo cual esta en concordancia con el hecho de que su asociación haya provenido de un estadio más avanzado de la enfermedad como es el CG respecto a las LPGs <sup>118 119</sup>.

El análisis simultáneo de los SNPs y haplotipos asociados significativamente con la evolución de las LPGs muestra que determinados haplotipos de *CD14* y *TFF1* tienen mayor probabilidad de progresión y un haplotipo *PTGS2* muestra menor probabilidad de la misma. Por otra parte, haplotipos de *CDH1*, *CDX2* así como *PTGES*, *TFF1*, *MAP3K14* y *PTPN11* se asocian positiva y negativa respectivamente con la regresión de las lesiones precursoras. De manera general, las combinaciones haplotípicas son más significativas que la de los SNPs individuales que contienen, y concordantes con el efecto individual de los mismos, lo que confirma el interés del análisis haplotípico para detectar asociaciones no siempre observables mediante el análisis de SNPs individuales. En un intento de restringir estos resultados a los más confiables se consideraron solo aquellos genes significativos simultáneamente a nivel de SNPs individuales y haplotipos. Evidenciándose que los genes *CD14* y *TFF1* están asociados positivamente mientras que *PTGS2* lo está negativamente con la progresión de las LPGs; mientras que *CDH1*, *CDX2* y *TFF1* se asocian positiva y negativa respectivamente con la regresión de las lesiones precursoras.

El análisis bioinformático mostró que SNPs asociados con la evolución de las LPGs se encuentran localizados en zonas regulatorias, UTR 5', UTR 3', o son eQTLs reguladores de la expresión génica. Aunque hay que tener presente que son predicciones funcionales y que la mayoría de los SNPs se seleccionaron por ser tagSNPs, por lo que su funcionalidad sería, en todo caso, casual. Sin embargo, es relevante destacar que *CD14* es uno de los genes con mayor número de SNPs asociados con la evolución de LPGs, los cuales a su vez fueron encontrados como eQTLs reguladores de la expresión génica, pero los mismos controlan los niveles de expresión de *WDR55* en lugar de *CD14*. Además, no existen diferencias significativas

en la distribución de dichas localizaciones o funcionalidades entre los SNPs asociados y los no asociados a la evolución de las LPGs, por lo que este análisis no evidencia una diferencia entre los mismos en cuanto a su estatus de asociación.

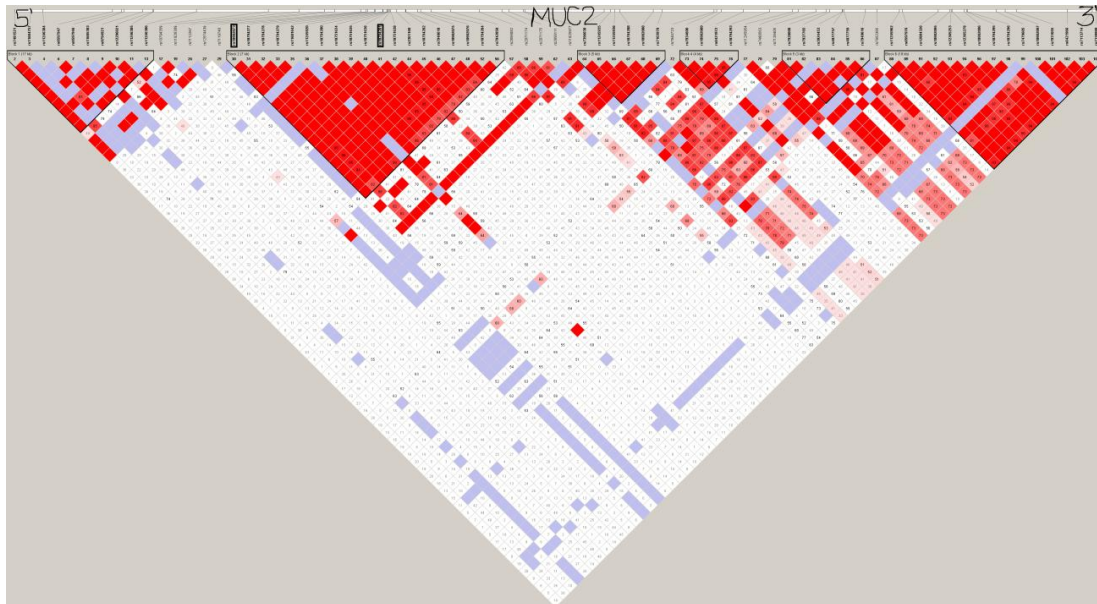
Hasta la fecha han sido publicados solamente cuatro estudios de asociación genética<sup>124 128 134 140</sup> que evalúan la variabilidad en genes candidatos respecto a la evolución (progresión/regresión) de las LPGs a lo largo del tiempo. Es relevante destacar la poca cantidad de los mismos, lo cual es lógico debido a la dificultad de llevar a cabo estudios de seguimiento de lesiones precursoras gástricas. Acorde a nuestra experiencia en el estudio multicéntrico de seguimiento de LPGs, las razones pudieran relacionarse con las dificultades en la búsqueda de información proveniente de historias clínicas antiguas donde los criterios patológicos, de clasificación y muestreo histológico eran diferentes a los de la actualidad, no disponibilidad de tinciones hematoxilina-eosina o bloques de parafina representativos de todas las localizaciones anatómicas del estómago, dificultades en el contacto con pacientes, así como tiempo disponible de los clínicos para realizar el trabajo investigativo.

Uno de los trabajos antes publicados se realizó con muestras del estudio de seguimiento de Soria y se evaluaron SNPs en los genes *MUC1*, *MUC2* y *MUC6*<sup>140</sup>. Los SNPs (rs10794293, rs3924453, rs4077759) en el extremo 3' de *MUC2* se asociaron con un menor riesgo de progresión de las lesiones. Mientras que los localizados en el extremo 5' se asociaron con el aumento (rs2071174) o disminución (rs10902073, rs10794281, rs7944723) de la probabilidad de regresión, indicando un rol de este gen en evitar la progresión de las LPGs a estadios de mayor severidad. El análisis haplotípico fue el más relevante, ya que ambas combinaciones CTCCCG y CATAGAAC de la región 5' y 3' se asociaron positivamente con la regresión e inversamente con la progresión, respectivamente<sup>140</sup>.

Nuestros resultados confirman la asociación inversa de los SNPs rs10902073 y rs10794281 (OR 0.55, IC95% 0.34-0.9) de *MUC2* con la regresión de las LPGs. El haplotipo AC formado por los alelos asociados de estos dos SNPs también está significativamente asociado con la regresión (OR 0.66, IC95% 0.50-0.86), con una frecuencia haplotípica alta (0.274) en los casos y su significación estadística es similar a la de sus SNPs individuales ( $p=0.002$ ). Ambos SNPs se encuentran en un bloque de haplotipos de elevado desequilibrio de ligamiento localizado en el extremo 5' del gen (Figura 52), añadiendo una evidencia de que estas variantes u otras causales de la asociación pudieran encontrarse en las secuencias regulatorias *upstream* que controlan la expresión génica de *MUC2*.

Las mucinas son proteínas altamente glicosiladas (50-80% O-glicosilación) sintetizadas por el epitelio gástrico y son los principales constituyentes del mucus gástrico, encargado de la protección de la mucosa contra bacterias, enzimas

proteolíticas y ácido clorhídrico. La sobre-expresión, localización intracelular aberrante y cambios en su patrón de glicosilación han sido asociados con diversos cánceres<sup>254</sup>. MUC2 forma parte del mucus protector en intestino delgado y grueso, sin embargo, no se expresa en estómago sano y sí lo hace en ambos subtipos histológicos de MI<sup>14</sup>.



**Figura 52.** Bloques de haplotipos de *MUC2*, el bloque que comprende a los SNPs rs10902073 y rs10794281 (marcados en negro) más significativos del estudio que se encuentran en el extremo 5' del gen.

En el estudio de expresión por microarrays, dos conjuntos génicos de glicosilación aberrante de mucinas se encuentran sobre-expresados en la MI-No CG. Es posible que un cambio anómalo y temprano en la glicosilación de MUC2 afecte las propiedades protectoras del mucus gástrico permitiendo la interacción directa de *Helicobacter pylori* con el epitelio gástrico y su internalización en el mismo incrementando la reacción inflamatoria. Dicha presencia de *H. pylori* intracelular e intersticial en el epitelio gástrico que incrementa la reacción inflamatoria ha sido reportada previamente<sup>75</sup>. Además, MUC2 es portadora del antígeno tumoral sialil-Tn producido por glicosilación aberrante en MI, CG y en otros tumores, siendo un indicador de mal pronóstico de los mismos<sup>255</sup>. La expresión de MUC2 en lavados peritoneales es un marcador de recurrencia de CG después de su resección quirúrgica<sup>256</sup>. Todo lo anterior apoya a nivel funcional la asociación genética de *MUC2* con la evolución las LPGs.

Otros trabajos han encontrado que los genes *COX2* -765C<sup>124</sup>, las citoquinas pro-inflamatorias *IL8* -251, *MIF* -173<sup>128</sup> y *XRCC1* y *OGG1*<sup>134</sup>, del mecanismo de reparación del ADN, se asocian con la evolución de las LPGs. La mayoría de estos

polimorfismos no han sido estudiados en otras cohortes de seguimiento de LPGs. Sin embargo, en concordancia con el último estudio se ha obtenido que las variantes rs5275 y rs4648276 de *PTGS2* (*COX2*) se asocian inversamente con la progresión, mientras que rs5277 lo hace con la regresión. El haplotipo GCCGAG formado por los alelos de riesgo rs5275 (C) y rs4648276 (C) asociados inversamente con la progresión se asocia también inversamente con la regresión aunque con un valor nominal, su frecuencia haplotípica no supera el 10% (0.098) y un p-valor nominal (0.046). *COX2* es la enzima principal en la síntesis de prostaglandinas, implicadas en inflamación y mitogénesis, y blanco de acción de los anti-inflamatorios no esteroideos. El ratón transgénico de *COX2* desarrolla MI que conlleva al desarrollo de CG, demostrando la importancia de este gen en la carcinogénesis gástrica<sup>164</sup>.

El gen con mayor cantidad de asociaciones a nivel de SNPs individuales y haplotipos es el supresor tumoral *TFF1*, cuya delección en ratón ocasiona el desarrollo de CG<sup>163</sup>. Es interesante señalar que codifica para otra proteína constituyente del mucus gástrico que interacciona directamente con *H. pylori*, reforzando la idea de que una estructura aberrante del mismo pudiera condicionar la evolución de las LPGs. *TFF1* pertenece a la familia de factores tipo trefoil, péptidos secretores que protegen la mucosa gastrointestinal y estabilizan el mucus gástrico, posiblemente por estabilización de glicoproteínas, e intervienen en la cicatrización epitelial. El único trabajo encontrado de asociación genética de *TFF1* con la carcinogénesis gástrica asocia el SNP C-394T localizado en un elemento de respuesta a estrógeno en el promotor de *TFF1* con un mayor riesgo de CG (TT: OR=8.78, IC95% 2.85-27.05)<sup>257</sup>.

En los presentes resultados varios SNPs de *TFF1* se asociaron directa (rs9976977) e inversamente (rs424694) con la progresión. Las variantes rs225357 y rs424694 se asocian directamente mientras que rs4920094 y rs13047838 lo hacen de manera inversa con la regresión. Las combinaciones haplotípicas CAAC y CAGC que contienen el alelo (A) de riesgo de rs9976977 y salvaje (C) de rs424694 y rs13047838 se asociaron con un mayor riesgo de progresión. La elevada frecuencia haplotípica (0.156) y significación (p=0.019) de CAAC la hacen una combinación de interés, sin embargo CAGC posee una frecuencia <10% y una significación nominal (p=0.039). El haplotipo IAAC, con una frecuencia de 0.225, también se asocia significativamente con una menor probabilidad de regresión, lo que es compatible con el mayor riesgo de progresión de los otros haplotipos que contienen el alelo C de rs424694. Esta combinación, aunque tiene una alta frecuencia (0.225) posee un valor nominal (p=0.025) de significación estadística.

También, los SNPs rs1583005 (OR 1.56, IC95% 1.13-2.16) y rs574445 (OR 1.63, IC95% 1.13-2.36) de *CD14* se asociaron con una mayor progresión de las LPGs. Lo cual fue corroborado en el haplotipo TGTG que se asocia igualmente con una mayor

progresión y contiene al alelo de riesgo T de rs574445 y los alelos salvajes (G) de rs11167532 y rs2569190 asociados con un menor riesgo de progresión. En este caso también esta combinación tiene mayor riesgo y significación estadística (OR 1.90, IC95% 1.26-2.87,  $p=0.0022$ ) que la de cada SNP por separado y posee una frecuencia haplotípica suficientemente alta en los casos (28%).

CD14 es una glicoproteína de membrana plasmática unida a glicosilfosfatidilinositol y expresada preferentemente por monocitos/macrófagos, que media la respuesta inmune al LPS bacteriano a través de MYD88, TIRAP y TRAF6, para activar a NFKB1, e inducir la secreción de citoquinas. Su expresión aumenta a consecuencia de la infección por *H. pylori* y los portadores del alelo -260T de rs2569190 manifiestan una inflamación más severa a consecuencia de la infección por *H. pylori*<sup>257</sup>; este alelo se ha asociado con un mayor riesgo de MI<sup>130</sup>. Es un hecho interesante que SNPs (rs11167532, rs1583005, rs2569190, rs5744455, rs778588) y haplotipos de *CD14* que se encuentran asociados respecto a un mayor riesgo de progresión de las LPGs también lo están respecto al CG tipo cardias e intestinal en el estudio Eurgast, confiriendo especial importancia a *CD14* en la carcinogénesis gástrica. La mayoría de estos SNPs son eQTLs, por lo que pudiera ser que la asociación observada reflejara en parte el control que estas variantes ejercen sobre el nivel de expresión génica. Como ha sido previamente comentado, el SNP rs2569190 regula los niveles de CD14 presentes en monocitos, siendo mayores en los homocigotos del alelo de riesgo<sup>243</sup>.

La variante de *CDH1* rs4783573 fue inversamente asociada con la progresión mientras que rs12597188 y rs16260 se asociaron negativamente con la regresión. El haplotipo AG (rs7186053, rs4783573) compuesto por el alelo de riesgo G de rs4783573 se asocia positivamente con la regresión (OR 2.14, IC95% 1.09-4.19), sin embargo presenta una frecuencia baja (0.074) y su significación está en el rango  $0.01 < p < 0.05$ .

Se observó que el haplotipo GAC (rs10760634, rs4837404, rs10739757) de *PTGES* presenta una asociación inversa con la regresión 0.55 (0.33-0.90,  $p=0.0183$ ), pero también es poco frecuente (0.059) y no muy significativo ( $p=0.018$ ). Este gen codifica para la sintasa de prostaglandina E, que cataliza el último paso de la síntesis de prostaglandina E (PGE2), induciendo desarrollo de cáncer por incremento de la angiogénesis, sobrevivencia celular, e inhibición de la respuesta inmune del huésped. El ratón transgénico de *COX2* y *PTGES* expresa al unísono a *COX2* y PGE2, desarrollando metaplasia y CG<sup>165</sup>. No se había analizado previamente la asociación de este gen con LPGs o CG.

El alelo T del SNP rs4503658 de *CDX2* presenta una asociación nominal (OR 1.54, IC95% 1.01-2.37,  $p=0.047$ ) y positiva con la regresión, sin embargo el haplotipo CAT que contiene a dicho alelo presenta un OR mayor (1.78, IC95% 1.11-2.85,  $p=0.0168$ )

que la variante asociada. *CDX2* es el agente causal más importante o *master gene* de la metaplasia intestinal <sup>52</sup>, ya que induce la expresión de los factores de reprogramación de desdiferenciación celular *SALL4* y *KLF5* <sup>63</sup> y genes de diferenciación intestinal como la sacarasa isomaltasa, cadherina LI, *MUC2*, *FURIN*, entre otros <sup>64</sup>. Además, es un inhibidor de la reparación del ADN no homóloga en cáncer de colón <sup>259</sup>, por lo que pudiera tener un rol similar en estómago.

MAP3K14 es una quinasa de tipo serina/treonina que forma parte de las vías de señalización ERK y JNK que activan a NFκB <sup>170</sup>. El alelo de riesgo C del SNP rs16939926 de este gen tiene bajo tamaño de muestra, el haplotipo CCTC muestra un efecto protector de la regresión (OR 0.52, IC95% 0.28-0.95) pero con baja frecuencia (0.039) y una significación no robusta por lo que no reviste importancia.

El único caso en que el análisis haplotípico aporta nueva información de asociación respecto al análisis de SNPs individuales es en el gen *PTPN11*. Su haplotipo GCAGGC se asocia inversamente con la regresión y, sin embargo, ningún SNP del mismo se asocia con las LPGs. Su frecuencia es mayor del 10% pero su nivel de significación no es robusto (0.034).

Al realizar el análisis de asociación estratificado por los factores de virulencia de *H. pylori* *cagA+/vacAs1/m1* aparecieron nuevos SNPs asociados respecto al análisis no estratificado; así como, un incremento del riesgo (ORs mayores) de otros previamente asociados, aunque de manera general el tamaño de muestra es muy pequeño como para emitir conclusiones sólidas (Tabla 28). Una excepción que incrementa el riesgo de progresión en el grupo de combinaciones no *cagA+/vacAs1m1* es rs6663310 de *RUNX3* que pasa de OR 1.61 IC95% 1.03-2.5 a OR 2.42, IC95% 1.04-5.66. Otra excepción es un SNP que decrementa el riesgo de regresión respecto al análisis general en el grupo *cagA+/vacAs1/m1*, la variante rs2070803 de *MUC1*. SNPs que incrementan el riesgo de regresión en otras combinaciones respecto al análisis general son rs778588 (*CD14*), rs11865086 (*MAPK3*) y rs424694 (*TFF1*). Mientras que variantes que decrementan el riesgo de regresión son rs13047838 y rs4920094 de *TFF1*. Figueiredo *et al.* encontraron que en pacientes portadores de combinaciones de factores de virulencia y alelos de riesgo existe un incremento importante de los ORs <sup>99</sup>.

En conclusión, la variabilidad genética en *MUC2* se asocia inversamente con la regresión de las LPGs y es el resultado más importante ya que su SNP rs10902073 es el único significativo después de aplicar el test de corrección para comparaciones múltiples, y dicha asociación se observó previamente en el estudio de seguimiento de LPGs de Soria, por lo que se replica en este estudio. Otros genes asociados a nivel de SNPs y haplotipos, teniendo en cuenta la significación estadística y frecuencia haplotípica, aunque no superan la corrección por corrección para comparaciones múltiples, son *CD14*, *TFF1* y *PTGS2* con el riesgo y menor

probabilidad de progresión de las LPGs respectivamente. Mientras que *CDH1*, *CDX2* y *TFF1* se asocian positiva y negativa respectivamente con la regresión de las LPGs. Se replican las asociaciones del estudio de Soria en *CD14* (rs2569190, rs5744455), *TFF1* (rs424694), *TFF2* (rs1079380) y *MUC2* (rs10902073, rs10794281). Son necesarios estudios de replicación y funcionales para confirmar dichas asociaciones e identificar posibles biomarcadores de evolución de las LPGs.

### **3. Estudio de expresión génica en metaplasia intestinal que progresa a cáncer gástrico**

#### **3.1 Genes diferencialmente expresados**

Hasta la fecha los estudios de expresión por microarrays en la MI no han diferenciado sus subtipos histológicos completa e incompleta, ni tenido en cuenta el porcentaje de extensión de la lesión metaplásica previo a la extracción del ARN, ni usado muestras bien caracterizadas clínicamente con respecto a su progresión a CG a lo largo de un seguimiento clínico, o realizado un análisis de enriquecimiento funcional en vías de señalización celular u otros procesos moleculares ya que los anteriores estudios de expresión génica se limitaron a exponer los genes diferencialmente expresados. Teniendo estas limitantes en cuenta se ha realizado un estudio de expresión génica por microarray comparando los subtipos histológicos MIC y MII que progresan a CG respecto a los mismos que no progresan a este tumor, así como el conjunto de las metaplasias que no progresan a CG respecto a la mucosa gástrica sana. Las muestras poseen una extensión de la lesión metaplásica  $\geq 75\%$  lo que garantiza una alta representación de transcritos característicos de la misma.

##### **3.1.1 Subtipos histológicos de metaplasia intestinal que progresan a cáncer gástrico**

Mediante el análisis del perfil de expresión génica por microarrays usando el array Almac-Xcel (Affymetrix, EUA) específico para muestras parafinadas, hemos encontrado 106 genes diferencialmente expresados en la MII que progresa a CG respecto a la que no progresa. Aunque las diferencias de expresión no son muy grandes, algunos genes poseen funciones en la carcinogénesis que pudieran influir en la progresión a CG desde la MII. Los candidatos de mayor interés son transcritos significativos en MII-CG vs MII-No CG (N=36) y en la comparación MII-CG vs Sanos (N=22), excluyendo transcritos no específicos compartidos con la comparación MII-No CG vs Sanos (Figura 32, Tabla 32).

Entre los genes diferencialmente expresados de mayor interés y que reflejan el proceso inflamatorio inducido por *H. pylori* están los que codifican para moléculas del procesamiento antigénico como *HLA-A*<sup>207</sup>, *HLA-C*<sup>223</sup>, *HLA-DRB1/3/5*<sup>195</sup>, *HLA-*

*DQA1*<sup>219</sup> y *HLA-DQB1*<sup>211</sup>, y *BPIFB1* que participa en la modulación de la respuesta al lipopolisacárido (GeneCards). El regulador negativo de la expresión de HLA-clase II, *IK*, se encuentra sub-expresado en la MII-CG, pudiendo influir parcialmente en la alta expresión de moléculas HLA clase II en este grupo y en el incremento de la presentación y procesamiento antigénico en la MII que progresa a CG. Además, la quemokina CXCL14, que participa en el reclutamiento de monocitos no activados y células dendríticas y se correlaciona positivamente con los estadios tumorales III/IV del CG<sup>204</sup>. Varias proteínas del sistema del complemento como C1R, C1QBP y C3 que participan en la respuesta inmunitaria ante microorganismos y potencian la respuesta inflamatoria, facilitan la fagocitosis e inducen la lisis celular del agente agresor. Componentes de dicho sistema como C1R, C1S, C1QR1, DAF y C1QL1 se sobre-expresan en el CG tipo intestinal<sup>260</sup>, y una inmunoterapia contra CD55 combinada con cirugía ha demostrado un aumento en la supervivencia al CG<sup>261</sup>.

Varias chaperonas sobre-expresadas (2.7>Fold Change>2) CANX, CCT6A, HSP90AA1, HSP90AB1 y ANP32 son encargadas del adecuado plegamiento tridimensional de proteínas oncogénicas. La actividad de HSP90AA1 y HSP90AB1 se encuentra elevada en células tumorales y su alta expresión ha sido asociada con la agresividad tumoral y mal pronóstico en el CG<sup>216</sup>. Además, se encuentra sobre-expresado un miembro de la familia de oncogenes RAS llamado RAN, el cual ha sido asociado a nivel de polimorfismos genéticos con un mayor riesgo de CG<sup>215</sup>. También, NHP2, una ribonucleoproteína componente de la telomerasa (TERT) que permite el alargamiento de los telómeros, su expresión es elevada en el CG y está correlacionada con un mal pronóstico clínico en el CG<sup>262</sup>.

Adicionalmente, se observaron sobre-expresados los factores de inicio de la traducción y reguladores del supresor tumoral RB1, EIF3D y EIF5B. Otros miembros de esta familia como E2F2 y E2F3 se encuentran sobre-expresados en CG tipo intestinal. Además, la caveolina 1 (CAV1), un candidato a supresor tumoral y regulador negativo de la cascada RAS-MAPK, cuya alta expresión se ha asociado con metástasis y estadificación del CG<sup>210</sup>.

Entre las moléculas sobre-expresadas con plausibilidad en la carcinogénesis pero que no han sido previamente asociadas con la carcinogénesis gástrica se encuentra el gen *MYOF*, implicado en la transducción de señales del VEGF, un inductor del mecanismo de angiogénesis (GeneCards). Y *GNL3*, una proteína que interactúa con TP53 y puede estar implicada en la carcinogénesis debido a su rol en la proliferación de células madres (GeneCards).

Entre las moléculas sub-expresadas se encontró la ciclofilina (*PPIA*) que acelera el plegamiento proteico en general, mientras que varias chaperonas se encuentran sobre-expresadas. La quinasa dependiente de ciclina CDKL3, un regulador del ciclo



celular ausente en la leucemia. El transportador de hierro ceruloplasmina (CP), ha sido reportado que la deficiencia de este metal se asocia con el CG no cardias <sup>263</sup>. Así como RBBP7, proteína que interactúa con el supresor tumoral BRCA1 regulando la diferenciación y proliferación celular (GeneCards).

Por lo tanto, en la MII-CG aparecen sobre-expresadas moléculas indicativas del proceso inflamatorio como HLA clase I y II, proteínas del sistema del complemento, moduladoras de la respuesta al LPS (*BPIFB1*) y la quemokina CXCL14. Y otras moléculas relacionadas con la carcinogénesis como chaperonas (HSP90AA1/AB1), el oncogén *RAN*, un componente de la telomerasa TERT (*NHP2*), reguladores del supresor tumoral RB1 (*EIF3D* y *EIF5B*), el inductor de metástasis gástrica *CAV1*, reguladores de la angiogénesis (*MYOF*) y de la proliferación de células madres (*GNL3*). Funciones sub-expresadas de importancia en la progresión a CG están dadas por el regulador negativo de la expresión de HLA clase II, *IK*, la función antimicrobiana (*LCT*), la ciclofilina, la regulación del ciclo y la proliferación celular (*CDKL3*, *RBBP7*) y del metabolismo del hierro (CP).

En la MIC que progresa a CG (MIC-CG) existe una menor cantidad de transcritos significativos (19) respecto a la MII-CG, la mayoría de los cuales están sub-expresados (11) con respecto a la MIC-No CG (Tabla 33). De manera similar con MII-CG no existen genes significativos después de aplicar la corrección para comparaciones múltiples y las diferencias de expresión son pequeñas ( $3.47 \geq \text{Fold Change} \geq 0.42$ ). Es relevante destacar el mayor número de genes diferencialmente expresados y con mayores valores de expresión en el grupo MII-CG respecto a la MIC-CG, lo que podría relacionarse con el mayor riesgo de progresión a CG de la MII en comparación con la MIC reportado en estudios epidemiológicos <sup>57</sup>.

Entre los transcritos específicos de la progresión a CG desde la MIC (Figura 34) y sub-expresados se encuentra *ANAPC5*, un componente del complejo promotor de la anafase que controla la progresión a través de la mitosis. A diferencia del grupo MII-CG, los componentes del sistema inmune *IL1R2*, y de los anticuerpos (*IGHG1/IGHGM/IGHV4-31*) presentan un bajo nivel de expresión indicando parcialmente una menor intensidad de la respuesta inmune en el grupo MIC-CG respecto al MII-CG. En similitud con lo anterior *IL1R2* también se encuentra sub-expresada en el CG intestinal <sup>214</sup>. Entre los genes sobre-expresados se encuentran *GP2*, *HOXA13*, *IGFBP5*, *OLFM4*, y *HLA-DRB1/3/4/5*. *HOXA13* es un factor transcripcional de la familia de genes homeobox como *CDX1/2*, encargados del desarrollo embrionario intestinal. Su alta expresión ha sido asociada con la agresividad del CG <sup>264</sup>, sugiriendo que quizás otros factores transcripcionales diferentes de *CDX1/2* pueden influir en la diferenciación intestinal metaplásica y la progresión tumoral. *IGFBP5* es una proteína de unión a IGFs (*insulin growth factors*), los cuales estimulan el transporte de glucosa y la proliferación celular. La

olfatomedina 4 (OLFM4) es una glicoproteína de la matriz extracelular que facilita la adhesión celular, está sobre-regulada en la MI<sup>81</sup>, y en el CG tiene función anti-apoptótica promoviendo el crecimiento tumoral<sup>265</sup>. El menor número y función de transcritos relacionados con la carcinogénesis en la MIC-CG en comparación con la MII-CG pudiera reflejar el hecho de que aún no ha ocurrido una desregulación relevante de la expresión en genes inflamatorios, oncogenes, supresores tumorales, etc., responsables de la progresión tumoral. Lo cual se corresponde con el hecho de que la MIC es un estadio de mayor diferenciación tisular previo a la MII<sup>64</sup>.

En conclusión, hemos encontrado que las metaplasias intestinales de ambos subtipos histológicos que progresan y no progresan a CG no se diferencian extensamente atendiendo a sus perfiles transcripcionales. Ya que se agrupan en el dendrograma (Figura 28), muestran pequeñas diferencias ( $5 \geq \text{Fold Change}$  absoluto  $\geq 0.26$ , Figuras 29-30) en el perfil de expresión, y dichos transcritos no son significativos después de ajustar por comparaciones múltiples. Sin embargo, es de destacar que existen diversos genes implicados en la carcinogénesis diferencialmente expresados. Se observa un mayor número de genes con funciones oncogénicas en la MII-CG respecto a la MIC-CG, lo cual apoya a nivel molecular el mayor riesgo de progresión a CG de la MII en comparación con la MIC<sup>60</sup>. Sin embargo, todas las MI son extremadamente diferentes de la mucosa gástrica sana en el dendrograma, heatmap y *Fold Changes* de genes diferencialmente expresados.

### **3.1.2 Metaplasia intestinal que no progresa a cáncer gástrico en comparación con mucosa gástrica sana**

Con el objetivo de caracterizar el perfil transcripcional de la metaplasia intestinal respecto a la mucosa gástrica sana, se agruparon ambos subtipos histológicos que no progresan a CG y se efectuó la comparación MI-No CG vs Sanos, la cual permite corroborar transcritos diferencialmente expresados previamente reportados en la MI. Se eligió para dicha comparación a la MI que no progresa a CG ya que es la más frecuente en la población y, además, para determinar su perfil transcripcional eliminando procesos oncogénicos de su caracterización. También, se realizó la comparación MI vs Sanos observándose que sus resultados son similares a MI-No CG vs Sanos en un 74.65% de los genes desregulados con una alta correlación entre dichos genes ( $r=0.92$ ,  $p<0.0001$ ).

A diferencia de las anteriores comparaciones, en este caso los transcritos son significativos atendiendo al p-valor ajustado por comparaciones múltiples ( $p<0.05$ ) y exhiben mayores diferencias de expresión ( $98.24 \geq \text{Fold Change} \geq 0.11$ ), lo que evidencia una diferencia radical entre el perfil transcripcional de la MI y la mucosa gástrica sana. Entre el total (579) se encuentran diversos genes previamente identificados desregulados en estudios de microarrays de MI (apartado 8.3 de la

sección Revisión Bibliográfica) como *CDX2*, *ACE2*, *KRT20*, *MUC13*, *OLFM4*, *REG4*<sup>81</sup>, *FABP1*, *MEP1B*, *SI*, *SLC6A19*<sup>214</sup>, *CDX1*, *MTTP*, *CEACAM6*<sup>79</sup>, *DGKQ*<sup>78</sup>, *APOA1*, *APOA4*, *APOB*, *ALDOB*, *CDH17*, *CLDN3*, *CLDN4*, *GGT1*, *HNF4A*, *VIL1*<sup>80</sup>, *MUC2*<sup>266</sup>, *MUC4*<sup>267</sup>, *TFF3*<sup>268</sup> y *CPS1*<sup>269</sup> (Tabla anexa 2). La amplia similitud encontrada con genes previamente hallados desregulados en la MI podemos considerarla como una validación de la veracidad de estos resultados.

Además, se han encontrado genes sobre-expresados respecto a los cuales se ha encontrado anteriormente una molécula de la misma familia génica (nueva molécula similar) asociada con la MI, como familias de transportadores de membrana SLC (*solute carrier*, N=36) especializados en el transporte de bicarbonato, glucosa, Na/Cl<sup>-</sup>, aminoácidos y oligopéptidos, lo cual es indicativo de un mayor metabolismo y quizás proliferación celular en la MI en comparación con la mucosa gástrica sana (Tabla anexa 2).

Entre los transcritos sobre-expresados con mayores diferencias (*Fold Change*>18) se encuentran en genes de expresión intestinal como las mucinas *MUC2*, *MUC3A/3B*, *MUC12* y *MUC17*, las cuales son secretadas por las células caliciformes, que incrementan drásticamente su número en la MI siendo el rasgo más distintivo de la misma<sup>50</sup>. Este incremento de los principales componentes del mucus gástrico es un mecanismo de defensa y a la vez consecuencia de la inflamación ocasionada por *H. pylori*<sup>64</sup>. *MUC2* no se expresa en el epitelio normal mientras que si lo hace en ambos subtipos histológicos de MI. En el estudio de asociación con la evolución de LPGs el resultado más significativo fue obtenido con variantes de este gen. Sin embargo, *MUC2* u otras mucinas no están diferencialmente expresadas en las metaplasias que progresan a CG, por lo que su desregulación es un evento temprano y específico de la progresión a MI pero no a cáncer gástrico.

*CDX1* y *CDX2* presentan *Fold Changes* no tan altos, de 4.5 y 3.28 respectivamente, por lo que es posible que su nivel de expresión no necesite ser tan elevado para seguir propiciando la diferenciación intestinal en el estómago una vez establecida la misma. De manera muy interesante encontramos la sobre-expresión de otro miembro de la familia génica homeobox, *HOXA13* (*Fold Change* 3.62, p-valor ajustado 9,0725E-07), el cual no ha sido reportado previamente desregulado en la MI, participa en el desarrollo embrionario del eje antero-posterior. Así como, la sub-expresión de *HOTAIRM1* (*Fold Change* 0.45, p-valor ajustado 0.014), un ARN antisentido que regula negativamente a esta familia génica y que ha sido asociado con mala sobrevivencia en CG<sup>270</sup>. Por lo que es posible que *HOXA13* contribuya sinérgicamente con *CDX1/2* a la diferenciación intestinal. La sobre-expresión de este gen también fue observada en la MIC-CG, resaltando su posible funcionalidad.

Es interesante destacar la relevancia del metabolismo lipídico entre los transcritos con mayor sobre-expresión, lo cual también está en correspondencia con resultados previos. Este proceso está representado por la proteína de transferencia de triglicéridos y colesterol MTP <sup>79 80</sup>, además de las apolipoproteínas APOA1/A4, APOB, APOBEC1, APOC3 y FABP1-2 <sup>80 81</sup>.

CPS1 (*Fold Change* 26.25), un biomarcador inmunohistoquímico específico y con alta expresión en CG <sup>273</sup>, también tiene alta expresión en la MI-No CG. Esta enzima elimina el exceso de amonio celular, mientras que *H. pylori* produce este compuesto como un mecanismo de protección ante el ambiente ácido estomacal <sup>27</sup>, reflejando la infección bacteriana en el grupo MI-No CG. El gen con mayor sobre-expresión (*Fold Change* 98.24) es el supresor tumoral en varios cánceres incluido el CG, *DMBT1*, con roles en la protección mucosal, unión a *H. pylori*, y diferenciación epitelial. Ha sido validado como un marcador inmunohistoquímico de MI <sup>275</sup> y se encuentra sobre-expresado en esófago de Barret y CG <sup>271</sup>. Es probable que su alta expresión sea interpretada como un mecanismo de defensa temprano y previo ante una posible desregulación oncogénica, la cual ocurre de hecho en las MI que progresan a CG.

Otros grupos de transcritos con alta sobre-expresión (*Fold Change* 2.12-4.73) son seis miembros de la familia de las sulfotransferasas (SULT1A1-4, SULT1E1, SULT1B1). Enzimas que participan en la conjugación de grupos sulfato a drogas y compuestos xenobióticos y carcinógenos. Además, existen varios genes de la familia de las UDP-glucuronosiltransferasas (UGTs) sobre-expresados (*Fold Change* 2.01-3.63) los cuales participan en la modificación química y eliminación de compuestos xenobióticos potencialmente tóxicos. Previamente se ha encontrado un aumento de la actividad sulfotransferasa en el cáncer gástrico <sup>272</sup>. También, se encuentran sobre-expresadas varias proteínas que contienen motivos tripartitas (TRIM14/15/31/36/40, *Fold Change* 2.01-3.05), las cuales activan a NFκB incrementando la respuesta inflamatoria <sup>273</sup>.

Respecto a los genes sub-expresados (*Fold Change*<0.19) en la MI-No CG respecto a la mucosa sana se observaron diversos transcritos con funciones gástricas como la gastrina (GAST), somatostatina (SST), ATPasa H<sup>+</sup>-K<sup>+</sup> (ATP4A/B), pepsinógeno C (PGC), gastroquinas 1 y 2 (GKN1-2), clusterina (CLU), ligasa de glutamato y amonio (GLUL), el factor intrínseco gástrico (GIF) y PSCA. Los cuales indican una hipoclorhidria en la mucosa gástrica que permite el crecimiento de bacterias generadoras de radicales libres y por ende daño al ADN, y una disminución de la digestión proteica por la pepsina. Una sub-expresión similar de transcritos con funciones gástricas ha sido observada en el CG respecto al pepsinógeno C, la anhidrasa carbónica II y la somatostatina <sup>80</sup>.

Diversos miembros (N=17) de los ARNs nucleolares pequeños C/D box (*SNORDs116*) se encuentran sub-expresados. Los mismos son reguladores de la metilación de ARNs ribosomales influyendo esta modificación química en la estructura tridimensional de los mismos y su interacción con proteínas ribosomales. Microdeleciones en el *cluster* de genes *SNORDs116* contribuyen al síndrome de Prader-Willi y este cluster controla la expresión génica de alrededor de 200 genes<sup>274</sup>. Otras proteínas gástricas reguladas negativamente son las mucinas MUC1, MUC5AC, MUC5B y MUC6, cuya disminución de la expresión ha sido reportada previamente en la MIC<sup>14</sup>.

En conclusión, hemos encontrado diferencias significativas de expresión entre la MI-No CG y la mucosa gástrica sana. Los genes sobre-expresados son responsables de la diferenciación intestinal (*CDX1/2*, *HOXA13*), así como transcritos expresados (mucinas, factores trefoil) por células intestinales que forman el mucus gástrico y genes del metabolismo lipídico, en concordancia con resultados previos. Como nuevos hallazgos se observa la sobre-expresión de múltiples transportadores de membrana SLC indicativos de un activo metabolismo y proliferación celular, un miembro de la familia génica homeobox (*HOXA13*) que pudiera ejercer un rol en la diferenciación intestinal además de los clásicos *CDX1/CDX2*, diversas proteínas (SULTs, UGTs) implicadas en el metabolismo de xenobióticos, proteínas TRIM activadoras de NFκB, proteínas TMEM de función desconocida y el supresor tumoral *DMBT1*. Los transcritos sub-expresados implican a proteínas que generan y regulan la secreción ácida gástrica, por lo que en consecuencia existe una hipoclorhidria y disminución de la digestión proteica. Nuevos genes encontrados con baja expresión incluyen a ARNs nucleolares pequeños tipo C/D box de la familia *SNORDs116*.

### **3.2 Vías de señalización celular y procesos biológicos implicados en la progresión de los subtipos histológicos de metaplasia intestinal a cáncer gástrico**

El análisis individual de genes diferencialmente expresados entre dos grupos presenta diversas limitantes. El mismo no tiene en cuenta los importantes efectos de vías de señalización celular y procesos moleculares<sup>183</sup>. Además, existe mayor concordancia entre estudios al evaluar vías en lugar de genes aislados y la interpretación de una larga lista de genes significativos es una tarea difícil dependiente del conocimiento del investigador<sup>183</sup>. Teniendo en cuenta lo anterior se han desarrollado métodos bioinformáticos que ayudan en la identificación de procesos biológicos a partir de genes diferencialmente expresados haciendo uso del análisis de enriquecimiento funcional. Este tipo de análisis se basa en que si un proceso biológico esta alterado los genes co-funcionales deben tener una alta probabilidad de estar significativamente enriquecidos en el fenotipo de estudio.

Existen diversos métodos para este análisis pero no brindan soluciones estadísticas exactas, sino que son algoritmos exploratorios que permiten dar una idea de los procesos implicados, los cuales deben ser evaluados sobre la base de su plausibilidad biológica. La diversidad de variables que influyen en estos métodos hacen que los tests de corrección para comparaciones múltiples sean solo una solución parcial a su baja especificidad <sup>275</sup>. Hemos utilizado dos programas bioinformáticos (*Gene Set Enrichment Analysis* [GSEA] e *Ingenuity Pathway Analysis* [IPA]) que hacen uso del enriquecimiento funcional con el objetivo de identificar vías de señalización celular y procesos biológicos desregulados en la progresión desde los subtipos de MI a CG (apartado 3.4.1 de la sección Resultados) y para comparar la MI que no progresa a CG con la mucosa gástrica sana.

### 3.2.1 Gene Set Enrichment Analysis

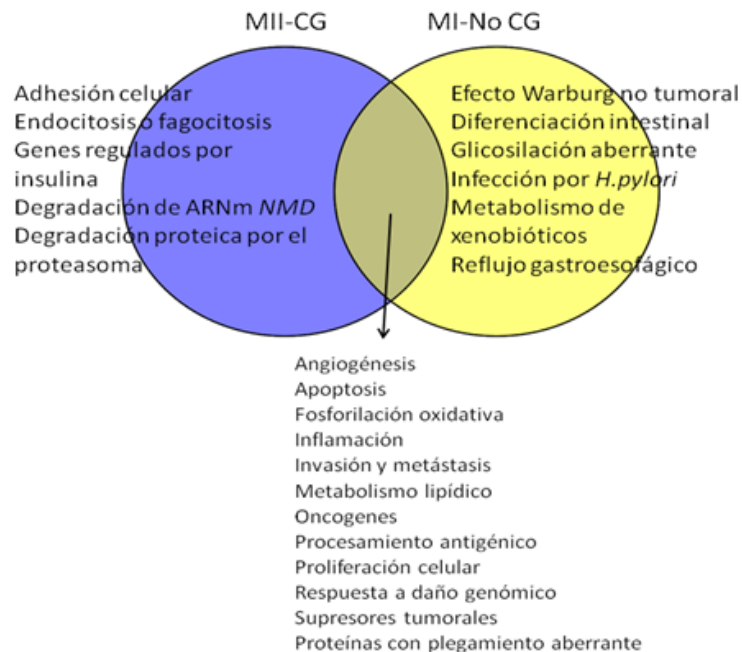
La principal ventaja del análisis del programa GSEA es que tiene en cuenta todos los genes analizados independientemente de su significación estadística o nivel de expresión. Permitiendo que genes con un efecto pequeño, que no pasarían los criterios de selección habituales ( $p < 0.05$  y  $Fold\ Change > 1.5$  o  $2$ ), contribuyan al enriquecimiento de un conjunto génico. Su principal limitante es que asume que los genes con mayores diferencias de expresión son los más importantes <sup>275</sup>.

#### MII y MIC que progresan a cáncer gástrico

Se observa una drástica diferencia entre la MII respecto a la MIC que progresa a CG; ya que en la primera se obtienen 124 conjuntos génicos sobre-expresados (Tabla anexa 3) y relacionados con diferentes procesos carcinogénicos, mientras que solo 2 sin plausibilidad biológica en el caso de la MIC. Esto concuerda con el hecho de un mayor número de genes desregulados implicados en la carcinogénesis y específicos de la MII-CG (N=58) en comparación con la MIC-CG (N=10). El mecanismo por el cual las MIC progresan a CG es desconocido, sin embargo es probable que en este caso sea relevante el papel de unos pocos genes diferencialmente expresados en lugar de vías de señalización celular, que sea necesario un mayor tiempo para que en la MIC comience un proceso de desregulación de la expresión génica como en el grupo MII-CG o que al pasar del tiempo la MIC se transforme en MII y exhiba una desregulación en la expresión de genes responsables de la progresión tumoral como la observada en esta última. Pudiera ser de importancia en este aspecto la molécula de adhesión celular que promueve la proliferación celular en el CG, OLFM4 <sup>265</sup>, la cual se encuentra sobre-expresada en la MIC-CG.

Con el objetivo de determinar los procesos comunes y específicos en el grupo MII-CG respecto a MI-No CG se realizó un diagrama de Venn (Figura 53). Los procesos comunes podrían desempeñar un papel en los pasos iniciales de la MI y continuar durante la progresión al CG. Sin embargo los específicos del grupo MII-CG

influyen sólo durante la progresión tumoral. Los procesos sobre-expresados específicamente en la MII-CG y no en la MI-No CG son la adhesión celular, endocitosis o fagocitosis, genes regulados por insulina, degradación de ARNm NMD y la degradación proteica por el proteasoma.



**Figura 53.** Diagrama de Venn que compara procesos funcionales sobre-expresados comunes y específicos entre los grupos MII-CG y MI-No CG, según los análisis GSEA e IPA.

El mayor número (24/144) de conjuntos sobre-expresados relacionados con la carcinogénesis en el grupo MII-CG están representados por el ciclo celular y moléculas reguladoras del mismo como TGFB1, PI3K, MAP quinasas y factores de transcripción E2F, lo cual indica una elevada proliferación celular en la MII-CG en comparación con la MII-No CG. El proyecto internacional Cancer Genome Atlas ha realizado la más extensa caracterización genómica del CG determinando mecanismos moleculares comunes en los mismos. Este estudio ha demostrado que, a semejanza de estos resultados, la vía de señalización PI3K se encuentra reiteradamente sobre-expresada en el CG <sup>46</sup>. Además, un trabajo con el marcador inmuno-histoquímico PCNA ha demostrado un incremento de la proliferación epitelial en la MI en comparación con la gastritis atrófica, incluso después de la erradicación de *H. pylori* <sup>276</sup>. Los factores transcripcionales E2F, EIF3D y EIF5B, también se encuentran sobre-expresados en la MII-CG y los resultados del catálogo c3.tft.v5.0 muestran que existen 19 de 27 conjuntos génicos compuestos por la familia E2F significativamente (q-valor FDR<0.05 y p-valor<0.01) sobre-expresados en la MII-CG.

A continuación, existen 20 conjuntos génicos compuestos por supresores tumorales (*RB1*, *BRCA1/2* y *TP53*) o dianas de los mismos. Estos genes inhiben la proliferación celular excesiva deteniendo el ciclo celular en respuesta generalmente a un daño en el ADN. La familia de RB1 o proteína del retinoblastoma, está compuesta por diversos miembros que presentan un dominio de unión a factores transcripcionales E2F (participan en el inicio de biosíntesis proteica) inhibiendo la acción de los mismos por bloqueo de las fases G1/S<sup>277</sup>. La expresión génica de RB1 aumenta acorde a la severidad de la cascada de Correa<sup>277</sup> y la vía de señalización del mismo se encuentra sobre-expresada en el CG<sup>46</sup>.

Además, se encontraron 20 conjuntos génicos desregulados compuestos por los oncogenes *MYC*, *RAS*, *NRAS*, *KRAS*, *MET* y *PTEN*, siendo *MYC* el más frecuente. De manera general los oncogenes son receptores de factores de crecimiento con actividad quinasa y factores transcripcionales. El aumento de su expresión implica un incremento en la proliferación y sobrevivencia celular debido a la regulación positiva que ejercen sobre el ciclo celular. En este sentido la infección por *H. pylori* induce la expresión de oncogenes *MYC* y  $\beta$ -catenina<sup>278</sup>. Se ha encontrado que *MYC* se expresa en el 77% de los adenocarcinomas gástricos asociándose con metástasis en nodo linfático y estadios tumorales III-IV<sup>279</sup>.

También se observaron una serie de conjuntos génicos (N=15) con un rol en la invasión y metástasis (*CDH1*, *ZEB1*, *ERBB2*, *MMP14*) y transición epitelio-mesénquima, proceso que tiene un papel relevante en la transformación tumoral gástrica<sup>44</sup>. Mutaciones en *CDH1* (cadherina E) han sido encontradas en el CG hereditario de tipo difuso y en hasta el 37% de los CGs con estabilidad genómica<sup>46</sup>. La pérdida de la función de *CDH1* contribuye a la progresión de la tumorigénesis por la ruptura de su unión con la  $\beta$ -catenina y pérdida de las uniones célula-célula con un consecuente incremento en la invasión y metástasis<sup>44</sup>. El gen *ERBB3* se encuentra mutado en el 20% de los CGs estudiados en el Cancer Genome Atlas<sup>46</sup>.

Otro proceso con diversos conjuntos desregulados (N=13) es la inflamación, como se evidencia en varios conjuntos compuestos por el factor transcripcional NF $\kappa$ B, el cual es activado por la infección de *H. pylori*, induciendo una señalización que culmina en la expresión de citoquinas pro-inflamatorias. Otro componente de los mismos es TNF, una citoquina pro-inflamatoria incrementada en la MI y otras LPGs.

Es interesante destacar que 8 conjuntos implicados en la adipogénesis aparecen a continuación y genes del metabolismo lipídico como *FABP5* y *FABP5P3* se encuentran sobre-expresados en el grupo MII-CG. La adipogénesis es el proceso de diferenciación por el cual los preadipocitos se convierten en adipocitos, los cuales participan en la homeostasis energética. La alta proliferación celular trae consigo una alta síntesis lipídica o lipogénesis para garantizar la generación de membranas



biológicas. Este fenómeno en el cáncer además conlleva a un incremento de lípidos saturados en las membranas protegiéndolas del daño oxidativo por peroxidación lipídica. También, los lípidos actúan como moléculas de señalización (Ej PIP3) en la proliferación celular, el colesterol es requerido para el funcionamiento de la vía de señalización hedgehog la cual es necesaria para el mantenimiento de la diferenciación y secreción ácida gástrica, y son componentes integrales del proceso de la autofagia, el cual permite a los tumores mantener su suplemento energético en condiciones de carencia de nutrientes. Además, participan en la regulación de la migración de células tumorales, interacción tumor-estroma, inducción de angiogénesis e hipoxia<sup>280</sup>.

La apoptosis esta representada en la MII-CG por solo 6 conjuntos en comparación con un mayor número (24) implicados en la proliferación celular, lo que indica un predominio de la proliferación sobre la apoptosis, característico de los procesos neoplásicos.

Aparecen 4 conjuntos génicos compuestos por integrinas, glicoproteínas de adhesión con la matriz extracelular y de la unión célula-célula. Las interacciones celulares mediadas por integrinas se han obtenido sobre-representadas en el CG<sup>46</sup>. Se observaron 5 conjuntos sobre-expresados que implican a la angiogénesis, constituidos la mayoría por el regulador positivo de este proceso VEGFA (vascular endothelial growth factor) el cual se sobre-expresa en el CG<sup>281 46</sup>, e incluso una terapia para disminuir la angiogénesis bloqueando su receptor (VEGFR2) ha sido aprobada para el tratamiento de CG en estados avanzados<sup>282</sup>.

Otro proceso con relevancia en la carcinogénesis es la degradación proteica vía proteasoma, un mecanismo que garantiza la degradación de entre el 80-90% de las proteínas intracelulares con una estructura tridimensional aberrante o de muy corto tiempo de vida. Los cánceres incrementan la degradación proteasomal de supresores tumorales para evadir la apoptosis y tener una proliferación elevada a pesar del daño genómico presente en sus células. *H. pylori* induce la degradación proteasomal de supresores tumorales gástricos como RUNX3<sup>283</sup> y CDKN1B (p27kip1)<sup>284</sup> y del inhibidor apoptótico BIRC5<sup>285</sup>.

Una estructura citoplasmática llamada PaCS (*particle-rich cytoplasmic structure*) consistente de acúmulos de proteínas ubiquitinadas y componentes proteasomales 19S y 20S ha sido detectada en la MII, sin embargo no se ha detectado en la MIC. Esta estructura modula la virulencia bacteriana induciendo la respuesta inmune inflamatoria a través del inmunoproteasoma y participa en la regulación del crecimiento neoplásico<sup>75 286</sup>.

Otros procesos sobre-expresados en la MII-CG pero con menor número de conjuntos génicos son la regulación por insulina, cáncer de esófago, cáncer gástrico,

degradación de ARNm aberrantes NMD y respuesta a daño genómico y a proteínas con un plegamiento aberrante. La señalización por insulina ha sido obtenida como un proceso desregulado en el CG. El cáncer de esófago progresa a partir del esófago de Barret, una metaplasia intestinal esofágica que presenta grandes similitudes con la gástrica. La presencia del mecanismo de degradación de ARNm aberrantes en la MII-CG es otro mecanismo similar a la degradación proteasomal pero a nivel de ARNm que confirma una alta tasa de degradación de transcritos o proteínas en la MII-CG y que previamente fue obtenida en el estudio de asociación genética en *NOD2*. La presencia de proteínas con plegamiento aberrante se confirma también a nivel de genes diferencialmente expresados con la sobre-expresión de varias chaperonas en la MII-CG.

De esta manera, se han hallado una diversidad de procesos moleculares sobre-expresados que pudieran condicionar y caracterizar la progresión desde la MII hacia el CG. Sin embargo, en el grupo MIC-CG no se identificaron conjuntos génicos significativos con plausibilidad en la carcinogénesis gástrica. Se ha planteado previamente la existencia de un mayor riesgo de progresar a CG para la MII respecto a la MIC. Este hecho se ha observado en dos estudios epidemiológicos de seguimiento de lesiones precursoras gástricas realizados, uno en la provincia de Soria <sup>57</sup> y otro en toda España, parte de las muestras analizadas en el microarray provienen de dichos estudios. Otros estudios epidemiológicos de seguimiento de LPGs y transversales han obtenido un resultado similar, apoyando la idea de que el subtipaje de la metaplasia intestinal pudiera contribuir a la detección de pacientes con MII que muestran una mayor probabilidad de desarrollar CG <sup>60</sup>.

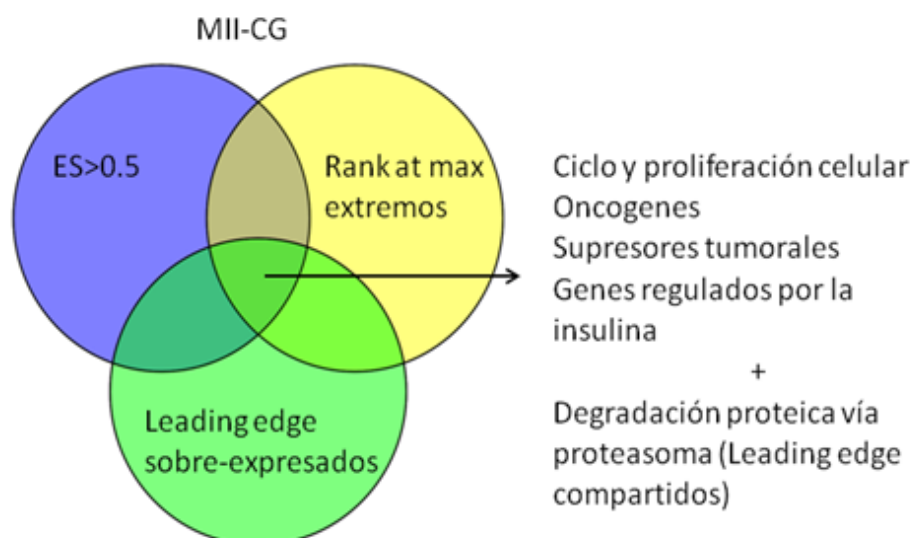
Otros estudios moleculares han mostrado diferencias entre la MII y la MIC que apoyan el hecho epidemiológico de un mayor riesgo de progresión para la MII respecto a la MIC. Por ejemplo, la presencia de *H. pylori* intraepitelial o intercelular se ha observado con mayor frecuencia en la MII en comparación con la MIC, induciendo una mayor respuesta inflamatoria debido a la existencia de una estructura formada por *cagA*, *vacA*, proteínas poliubiquitinadas, componentes del proteasoma y proteínas oncogénicas SHP2 y ERK <sup>75 286</sup>. Otro proceso es una menor expresión de CDX2 en la MII en comparación con la MIC, induciendo un estadio indiferenciado susceptible de transformación tumoral. En el mismo la acción supresora tumoral de CDX2 sería menor en la MII permitiendo la proliferación de células metaplásicas susceptibles de convertirse en tumorales <sup>65 287</sup>. SOX2, una proteína esencial para mantener la pluripotencia de las células madres embrionarias, muestra una alta expresión en la mayoría de las MII (85%) mientras que solo en el 7% de las MIC <sup>288</sup>. También, la vía de señalización Shh tiene un menor nivel de activación en la MII <sup>70</sup>. Esta vía regula la proliferación y diferenciación gástrica, y disminuye su activación a partir de la atrofia gástrica. La inhibición de esta vía conlleva a un programa de diferenciación alterado y pérdida

de la secreción ácida gástrica; así como, a un incremento de la proliferación epitelial y menor expresión de reguladores de la expresión tejido específica <sup>70</sup>. También, se ha reportado una mayor actividad telomerasa <sup>74 289</sup>, incremento de mutaciones en *TP53* e inestabilidad de microsatélites <sup>52</sup> en la MII en comparación con la MIC.

### Genes *leading edge* y filtrado de conjuntos génicos en el grupo MII-CG

En el análisis del GSEA no todos los genes contribuyen de igual manera a un proceso identificado, sino que existe un subconjunto de los mismos (*leading edge*) que contribuyen más a la señal de enriquecimiento (*enrichment score*). Debido a lo anterior, a la baja especificidad del análisis de enriquecimiento funcional y la alta cantidad de conjuntos génicos sobre-expresados (N=124) en la MII-CG, y en un intento por restringir los mismos a los más relevantes se identificaron los genes *leading edge* de los mismos compartidos con los genes sobre-expresados en este grupo <sup>260</sup>.

Además, se filtraron los conjuntos génicos atendiendo a los mayores valores de ES (>0.5) y *Rank at max extremos*. Los procesos moleculares compartidos de los análisis anteriores y, por tanto, de mayor relevancia en la progresión a CG desde la MII son el incremento de moléculas del ciclo y la proliferación celular, oncogenes, supresores tumorales y genes regulados por la insulina (Figura 54). A estos procesos relevantes se les añade el rol de la degradación proteica proteasomal identificado a partir de genes *leading edge* compartidos entre todos los conjuntos génicos. Los conjuntos génicos representativos de dichos procesos funcionales con ES >0.5 (Tabla anexa 3), *Rank at max extremos* (Tabla 35), o genes *leading edge* sobre-expresados en la MII-CG (Tabla 37) pudieran ser validados en futuros estudios de expresión génica o inmunohistoquímicos.



**Figura 54.** Procesos funcionales comunes a conjuntos génicos con diferentes criterios de selección en la MII-CG.

### MI-No CG en comparación con mucosa gástrica sana

En el grupo MI-No CG se observó en comparación con MII-CG una menor cantidad de conjuntos génicos que representan procesos carcinogénicos como la proliferación celular, oncogenes y supresores tumorales, lo cual se corresponde con el estadio de la MI que no progresa a CG (Tablas anexas 3-4). No se observaron conjuntos génicos sub-expresados y con plausibilidad biológica en el fenómeno de estudio. Los procesos específicos de la MI-No CG son el efecto Warburg no tumoral, diferenciación intestinal, glicosilación proteica aberrante, infección por *H. pylori*, metabolismo de xenobióticos y reflujo gastroesofágico. Sin embargo, también existen procesos comunes a ambos grupos como el metabolismo lipídico que pudieran ejercer un rol en ambos estadios de la MI.

El proceso funcional más representado es el efecto Warburg o glicólisis aeróbica (Tabla anexa 4). El mismo consiste en que en los procesos tumorales la obtención de energía ocurre por un incremento de la glicólisis respecto a la fosforilación oxidativa. Para contrarrestar el uso preferente de un proceso menos eficiente energéticamente, las células aumentan la generación de intermediarios como la glucosa y sus transportadores, la gluconeogénesis y la biosíntesis de enzimas glicolíticas como la hexoquinasa que cataliza el paso limitante de la glicólisis. En el grupo MI-No CG se observa la sobre-expresión de transportadores de glucosa *SLC2A5*, *SLC5A9*, la enzima glicolítica hexoquinasa *KHK* y del glucagón, un activador de la gluconeogénesis.

En el efecto Warburg se incrementa la biosíntesis de glucosa a partir de precursores no glucídicos (gluconeogénesis) para facilitar la formación de macromoléculas en condiciones de activa proliferación celular <sup>45</sup>. En este sentido observamos varios conjuntos sobre-expresados del metabolismo de los aminoácidos alanina, aspartato, glutamato, arginina y prolina. Los productos finales de la glicólisis, el lactato y piruvato activan los factores inducibles por hipoxia HIF1A y HIF2A, los cuales a su vez inducen genes glicolíticos, transportadores de glucosa e inductores de angiogénesis. En este aspecto se observaron 2 conjuntos génicos significativos formados por HIF1A.

El efecto Warburg es ocasionado por una fosforilación oxidativa disfuncional debido a mutaciones en el ADNm causadas por radicales libres generados en la misma y que están presentes en la infección por *H. pylori* <sup>290</sup> y en el CG <sup>291</sup>, afectando la integridad mitocondrial y disminuyendo la fosforilación oxidativa. Sin embargo, el efecto Warburg es universal en células normales en condiciones de activa proliferación, con el objetivo de cambiar el metabolismo de fosforilación oxidativa a glicólisis para evitar la generación de radicales libres y proteger al genoma de daño mutacional durante la activa síntesis de ADN <sup>292</sup>. Atribuimos a este “mecanismo de defensa” el predominio del efecto Warburg en la MI que no avanza a CG en lugar de

la que si avanza, hecho que sería lo más plausible biológicamente. La presencia de un solo conjunto significativo relacionado con la fosforilación oxidativa en comparación con 10 relacionados con el metabolismo glicolítico muestra que es plausible un efecto Warburg no tumoral reflejo de una activa proliferación celular en la MI-No CG y, coincidentemente, existen 6 conjuntos génicos sobre-expresados de proliferación celular.

El segundo proceso más representado en la MI-No CG es el metabolismo lipídico a través del metabolismo de ácidos grasos, glicerolípidos, triacilglicerol, transporte lipídico, etc. Debido a que los ácidos grasos son usados como fuente de energía gluconeogénica, el incremento del metabolismo lipídico en este grupo pudiera ser una consecuencia del efecto Warburg. Lo cual esta en consonancia con varios estudios de expresión por microarray que evidencian la sobre-expresión de genes de dicho proceso como *FABP1*, *MTTP*, *APOA1*<sup>80</sup>, *AKR1B10*, *ALDH3A2*, *ADH1B*, *CDS1* y *DGKQ*<sup>78</sup> en la MI, y con el hecho de la sobre-expresión de diversos genes (*APOA1/4*, *APOB*, *APOC3*, etc.) de este proceso en la MI-No CG. La plausibilidad de este mecanismo en la carcinogénesis gástrica está dada porque en condiciones de activa proliferación, las membranas biológicas que poseen un importante componente lipídico necesitan ser sintetizadas para las nuevas células<sup>280</sup>. En este sentido también se observaron 7 conjuntos formados por factores de transcripción hepáticos HNF1A, HNF3A y HNF4A (Tabla anexa 4), los cuales inducen la expresión de genes del metabolismo lipídico<sup>80</sup> y que *HNF4A* y *HNF4G* están sobre-expresados a nivel de genes individuales en la MI-No CG. Es válido recordar que conjuntos sobre-expresados en el proceso de adipogénesis se hallaron en la MII-CG.

A continuación aparece el proceso común (entre MII-CG y MI-No CG) de la inflamación, con diversos conjuntos génicos (N=13) formados por vías de señalización de citoquinas IL2, IL4, IL8, el factor regulador de IFN IRF4, quemokinas, neutrófilos y el factor transcripcional NFkB, cuya activación por la infección de *H. pylori* conlleva a la expresión de citoquinas pro-inflamatorias en la MI y otras LPGs. Varios conjuntos génicos cuyo conocimiento por estudios previos validan a la diferenciación intestinal por CDX2, factores nucleares hepáticos (HNF1A/3A/4A), al esófago de Barret (lesión precursora metaplásica de cáncer esofágico) y el reflujo gastro-esofágico, un factor de riesgo de MI<sup>293</sup>. Posteriormente aparecen conjuntos génicos compartidos con la MII-CG como el procesamiento antigénico, oncogenes, apoptosis, proliferación celular, respuesta a daño genómico, supresores tumorales, angiogénesis e invasión y metástasis.

Otro proceso específico del grupo MI-No CG es el metabolismo de xenobióticos. Estos son compuestos inexistentes en la naturaleza incluyendo a fármacos, aditivos alimentarios y contaminantes ambientales, los cuales son metabolizados en el hígado para hacerlos menos tóxicos y fácilmente excretables. Paradójicamente, el

metabolismo de xenobióticos induce daño al ADN, lípidos y proteínas mediante la generación de radicales libres del O<sub>2</sub> y el N<sub>2</sub>; así como, la transformación de los xenobióticos en carcinógenos<sup>294</sup>. Parte de la regulación de la transformación de xenobióticos es ejercida por la señalización de HNF4A<sup>295</sup>, la cual se encuentra sobre-expresada en la MI-No CG. Este mecanismo se divide en biotransformación de fase I y II. La fase I consistente en reacciones de mono-oxigenación realizadas por citocromos P450 (CYPs) implicados en el metabolismo de drogas y esteroides; mientras que la fase II consiste en reacciones de glucuronidación, sulfación, metilación, acetilación, conjugación de glutatión y de aminoácidos llevados a cabo por enzimas transferasas como UGTs (UDP glucuronosiltransferasas), sulfotransferasas (STs), N-acetiltransferasas (NATs), glutatión S-transferasas (GSTs) y metiltransferasas. Además de observarse este proceso molecular enriquecido en 3 conjuntos génicos, a nivel de transcritos individuales se observó previamente que diversos miembros de las enzimas UGTs, sulfotransferasas y citocromos P450 se encontraban sobre-expresadas en la MI-No CG. Se ha observado asociación entre polimorfismos de genotipos *GSTT1* y *GSTM1* con las lesiones preneoplásicas<sup>136</sup>, y *COMT* Val158Met con la severidad de la MI<sup>296</sup>.

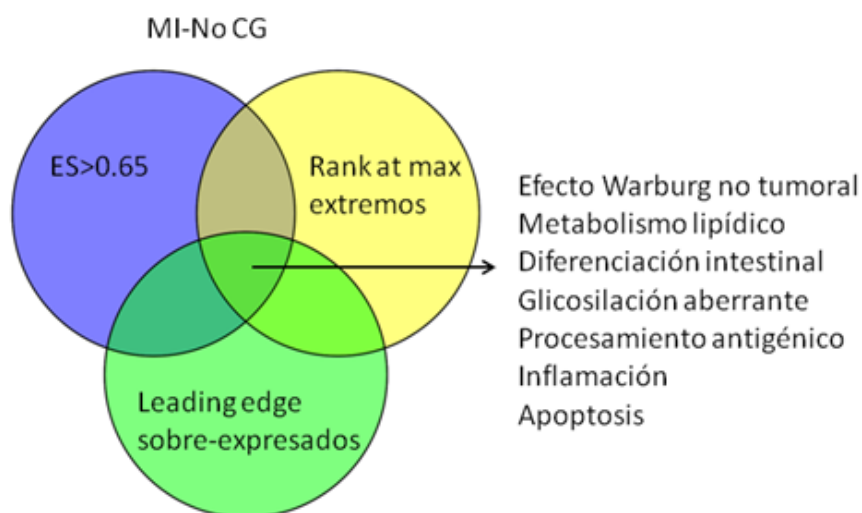
Un par de procesos específicos de la MI-No CG son la terminación de biosíntesis de O-glicanos y la O-glicosilación de mucinas, compuestas por diversas mucinas (MUC2/4/12/13/17) y enzimas acetil (GCNT3) y galactosiltransferasas (B3GNT7) encargadas de su O-glicosilación. Las mucinas son moléculas con alta O-glicosilación que contribuyen a la protección de la mucosa gástrica. Las mismas mimetizan la presencia de antígenos tisulares y sanguíneos en su estructura permitiendo la adhesión bacteriana por lo que actúan como un señuelo molecular (*molecular decoy*)<sup>256</sup>. En este sentido, *H. pylori* reconoce los antígenos sialil Le<sup>a</sup> y Le<sup>x</sup> presentes en O-glicanos de las mucinas gástricas y esta bacteria modula la glicosilación creando un microambiente favorable para la infección. Así, en pacientes con metaplasia intestinal incompleta, una alteración en la glicosilación de MUC5AC y MUC6 se correlaciona con un aumento en la hidrosolubilidad de las mismas y del mucus gástrico y alta expresión de oligosacáridos que expresan antígenos sialil Le<sup>a</sup> y Le<sup>x</sup>. Mientras que pacientes infectados asintomáticos clínicamente son capaces de evitar estadios más avanzados de la infección por la adaptación de la glicosilación de dichas mucinas; como la expresión de glicanos αGlcNAc (α1,4-N-acetylglucosamina) con acción antibiótica en la mucosa gástrica con el objetivo de impedir su penetración en el epitelio y limitar la infección al mucus gástrico<sup>297</sup>. El ratón knock-out *A4gnt(-/-)* que no expresa glicanos αGlcNAc manifiesta la casaca de Correa sin infección por *H.pylori*<sup>298</sup>.

Un reciente trabajo también halló una O-glicosilación anormal en plasminógeno sérico de pacientes con gastritis y MI<sup>299</sup>. Adicionalmente, en diversos adenocarcinomas MUC1 es sobre-expresada en forma sub-glicosilada con O-

glicanos truncados alterando su tráfico endocítico y localización subcelular <sup>300</sup>. También es conocido que una glicosilación alterada ocurre en el cáncer en general debido a la desregulación de la expresión de glicosiltransferasas y alteración en la actividad de glicosidasas y chaperonas; contribuyendo dicha alteración a la inflamación, adhesión y señalización celular de los procesos tumorales <sup>301</sup>. Basado en las anteriores evidencias en los primeros estadios de la MI una glicosilación anómala trae consigo un mucus gástrico más hidrofílico que pudiera imposibilitar el atrapamiento de *H. pylori* y permitir su interacción directa y entrada al epitelio gástrico, hecho que traería como consecuencia una mayor severidad de la reacción inflamatoria antes descrita en el caso de la presencia intracelular de *H. pylori* <sup>75</sup>.

### Genes *leading edge* y filtrado de conjuntos génicos en el grupo MI-No CG

Al igual que se realizó previamente para el grupo MII-CG y con el objetivo de restringir los procesos funcionales, se determinaron en la MI-No CG los comunes a aquellos conjuntos génicos con  $ES > 0.65$  (Tabla anexa 4), valores *Rank at max* extremos de la distribución (Tabla 36) y que poseyeran genes *leading edge* sobre-expresados (Tabla 38). Los procesos comunes y más relevantes en la MI-No CG son el efecto Warburg no tumoral, metabolismo lipídico, diferenciación intestinal, glicosilación proteica aberrante, procesamiento antigénico, inflamación y apoptosis (Figura 55). Los conjuntos génicos de dichos procesos pudieran ser candidatos para una futura validación por qRT-PCR o inmuno-histoquímica.



**Figura 55.** Procesos funcionales comunes a conjuntos génicos con diferentes criterios de selección en la MI-No CG.

En conclusión, el análisis de enriquecimiento funcional mediante el programa GSEA ha permitido identificar una gran cantidad de procesos moleculares implicados en la carcinogénesis sobre-expresados en la MII-CG con respecto a la MIC-CG, lo que confirma su mayor riesgo de progresar a CG. Una parte de estos procesos son

específicos y otros comunes a los grupos MII-CG y MI-No CG (Figura 53), siendo estos últimos característicos de la MI independientemente de su progresión tumoral. Los procesos funcionales sobre-expresados en la MI-No CG poseen una alta similitud con resultados previos de la literatura. Mediante la superposición de diferentes criterios de selección hemos reducido la gran cantidad de procesos funcionales existentes en MII-CG y MI-No CG a los más relevantes.

### 3.2.2 Ingenuity Pathway Analysis

#### MII y MIC que progresan a cáncer gástrico

El programa informático IPA permitió identificar por una metodología alternativa vías de señalización celular y mecanismos funcionales implicados en la progresión de los subtipos histológicos de metaplasia intestinal completa e incompleta al cáncer gástrico. Estos resultados son parcialmente diferentes de los obtenidos con el GSEA, pudiendo deberse a sus diferentes metodologías, ya que el IPA trabaja solamente con genes significativos diferencialmente expresados y sus *Fold Changes*, los cuales son buscados en la base de datos *Ingenuity® Knowledge Base*, la cual posee información de mecanismos patológicos a nivel molecular, celular y de organismo. Los resultados del IPA evidencian una serie de vías canónicas sobre-expresadas en ambos subtipos histológicos de MI que progresan a CG, a diferencia del GSEA donde solo fueron obtenidas vías sobre-expresadas en la MII-CG. Dichas vías comunes son la presentación antigénica, señalización por TNFRSF4, comunicación entre células innata y adaptativa, enfermedad tiroidea autoinmune y maduración del fagosoma, evidenciando la relevancia de la respuesta inmune innata y adquirida en la progresión tumoral (Tabla 40). Sin embargo, se observan una menor cantidad de procesos carcinogénicos respecto a los resultados del GSEA.

La mayoría de las vías obtenidas son más significativas y con un mayor número de moléculas en la MII-CG que en la MIC-CG (Figura 40); hecho que coincide con el mayor riesgo de progresión a CG de la MII en comparación con la MIC obtenido en estudios epidemiológicos<sup>57 60</sup>. Todas las vías sobre-expresadas en MII-CG y MIC-CG (Tabla 40) contienen a moléculas sobre-expresadas de HLA clase I/II y componentes de anticuerpos. De manera coincidente, en el análisis del GSEA también se obtienen vías implicadas en la presentación antigénica (MII-CG y MI-No CG), maduración del fagosoma (fagocitosis, MII-CG) y maduración de células dendríticas (MII-CG).

TNFRSF4 es el receptor del factor de necrosis tumoral el cual participa en la respuesta de células T CD4+, la proliferación de células B dependiente de células T, e inhibición de la apoptosis; respecto al mismo no existen trabajos previos en la MI o CG. Las células dendríticas gástricas reconocen a *H. pylori* y activan a células T para iniciar la respuesta inmune<sup>302</sup>. El fagosoma es una vesícula formada alrededor



de una partícula absorbida por fagocitosis, en el interior de la cual los microorganismos son eliminados por enzimas lisosomales y la acidificación. *Helicobacter pylori* altera la maduración del fagosoma a través de *vacA* y resiste el ambiente ácido mediante la producción de amonio por la ureasa, contribuyendo a la evasión de la respuesta inmune <sup>303</sup>.

Otra de las vías canónicas sobre-expresadas en MIC-CG y MII-CG es la enfermedad tiroidea autoinmune. Basándose en la alta correlación entre la incidencia de la enfermedad tiroidea y la mortalidad por CG, se ha propuesto la hipótesis de que algún factor sea producido en la tiroides como resultado de esta enfermedad, el cual afecte la mucosa gástrica contribuyendo a la gastritis crónica <sup>304</sup>. Es muy interesante que existan 3 conjuntos génicos sobre-expresados en MII-CG compuestos por cáncer de tiroides, del cual la enfermedad tiroidea autoinmune es su principal factor de riesgo. Previamente un estudio proteómico había identificado la sobre-expresión de la principal enzima tiroidea, triyodotironina (T3), y HIF1A en CG proponiendo que la expresión de este último sea inducida por T3 vía PI3K conllevando a un efecto Warburg <sup>305</sup>.

En ambos grupos, MII-CG y MIC-CG, las enfermedades reconocidas por el análisis funcional del IPA fueron las inmunológicas e inflamatorias, mientras que las funciones moleculares comunes fueron la proliferación celular, señalización e interacción célula-célula y muerte y sobrevivencia celular. El análisis de reguladores *upstream* en la MII-CG confirma que es posible que otros genes homeobox diferentes de *CDX1/2*, como en este caso *HOXC11*, contribuyan a la diferenciación intestinal. Se ha observado previamente que otro miembro de esta familia, *HOXA13*, se encuentra sobre-expresado en la MIC-CG y MI-No CG <sup>264</sup>.

La red molecular construida por el IPA con mayor *score* en la MII-CG fue asociada con enfermedades del tejido conectivo e inflamatorias, estando compuesta por 21 moléculas centrales diferencialmente expresadas en la MII-CG y con una extensa inter-relación (Figura 40A). Por ejemplo, HLA-A, C, DRB1, DQB1, IL1R2, FABP5 interactúan con inmunoglobulinas. También varias moléculas interactúan con el oncogén *RHOA*, mutado frecuentemente en CG <sup>46</sup> y sub-expresado en la MII-CG. Otras moléculas sobre-expresadas interaccionan con el factor transcripcional NFκB, inductor de la expresión de citoquinas pro-inflamatorias. El IPA no construyó ninguna red molecular de alto *score* (>15) en la MIC-CG, debido a la pequeña cantidad y baja inter-relación de los genes diferencialmente expresados en este grupo, lo cual apoya el fenómeno de un mayor riesgo de progresión a CG de la MII respecto a la MIC.

### MI-No CG en comparación con mucosa gástrica sana

Al igual que en el análisis con el GSEA las muestras de MII y MIC que no progresan a CG fueron unidas y comparadas con los controles sanos, con el objetivo de identificar mecanismos funcionales característicos del grupo MI-No CG. Observamos que en esta comparación aparecen sobre-expresadas 3 vías pertenecientes a la familia de receptores nucleares pregnano X (PXR/NR1I2), farnesoide X (FXR/NR1H4), hepático X (LXR) y retinoide X (RXRG). PXR es un factor transcripcional activado por compuestos xenobióticos, tras lo cual actúa como un regulador transcripcional de los citocromos P450, encargados de metabolizar y eliminar los mismos. Coincidentemente está sobre-expresada la vía metabolismo de xenobióticos, obtenida como desregulada empleando el programa GSEA y a nivel de genes individuales. El receptor FXR es miembro de la subfamilia de receptores nucleares activados por esteroides, siendo un receptor de ácidos biliares y activador transcripcional de la síntesis y transporte de los mismos.

Las funciones moleculares del IPA confirman, al igual que el GSEA y los genes diferencialmente expresados, que el metabolismo lipídico está incrementado en la MI-No CG, siendo dos de los reguladores *upstream* en este grupo los factores transcripcionales nucleares hepáticos HNF1A/4A, inductores de este proceso. En la red de mayor *score* (29) se observa a dicho proceso como el más relevante, existiendo un nodo central compuesto por HNF4A sobre-expresado e interactuando directamente con otras moléculas sobre-expresadas del metabolismo lipídico como APOA, APOB, MTP, APOC3 y ABCG8; así como, con el citocromo P450 CYP3A4 y el receptor pregnano NR1I2, evidenciando también la relevancia del metabolismo de xenobióticos en dicha red. Además, la función transporte molecular y bioquímica de moléculas pequeñas pudiera estar reflejando el hecho de que en este grupo fue obtenido el efecto Warburg como el proceso molecular más frecuentemente sobre-expresado en el análisis de GSEA. De esta manera, en la MI existe un incremento del metabolismo lipídico corroborado mediante el análisis de genes diferencialmente expresados, el de funciones y redes moleculares del IPA y el GSEA.

Otra vía de interés implica al metabolismo de la hormona tiroidea triiodotironina (T3), reguladora de la tasa metabólica general y de proteínas, grasas y carbohidratos, la que está sobre-expresada en el CG<sup>305</sup>. T3 induce la expresión de HIF1A, un regulador transcripcional de la respuesta hipóxica con dos vías sobre-expresadas en el GSEA. HIF1A induce la expresión de enzimas glicolíticas y transportadores de glucosa, como la hexoquinasa, SLC2A5, SLC5A9 y SLC5A1 que están sobre-expresados en la MI-No CG y son una manifestación del efecto Warburg en este grupo. De manera muy interesante se observan dos vías dedicadas a la degradación de melatonina, una hormona con un rol en la motilidad gastrointestinal, protección mucosal ante la secreción ácida y eliminación de

radicales libres<sup>306</sup>, por lo que al encontrarse sub-expresada en la mucosa gástrica no ejercería su actividad protectora y anti-oxidante.

En conclusión, el análisis mediante IPA identificó una serie de procesos comunes con el GSEA, lo que les confiere mayor importancia. En el caso de la MI-No CG, existe un incremento del metabolismo lipídico previamente reportado. Nuevos procesos identificados incluyen al metabolismo de xenobióticos, hormona tiroidea (probablemente mediadora del efecto Warburg) y la degradación de melatonina. Con respecto a ambos subtipos de MI que progresan a CG se identificaron procesos relacionados con la respuesta inmune (pero sin un rol carcinogénico específico) y la enfermedad tiroidea autoinmune, esta última puede interpretarse como un estadio de mayor severidad respecto al observado con la desregulación de la hormona tiroidea en la MI-No CG. Además, es posible que otros genes homeobox como *HOXC11* contribuyan a la diferenciación intestinal en la MI.

### **3.3. Validación de la expresión génica por qRT-PCR**

Se intento validar por qRT-PCR las diferencias de expresión entre casos y controles de 37 genes (MII-CG N=19, MIC-CG N=3, MI-No CG N=18) significativos del microarray de expresión, así como de 3 integrantes de la familia *Schlafen* entre los que se encuentra *Schlafen5*. El mismo no estaba diferencialmente expresado en el microarray; sin embargo se encuentra sobre-expresado en la qRT-PCR; lo cual concuerda con su sobre-expresión inmuno-histoquímica en ambos subtipos histológicos de MI-CG en comparación con MI-No CG descrita en la sección 4 de Discusión. Se han validado las diferencias de expresión en un 56.75% de los genes elegidos, la baja efectividad de la qRT-PCR la atribuimos a la mala calidad de las muestras de ARN procedentes de tejido parafinado con una media de 16.4 años de antigüedad y al hecho de que la plataforma utilizada no ha sido validada con dicho ARN de tan mala calidad sino que aconsejan usar muestras de RIN alto (7-8) (comunicación personal).

## **4. Estudio de expresión génica e inmuno-histoquímica de *Schlafen5* en lesiones precursoras y cáncer gástrico**

Existen numerosos biomarcadores inmuno-histoquímicos de metaplasia intestinal como las mucinas MUC1, MUC5AC, MUC2<sup>14</sup> y TFF3<sup>52</sup>. Estos se han establecido empleando muestras de las que no se conoce su evolución durante el seguimiento clínico. Sin embargo, biomarcadores de la progresión desde metaplasia intestinal a cáncer gástrico son escasos<sup>307-309</sup>. Ello es debido, en parte, a la dificultad de realizar estudios de seguimiento con colecta de muestras biológicas, cuya ejecución requiere mucho tiempo y dinero, así como por los esfuerzos que han de llevarse a cabo para que la calidad del seguimiento minimice la pérdida de los pacientes que se monitorizan.

En un modelo murino de MI donde la vía de señalización celular Shh se encuentra activada y desarrolla metaplasia tipo SPEM debido a la infección por *Helicobacter felis*, existe una expresión incrementada de la proteína Schlafen4 en la mucosa gástrica en comparación con el modelo murino *knock-out* (Gli1<sup>-/-</sup>) que tiene inactivada dicha vía y no desarrolla metaplasia<sup>85</sup>. En un intento traslacional de corroborar dichos resultados en pacientes afectados de LPGs se examinó si el homólogo humano de Schlafen4, Schlafen 12, se expresaba en LPGs humanas obteniéndose un resultado negativo (comunicación personal). Sin embargo, otro miembro de esta familia, Schlafen5, es inducido por IFN $\alpha$  en melanocitos y melanoma humano<sup>88</sup>. El IFN $\alpha$  es una citoquina sobre-expresada en LPGs a consecuencia de la infección por *H. pylori*<sup>310</sup>. Basado en este hecho, la hipótesis fue investigar si Schlafen5 es inducido por IFN $\alpha$  en células inflamatorias del estroma gástrico y conocer su expresión en diversas LPGs humanas, con especial énfasis en los subtipos histológicos de MI que progresan a CG, provenientes de un estudio longitudinal de seguimiento de LPGs durante 12.5 años realizado en el hospital de Soria.

Las guías de manejo clínico de la metaplasia intestinal recomiendan la vigilancia endoscópica cada 3 años para una metaplasia intestinal extensa<sup>58</sup>. Teniendo en cuenta que las tasas de progresión a CG para esta lesión pueden llegar hasta un máximo del 10%<sup>56</sup>, un biomarcador inmuno-histoquímico que pudiera brindar información sobre cuales pacientes con esta lesión son más susceptibles a desarrollar CG y que se complementa con el subtipo histológico de MI (completa o incompleta) es de gran relevancia.

Schlafen5 es inducido por IFN $\alpha$  en melanocitos y melanoma humano<sup>88</sup>; sin embargo, no se conocía si el mismo era inducido por esta citoquina en células inflamatorias que componen el estroma gástrico. Hecho que se confirmó a nivel de ARNm en las líneas celulares humanas de monocitos (HL60) y células T (Jurkat) (apartado 4.1 de la sección Resultados). Además, en las células Jurkat se confirmó que dicha expresión también ocurría a nivel proteico. Ello indica que es posible que en la mucosa gástrica humana Schlafen5 sea inducido por la acción del IFN $\alpha$ , una citoquina pro-inflamatoria elevada como consecuencia de la infección por *Helicobacter pylori*<sup>310</sup>. La incubación de ambas líneas celulares con *H. pylori* evidencia un incremento apenas significativo (HL60, p=0.039 y Jurkat p=0.044) de la expresión de Schlafen5, por lo que es probable que el aumento de este ocurra indirecta y posteriormente en el tiempo como consecuencia de la inducción de IFN $\alpha$  debido a la infección por *H. pylori*.

Es conocido que los miembros de la familia Schlafen se expresan en células T murinas<sup>83</sup>; sin embargo, se desconocía si esto también ocurriría en la mucosa gástrica humana. La doble tinción inmuno-histoquímica realizada para detectar

marcadores de superficie (*cluster of differentiation*) de células T, B y mieloides demostró que Schlafen5 se expresa mayoritariamente por células T y en mucha menor cuantía por macrófagos u otras células mieloides (apartado 4.6 de la sección Resultados) presentes en el infiltrado inflamatorio gástrico, sobre todo en los centros germinales indicativos de una infección por *H. pylori*.

A continuación, se valoró el patrón de expresión de Schlafen5 en un microarray de tejido formado por diferentes LPGs (GNA, GCA, MI, DISP) y CG. Mostrando una tendencia a aumentar según los estadios de la cascada de Correa y una disminución en el CG. Estas muestras no tienen datos clínicos y mostraron baja tinción, debido a que provienen de gastrectomías con un evidente daño tisular, por lo que es esperable una menor reactividad antigénica. Sin embargo, este resultado preliminar fue un primer indicio para la posterior evaluación de Schlafen5 en muestras bien caracterizadas clínicamente del estudio de seguimiento de Soria. Las mismas confirmaron el mismo incremento de expresión en el estroma gástrico de acuerdo con la severidad (GNA, GCA, MI) de las lesiones precursoras y una disminución en el cáncer gástrico (Tabla 45, Figura 48). Esta tinción es significativamente diferente (Kruskal-Wallis,  $p=0.0033$ ) en los grupos analizados (GNA, GMA, MIC, MII y CG). Es importante destacar que el grupo de metaplasia intestinal que progresa a cáncer gástrico (MI-CG) muestra una expresión significativamente mayor con respecto a la metaplasia intestinal que no progresa a cáncer gástrico (MI-No CG) (Mann-Whitney  $p<0.0001$ , Tabla 45). Similar resultado se obtiene al comparar los subtipos de MI que progresan (MIC/MII-CG) respecto a los que no progresan (MIC/MII-No CG) ( $p<0.0001$ ).

Adicionalmente, el score estromal H de Schlafen5 fue categorizado y se construyeron dos modelos logísticos de predicción de la ocurrencia de CG que se distinguían por la inclusión, o no, de la variable categórica de SLFN5 en el modelo, que incluía las variables sexo, edad y diagnóstico histológico. La probabilidad de clasificar correctamente a los pacientes que desarrollarán CG desde la MI es significativamente mayor cuando se añade el score estromal de Schlafen5 al diagnóstico histológico (AUC=91.2% vs 78.3%,  $p<0.0001$ ) (Figura 49), siendo el incremento del area bajo la curva de un 12.9%. Este mismo comportamiento ocurre para la MII (AUC=95.3% vs 75.6%,  $p<0.0001$ ), pero en este caso el incremento es mayor alcanzando un 19.7% (Figura 50). Ello indica que la adición de Schlafen5 al modelo de predicción incrementa significativamente la probabilidad de clasificación de la muestra como potencial a desarrollar cáncer gástrico a lo largo del tiempo, y sugiere que esta proteína pudiera ser un biomarcador para identificar pacientes que desarrollarán CG a partir de metaplasia intestinal y que pudieran ser sujetos a una vigilancia clínica mayor.

Es importante destacar que el grupo MII-CG posee más muestras y mayores valores de score H en las categorías Débil+ (73.68) y Moderado+ (10.53) en comparación con el grupo MIC-CG (Débil+ (66.67) y Moderado+ (0)), y que el incremento del área bajo la curva para el modelo de MII-CG es mayor que el que se obtiene para el modelo MI-CG que agrupa los dos subtipos de metaplasia (MII y MIC). Todo ello evidencia una mayor expresión de Schlafen5 en las MII con respecto a las MIC que progresan a CG; así como, que la expresión de SLFN5 contribuye a mejorar el modelo de predicción del CG en comparación a cuando se analizan sólo ambos tipos histológicos de MI. El modelo de predicción basado solamente en la MIC no se realizó debido al bajo tamaño de muestra de este grupo (N=6).

Se ha mencionado que la MII presenta un mayor riesgo de progresión a CG en comparación con la MIC, basado en una revisión de estudios epidemiológicos<sup>60</sup>, así como en estudios de seguimiento de LPGs realizados en Soria (MII vs MIC HR 11.3 [3.8-33.9])<sup>57</sup> y en toda España (MII vs MIC HR 2.75 [1.06-6.26]) (artículo pendiente de aceptación). Esta diferencia de riesgo es apoyada por diversos procesos moleculares relacionados con la oncogénesis que diferencian los subtipos histológicos de la MI y que pudieran condicionar el mayor riesgo de progresión a CG desde la MII, los cuales fueron discutidos previamente en el estudio de microarray de expresión. Estos son la presencia de *H. pylori* de localización intracelular en la MII y no en la MIC<sup>75</sup>, menor expresión de *CDX2* en la MII vs MIC<sup>65</sup>, menor activación de la vía de señalización Shh en la MII<sup>70</sup>, mayor actividad telomerasa en la MII<sup>74</sup>, mayor expresión de *SOX2* en la MII<sup>288</sup>, incremento de mutaciones en TP53 e inestabilidad de microsatélites<sup>52</sup> en la MII. En adición a los mismos, el estudio de expresión por microarray identificó una desregulación de otros procesos carcinogénicos en la MII en lugar de la MIC. Y la tinción estromal de Schlafen5 también indica un efecto diferencial de mayor riesgo de progresión a CG para la MII respecto a la MIC, con plausibilidad biológica en la carcinogénesis como se verá más adelante.

La tinción de Schlafen5 fue consistentemente mayor en el estroma con respecto a las glándulas gástricas, en consonancia con la expresión de los miembros de la familia Schlafen en las células T<sup>83</sup>, las cuales forman parte del estroma gástrico inflamado. Las glándulas gástricas ubicadas alrededor de un estroma muy positivo también se tiñen, pero con menor intensidad y no muestran un incremento acorde a la cascada de Correa, por lo que concluimos que su expresión en esta localización no era informativa y por tanto no fue analizada estadísticamente.

Schlafen5 pertenece al grupo III (masa molecular 100-104 kDa) de la familia de genes Schlafen; caracterizado por la presencia de un dominio helicasa de ARN y una señal de localización nuclear (RKRRR)<sup>89</sup> lo que sugiere que son ATPasas dependientes de ARN pertenecientes a la superfamilia de las helicasas. En

consonancia con dicha función observamos la expresión del mismo en el núcleo de las células T en lugar del citoplasma. También de manera coincidente con nuestros resultados Schlafen5 también es inducido por IFN $\alpha$  en carcinoma renal <sup>311</sup>, melanocitos y melanoma humano, estando sub-expresado en este último con respecto a melanocitos <sup>88</sup>. Schlafen5 tiene un efecto inhibitorio sobre el crecimiento e invasión tumoral en melanoma maligno, por lo que en este tumor Schlafen5 funciona como supresor tumoral y su baja expresión endógena favorece la tumorigénesis <sup>88</sup>.

Un estudio muy reciente implica a Schlafen5 como un represor de la motilidad e invasión tumoral (a través del control negativo de las metaloproteinasas MMP1 y MMP13) en líneas celulares de carcinoma renal humano. A diferencia del estudio anterior, cuando Schlafen5 fue silenciado no hubo una disminución en la proliferación celular sino un cambio morfológico hacia células más pequeñas con pérdida de fibras de estrés (la proteína ejerce un control sobre el citoesqueleto) y una motilidad celular incrementada. Los procesos moleculares más afectados por su silenciamiento fueron de cáncer, mecanismos de adhesión focal y uniones célula-célula, estando correlacionada su mayor expresión con una mayor supervivencia en pacientes de carcinoma renal, posiblemente ejerciendo un rol de supresor tumoral <sup>315</sup> al igual que en melanoma maligno. Los posibles mecanismos moleculares mediante el cual Schlafen5 regula a MMP1 y MMP13 es que funcione como un represor transcripcional de las mismas y/o ejerza efecto en las vías de señalización NOTCH/TGF- $\beta$  y MAPK las cuales controlan la expresión de dichas metaloproteinasas <sup>311</sup>.

La mayor expresión de Schlafen11 también esta correlacionada con un aumento en la supervivencia de pacientes con cáncer de ovario tratados con cisplatino, sugiriendo nuevamente un rol de las proteínas Schlafen como potenciales supresores tumorales <sup>312</sup>. En el CG también existe una baja expresión de Schlafen5 similar a la existente en melanoma y carcinoma renal, por lo que al parecer también se comporta como supresor tumoral en esta patología.

En las muestras que progresan a CG desde la MI Schlafen5 se encuentra sobre-expresado respecto a las que no progresan, por lo que a diferencia del CG, carcinoma renal y melanoma en las mismas no se comporta como un supresor tumoral sino posiblemente promoviendo la progresión tumoral indirectamente como un regulador negativo de la activación de células T. A diferencia de los estudios anteriores, las muestras de LPGs donde se evaluó la expresión de Schlafen5 no son tumorales, sino que están afectadas por cambios preneoplásicos; provienen de un tejido diferente a los anteriores y las células donde tiene mayor expresión no son epiteliales sino estromales, razones que pudieran explicar las diferencias de comportamiento.

Los genes Schlafen (grupos I, II y III) se encuentran sobre-expresados durante la diferenciación de timocitos a células T y en células T periféricas no activadas, sin embargo ocurre una disminución en su expresión en células T activadas por antígenos <sup>89</sup>. La expresión transgénica de SLFN1 (prototipo de esta familia) en timocitos provoca arresto del ciclo celular y alta tasa de muerte celular lo cual se traduce en una celularidad reducida del timo. Por lo que es un inhibidor del crecimiento celular y esta correlacionada positivamente su alta expresión con el estado de quiescencia <sup>83</sup>. SLFN1 causa arresto del ciclo celular en fibroblastos inhibiendo la transcripción de la ciclina D1 <sup>82</sup>. La sobre-expresión de SLFN8 (miembro del grupo III al que pertenece Schlafen5) en células T, bloquea parcialmente la proliferación de las mismas <sup>89</sup>. Schlafen11, miembro del mismo grupo III al que pertenece Schlafen5, también es responsable del arresto del ciclo y muerte celular de líneas celulares tumorales NCI-60 tratadas con quimiofármacos <sup>311</sup>.

Teniendo en cuenta la similitud funcional de los miembros de la familia Schlafen, la elevada expresión de Schlafen5 posiblemente inducida por IFN $\alpha$  a consecuencia de la infección por *H. pylori* en pacientes con metaplasia intestinal podría disminuir la respuesta inmune ante el crecimiento tumoral, manteniendo a las células T en un estado de inactivación que facilite la progresión tumoral a lo largo del tiempo. En este sentido es importante destacar que la respuesta inmune ante la infección por *H. pylori*, en la gastritis crónica y el CG es ejercida fundamentalmente por células T en lugar de B <sup>32</sup>. Ratones sin células T (RAG2 <sup>-/-</sup>) tienen mayor incidencia de cáncer (tanto inducidos como espontáneos) en comparación con ratones inmunocompetentes <sup>313</sup>. Lo cual apunta a la importancia de las células T en el proceso tumoral y que Schlafen5 pudiera inducir una inmunodeficiencia similar a la reportada en modelos murinos que influya positivamente en la progresión desde MI a CG.

Teniendo en cuenta que las guías de manejo clínico de LPGs recomiendan la vigilancia endoscópica cada 3 años para la metaplasia intestinal extensa <sup>58</sup>, los pacientes con dicha lesión y un alto nivel de Schlafen5 podrían ser considerados para una vigilancia endoscópica en un intervalo de tiempo menor.

Los estudios inmunohistoquímicos con marcadores pronósticos utilizando muestras de estudios de seguimiento en la carcinogénesis gástrica son muy escasos; tan solo 3 artículos han resultado de una extensa búsqueda en PubMed con diferentes combinaciones de términos. De manera similar a nuestros resultados, *Goto T et al* realizaron inmuno-histoquímica de la óxido nítrico sintasa inducible (iNOS) y nitrotirosina en 93 pacientes con atrofia gástrica y MI posteriormente diagnosticados con cáncer gástrico (N=34, *H. pylori* +: grupo A) o no (59 *H. pylori* +: grupo B, *H. pylori* -: grupo C) al menos 2 años más tarde. La expresión de estas



proteínas fue significativamente mayor en el grupo A los que más tarde desarrollaron el proceso tumoral<sup>307</sup>. En otro estudio fue evaluada la expresión inmuno-histoquímica del antígeno tumoral MG7-Ag y COX2 entre la MI que progresa a displasia y MI que se mantiene estable en el seguimiento. Las muestras con una tinción positiva para ambas proteínas tienen un riesgo significativo de progresión (OR 22.7,  $p < 0.005$ ) en comparación con los negativos<sup>308</sup>.

También la expresión inmuno-histoquímica de c-Met fue testada en pacientes con MI que progresaron a displasia (MI-DISP) y en displasia que progreso a CG durante un seguimiento máximo de 10 años en China. c-Met se sobre-expresó significativamente en la MI-DISP (71.4%) en comparación con la MI-MI (40%), OR 7.42 (2.08-26.4),  $p < 0.05$ <sup>309</sup>. Sin embargo ninguno de estos genes son significativos en el estudio de microarray de expresión de los subtipos de MI que progresan a CG respecto a los que no progresan. Las causas de dichas diferencias pueden estar en múltiples factores como los tipos de lesiones al reclutamiento y final de seguimiento, la susceptibilidad genética, infección por *H. pylori* y sus factores de virulencia *cagA* y *vacA*, hábitos alimenticios de las poblaciones de estudio, el tiempo de seguimiento y de establecida cada lesión, edad, etc. Coincidentemente con estos estudios y el nuestro propio se obtuvo un incremento significativo de la expresión de ARNm por qRT-PCR de *Schlafen5* ( $p = 0.0025$ ) y *Schlafen12* ( $p = 0.055$ ) en ambos tipos de MI que progresan a CG respecto a las que no progresan. Nuestro trabajo y los anteriores evidencian que en la metaplasia intestinal que avanza a displasia o CG existe una expresión incrementada de ciertas proteínas años antes de la aparición del cáncer gástrico.

Concluimos que *Schlafen5* pudiera ser inducido por la acción del  $IFN\alpha$  en células T de la mucosa gástrica, e influir en la progresión desde MI a CG debido a la disminución de la respuesta anti-tumoral de células T. Su tinción inmunohistoquímica e interpretación cuantitativa en el estroma gástrico de pacientes con metaplasia intestinal, especialmente en los que tienen una MII, podría ser un biomarcador útil para detectar aquellos con mayor riesgo de desarrollar un cáncer gástrico en el futuro. Estudios de validación en series independientes de pacientes con metaplasia intestinal que progresan y no progresan a CG; así como tinción inmunohisto-química de marcador CD30 para chequear si las células T se encuentran activas funcionalmente<sup>320</sup> son necesarios para apoyar esta afirmación.

# **CONCLUSIONES**



1. El análisis de SNPs y haplotipos indica que el gen *CD14* se asocia estadísticamente como un factor de riesgo del CG del cardias y del tipo intestinal mientras que el gen *NOD2* es un factor protector del CG no cardias.
2. Se confirma que variabilidad genética en *MUC2*, observada en un estudio previo, se asocia con la regresión de las lesiones precursoras de cáncer gástrico. Aunque la asociación no supera la corrección para comparaciones múltiples, también se replican las asociaciones entre variantes y/o haplotipos de *CD14* y *TFF2* con un mayor y menor riesgo, respectivamente, de progresión de las lesiones; y la asociación entre la variabilidad en *TFF1* y mayor riesgo de progresión o menor probabilidad de regresión de las LPGs.
3. Otros genes asociados con la evolución de lesiones precursoras del CG pero que no superan la corrección para comparaciones múltiples, son *CD14*, *PTGS2* y *TFF1* que se asocian con la progresión, mientras que *CDH1*, *TFF1* y *MUC2* se asocian con la regresión.
4. Otros genes que se observan asociados por primera vez con la evolución de las lesiones precursoras del CG aunque no superan la corrección para comparaciones múltiples, son *DNAH11*, *PRKAA1* y *PTGS2*, que se asocian con la progresión; y *PTGES*, *CDX2*, *CDH1*, *MAP3K14* y *PTPN11*, que se asocian con la regresión.
5. El perfil transcripcional de la metaplasia intestinal que no progresa a CG respecto a la mucosa gástrica sana muestra grandes diferencias en los niveles de expresión. Nuevos genes identificados como sobre-expresados incluyen a los transcritos *TRIM*, *TMEM* y una sub-expresión de ARNs nucleolares *SNORDs116*. Los procesos moleculares más frecuentes o específicos que se confirman en nuestro estudio son el metabolismo lipídico, la diferenciación intestinal, metabolismo de xenobióticos y reflujo gastroesofágico. Nuevos procesos moleculares identificados son el efecto Warburg no tumoral, glicosilación aberrante de mucinas, desregulación de la hormona tiroidea y degradación de melatonina.
6. El perfil transcripcional de las MI que progresan a CG respecto a las que no progresan muestra pequeñas diferencias en los niveles de expresión en comparación con la MI que no progresa respecto a la mucosa gástrica sana.
7. La MII que progresa a CG presenta una mayor cantidad de genes sobre-expresados relacionados con el proceso tumoral con respecto a la MIC que progresa a CG, lo cual coincide a nivel molecular con su mayor riesgo de progresión a CG. Los genes sobre-expresados relacionados con la carcinogénesis son moléculas inflamatorias y del sistema del complemento, chaperonas, el oncogen *RAN*, un componente de la telomerasa *TERT*, reguladores del supresor tumoral *RB1* (*EIF3D* y *EIF5B*), el inductor de metástasis gástrica *CAV1*, un regulador de la angiogénesis

(MYOF) y de la proliferación de células madres (GNL3). Funciones sub-expresadas están dadas por el regulador negativo de la expresión de HLA, IK, la función antimicrobiana (LCT) y la regulación del ciclo celular (CDKL3, RBBP7) y del metabolismo del hierro (CP).

8. Los procesos moleculares enriquecidos funcionalmente más frecuentes o específicos de la MII-CG están representados por la activación del ciclo y la proliferación celular, oncogenes, supresores tumorales, invasión y metástasis, adhesión celular, endocitosis o fagocitosis, genes regulados por la insulina, degradación de ARNm por mecanismo *Non sense mediated decay* y proteica por el proteasoma, y enfermedad tiroidea autoinmune.

9. Los genes relacionados con la carcinogénesis en la MIC-CG se restringen a un promotor de la proliferación celular (GFBP5) y un factor anti-apoptótico que promueve el crecimiento tumoral (OLFM4). No se encuentran procesos moleculares enriquecidos significativamente en este grupo.

10. Es posible que otros genes *homeobox* diferentes de CDX1/CDX2 como HOXA13 (MI-No CG) y HOXC11 (MII-CG) contribuyan a la diferenciación intestinal en la MI.

11. Se han validado las diferencias de expresión de algunos genes desregulados en MII-CG, MIC-CG y MI-No CG por la técnica qRT-PCR.

12. La expresión génica de *Schlafen5* es inducida por IFN $\alpha$  en monocitos y células T humanas, y en la metaplasia intestinal la proteína *Schlafen5* es expresada mayoritariamente también por células T del estroma gástrico.

13. *Schlafen5* se sobre-expresa en la MI que progresa a CG con respecto a la que no progresa a CG y es posible que contribuya a la progresión tumoral. Su tinción inmunohistoquímica en pacientes con MI, mayormente en la MII, aumenta significativamente la probabilidad de identificar a pacientes con mayor potencial riesgo de desarrollar cáncer gástrico a lo largo del tiempo.

## **REFERENCIAS BIBLIOGRÁFICAS**



1. Piazuolo M. B. & Correa P. Gastric cancer: Overview. *Colomb. Med.* **44**, 192–201 (2013).
2. IARC *Helicobacter pylori* Working Group (2014). *Helicobacter pylori eradication as a strategy for preventing gastric cancer*. Lyon: International Agency for Research on Cancer (IARC Working Group Reports, No. 8).
3. Camargo, M. C. *et al.* Sex hormones, hormonal interventions, and gastric cancer risk: A meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **21**, 20–38 (2012).
4. Aragonés, N. *et al.* The striking geographical pattern of gastric cancer mortality in Spain: Environmental hypotheses revisited. *BMC Cancer* **9**, 316 (2009).
5. Correa, P. Gastric cancer: Overview. *Gastroenterol. Clin. North Am.* **42**, 211–217 (2014).
6. Miyahara, R. *et al.* Prevalence and prognosis of gastric cancer detected by screening in a large Japanese population: Data from a single institute over 30 years. *J. Gastroenterol. Hepatol.* **22**, 1435–1442 (2007).
7. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
8. Hall, J. E. (2010). *Guyton and Hall Textbook of Medical Physiology*. 12<sup>ma</sup> edición. St. Louis: W. B. Saunders.
9. Schubert, M. L. & Peura, D. A. Control of gastric acid secretion in health and disease. *Gastroenterology* **134**, 1842–1860 (2008).
10. Hansson, L. E., Sparén, P. & Nyrén, O. Increasing incidence of carcinoma of the gastric cardia in Sweden from 1970 to 1985. *Br. J. Surg.* **80**, 374–377 (1993).
11. González, C. A. *et al.* *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**, 1320–1324 (2012).
12. Anderson, W. F. *et al.* Age-specific trends in incidence of noncardia gastric cancer in US adults. *JAMA* **303**, 1723–1728 (2010).
13. Lauren, P. The two histological main types of gastric carcinoma: Diffuse and so called intestinal-type carcinoma. *Acta Pathol. Microbiol. Scand.* **64**, 31–49 (1965).
14. Correa, P. & Piazuolo, M. B. The gastric precancerous cascade. *J. Dig. Dis.* **13**, 2–9 (2012).



15. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* **118**, 3030–3044 (2006).
16. Ghoshal, U. C., Chaturvedi, R. & Correa, P. The enigma of *Helicobacter pylori* infection and gastric cancer. *Indian J. Gastroenterol.* **29**, 95–100 (2011).
17. Stein, M., Ruggiero, P., Rappuoli, R. & Bagnoli, F. *Helicobacter pylori* CagA: From pathogenic mechanisms to its use as an anti-cancer vaccine. *Front. Immunol.* **4**, 328 (2013).
18. Bravo, L. E., van Doorn, L. J., Realpe, J. L. & Correa, P. Virulence-associated genotypes of *Helicobacter pylori*: Do they explain the African enigma? *Am. J. Gastroenterol.* **97**, 2839–2842 (2002).
19. Kodaman, N. *et al.* Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1455–1460 (2014).
20. Palframan, S. L., Kwok, T. & Gabriel, K. Vacuolating cytotoxin A (VacA), a key toxin for *Helicobacter pylori* pathogenesis. *Front. Cell. Infect. Microbiol.* **2**, (2012).
21. Higashi, H. *et al.* SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* **295**, 683–686 (2002).
22. Cover, T. L. & Blanke, S. R. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat. Rev. Microbiol.* **3**, 320–332 (2005).
23. Kaparakis, M. *et al.* Bacterial membrane vesicles deliver peptidoglycan to NOD1 in epithelial cells. *Cell. Microbiol.* **12**, 372–385 (2010).
24. Correa, P. & Houghton, J. Carcinogenesis of *Helicobacter pylori*. *Gastroenterology* **133**, 659–672 (2007).
25. Donnelly, J. M. *et al.* Mesenchymal stem cells induce epithelial proliferation within the inflamed stomach. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306**, G1075–G1088 (2014).
26. Camilo, V. *et al.* *Helicobacter pylori* and the BMP pathway regulate CDX2 and SOX2 expression in gastric cells. *Carcinogenesis* **33**, 1985–1992 (2012).
27. Mobley, H. L. T., Mendz, G. L. & Hazell, S. L. (2001). *Helicobacter pylori. Physiology and Genetics*. Washington DC: ASM.
28. Higashi, H. *et al.* Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14428–14433 (2002).

29. Kwok, T. *et al.* *Helicobacter* exploits integrin for type IV secretion and kinase activation. *Nature* **449**, 862–866 (2007).
30. Selbach, M., Moese, S., Hauck, C. R., Meyer, T. F. & Backert, S. Src is the kinase of the *Helicobacter pylori* CagA protein in vitro and in vivo. *J. Biol. Chem.* **277**, 6775–6778 (2002).
31. Brandt, S., Kwok, T., Hartig, R., König, W. & Backert, S. NF- $\kappa$ B activation and potentiation of proinflammatory responses by the *Helicobacter pylori* CagA protein. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9300–9305 (2005).
32. Moyat, M. Immune responses to *Helicobacter pylori* infection. *World J. Gastroenterol.* **20**, 5583 (2014).
33. De Elios, M. M. *et al.* Different cytokine profile and antigen-specificity repertoire in *Helicobacter pylori*-specific T cell clones from the antrum of chronic gastritis patients with or without peptic ulcer. *Eur. J. Immunol.* **27**, 1751–1755 (1997).
34. Fox, J. G. *et al.* Concurrent enteric helminth infection modulates inflammation and gastric immune responses and reduces *Helicobacter*-induced gastric atrophy. *Nat. Med.* **6**, 536–542 (2000).
35. Wang, Q. *et al.* Consumption of fruit, but not vegetables, may reduce risk of gastric cancer: Results from a meta-analysis of cohort studies. *Eur. J. Cancer* **50**, 1498–509 (2014).
36. Mendez, M. A. *et al.* Cereal fiber intake may reduce risk of gastric adenocarcinomas: The EPIC-EURGAST study. *Int. J. Cancer* **121**, 1618–1623 (2007).
37. González, C. A. & Agudo, A. Carcinogenesis, prevention and early detection of gastric cancer: Where we are and where we should go. *Int. J. Cancer* **130**, 745–753 (2012).
38. Agudo, A. *et al.* Impact of cigarette smoking on cancer risk in the European prospective investigation into cancer and nutrition study. *J. Clin. Oncol.* **30**, 4550–4557 (2012).
39. Abnet, C. C. *et al.* Non-steroidal anti-inflammatory drugs and risk of gastric and oesophageal adenocarcinomas: Results from a cohort study and a meta-analysis. *Br. J. Cancer* **100**, 551–557 (2009).
40. Gutiérrez-González, L. *et al.* The clonal origins of dysplasia from intestinal metaplasia in the human stomach. *Gastroenterology* **140**, 1251–1260 (2011).
41. Hayakawa, Y. *et al.* Mouse models of gastric cancer. *Cancers (Basel)*. **5**, 92–130 (2013).

42. Yuasa, Y. Control of gut differentiation and intestinal-type gastric carcinogenesis. *Nat. Rev. Cancer* **3**, 592–600 (2003).
43. Nobili, S. *et al.* Genomic and genetic alterations influence the progression of gastric cancer. *World J. Gastroenterol.* **17**, 290–299 (2011).
44. Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R. & Carneiro, F. E-cadherin alterations in hereditary disorders with emphasis on hereditary diffuse gastric cancer. *Prog. Mol. Biol. Transl. Sci.* **116**, 337–359 (2013).
45. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
46. Bass, A. J. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
47. Dixon, M. F., Genta, R. M., Yardley, J. H. & Correa, P. Classification and grading of gastritis. The updated Sydney System. *Am. J. Surg. Pathol.* **20**, 1161–1181 (1996).
48. Den Hoed, C. M. *et al.* Follow-up of premalignant lesions in patients at risk for progression to gastric cancer. *Endoscopy* **45**, 249–256 (2013).
49. Vannella, L. *et al.* Risk factors for progression to gastric neoplastic lesions in patients with atrophic gastritis. *Aliment. Pharmacol. Ther.* **31**, 1042–1050 (2010).
50. Filipe, M. I. *et al.* Intestinal metaplasia types and the risk of gastric cancer: A cohort study in Slovenia. *Int. J. Cancer* **57**, 324–329 (1994).
51. Slack, J. M. W. Metaplasia and somatic cell reprogramming. *J. Pathol.* **217**, 161–168 (2009).
52. Barros, R., Freund, J. N., David, L. & Almeida, R. Gastric intestinal metaplasia revisited: Function and regulation of CDX2. *Trends Mol. Med.* **18**, 555–563 (2012).
53. De Vries, A. C. *et al.* Epidemiological trends of pre-malignant gastric lesions: A long-term nationwide study in the Netherlands. *Gut* **56**, 1665–1670 (2007).
54. De Vries, A. C. *et al.* Gastric cancer risk in patients with premalignant gastric lesions: A nationwide cohort study in the Netherlands. *Gastroenterology* **134**, 945–952 (2008).
55. Den Hoed, C. M. *et al.* The prevalence of premalignant gastric lesions in asymptomatic patients: Predicting the future incidence of gastric cancer. *Eur. J. Cancer* **47**, 1211–1218 (2011).

56. De Vries, A. C., Haringsma, J. & Kuipers, E. J. The detection, surveillance and treatment of premalignant gastric lesions related to *Helicobacter pylori* infection. *Helicobacter* **12**, 1–15 (2007).
57. González, C. A. *et al.* Gastric cancer occurrence in preneoplastic lesions: A long-term follow-up in a high-risk area in Spain. *Int. J. Cancer* **127**, 2654–2660 (2010).
58. Dinis-Ribeiro, M. *et al.* 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. *Virchows Arch.* **460**, 19–46 (2012).
59. Kang, K. P. *et al.* Role of intestinal metaplasia subtyping in the risk of gastric cancer in Korea. *J. Gastroenterol. Hepatol.* **24**, 140–148 (2009).
60. González, C. A., Sanz-Anquela, J. M., Gisbert, J. P. & Correa, P. Utility of subtyping intestinal metaplasia as marker of gastric cancer risk. A review of evidence. *Int. J. Cancer* **133**, 1023–1032 (2013).
61. You, W. C. *et al.* Evolution of precancerous lesions in a rural Chinese population at high risk of gastric cancer. *Int. J. Cancer* **83**, 615–619 (1999).
62. Correa, P., Piazuelo, M. B. & Wilson, K. T. Pathology of gastric intestinal metaplasia: Clinical implications. *Am. J. Gastroenterol.* **105**, 493–498 (2010).
63. Fujii, Y. *et al.* CDX1 confers intestinal phenotype on gastric epithelial cells via induction of stemness-associated reprogramming factors SALL4 and KLF5. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20584–20589 (2012).
64. Mesquita, P. *et al.* Metaplasia, a transdifferentiation process that facilitates cancer development: The model of gastric intestinal metaplasia. *Crit. Rev. Oncog.* **12**, 3–26 (2006).
65. Liu, Q. *et al.* CDX2 expression is progressively decreased in human gastric intestinal metaplasia, dysplasia and cancer. *Mod. Pathol.* **20**, 1286–1297 (2007).
66. Farinati, F. *et al.* *Helicobacter pylori*, inflammation, oxidative damage and gastric cancer: A morphological, biological and molecular pathway. *Eur. J. Cancer Prev.* **17**, 195–200 (2008).
67. Verhulst, M. L., Van Oijen, A. H. A. M., Roelofs, H. M. J., Peters, W. H. M. & Jansen, J. B. M. J. Antral glutathione concentration and glutathione S-transferase activity in patients with and without *Helicobacter pylori*. *Dig. Dis. Sci.* **45**, 629–632 (2000).
68. Osaki, M. *et al.* Expression of RUNX3 protein in human gastric mucosa, intestinal metaplasia and carcinoma. *Eur. J. Clin. Invest.* **34**, 605–612 (2004).

69. Shiao, Y. H., Rugge, M., Correa, P., Lehmann, H. P. & Scheer, W. D. P53 alteration in gastric precancerous lesions. *Am. J. Pathol.* **144**, 511–517 (1994).
70. Shiotani, A. *et al.* Evidence that loss of sonic hedgehog is an indicator of *Helicobacter pylori*-induced atrophic gastritis progressing to gastric cancer. *Am. J. Gastroenterol.* **100**, 581–587 (2005).
71. Lee, J. H. *et al.* Frequent CpG island methylation in precursor lesions and early gastric adenocarcinomas. *Oncogene* **23**, 4646–4654 (2004).
72. Hamamoto, T. *et al.* Altered microsatellites in incomplete-type intestinal metaplasia adjacent to primary gastric cancers. *J. Clin. Pathol.* **50**, 841–846 (1997).
73. Jeong, C. W., Lee, J. H., Sohn, S. S., Ryu, S. W. & Kim, D. K. Mitochondrial microsatellite instability in gastric cancer and gastric epithelial dysplasia as a precancerous lesion. *Cancer Epidemiol.* **34**, 323–327 (2010).
74. Kameshima, H. *et al.* *Helicobacter pylori* infection: Augmentation of telomerase activity in cancer and noncancerous tissues. *World J. Surg.* **24**, 1243–1249 (2000).
75. Necchi, V. *et al.* Intracellular, intercellular, and stromal invasion of gastric mucosa, preneoplastic lesions, and cancer by *Helicobacter pylori*. *Gastroenterology* **132**, 1009–1023 (2007).
76. Meireles, S. I. *et al.* Molecular classifiers for gastric cancer and nonmalignant diseases of the gastric mucosa. *Cancer Res.* **64**, 1255–1265 (2004).
77. Zhao, Y. *et al.* A potential role of collagens expression in distinguishing between premalignant and malignant lesions in stomach. *Anat. Rec.* **292**, 692–700 (2009).
78. Gomes, L. I. *et al.* Expression profile of malignant and nonmalignant lesions of esophagus and stomach: Differential activity of functional modules related to inflammation and lipid metabolism. *Cancer Res.* **65**, 7127–7136 (2005).
79. Boussioutas, A. *et al.* Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. *Cancer Res.* **63**, 2569–2577 (2006).
80. Chen, X. *et al.* Variation in gene expression patterns in human gastric cancers. *Mol. Biol. Cell* **14**, 3208–3215 (2003).
81. Lee, H. J. *et al.* Gene expression profiling of metaplastic lineages identifies CDH17 as a prognostic marker in early stage gastric cancer. *Gastroenterology* **139**, 358–366 (2010).

82. Mavrommatis, E., Fish, E. N. & Plataniias, L. C. The Schlafen family of proteins and their regulation by interferons. *J. Interf. Cytokine Res.* **33**, 206–210 (2013).
83. Schwarz, D. A, Katayama, C. D. & Hedrick, S. M. Schlafen, a new family of growth regulatory genes that affect thymocyte development. *Immunity* **9**, 657–668 (1998).
84. Katsoulidis, E. *et al.* Role of Schlafen 2 (SLFN2) in the generation of interferon  $\alpha$ -induced growth inhibitory responses. *J. Biol. Chem.* **284**, 25051–25064 (2009).
85. El-Zaatari, M. *et al.* Gli1 deletion prevents *Helicobacter*-induced gastric metaplasia and expansion of myeloid cell subsets. *PLoS One* **8**, e58935 (2013).
86. Van Zuylen, W. J. *et al.* Macrophage activation and differentiation signals regulate Schlafen-4 gene expression: Evidence for Schlafen-4 as a modulator of myelopoiesis. *PLoS One* **6**, (2011).
87. Gabrilovich, D. I., Ostrand-Rosenberg, S. & Bronte, V. Coordinated regulation of myeloid cells by tumours. *Nat. Rev. Immunol.* **12**, 253–268 (2012).
88. Katsoulidis, E. *et al.* Role of interferon  $\alpha$  (IFN $\alpha$ )-inducible Schlafen-5 in regulation of anchorage-independent growth and invasion of malignant melanoma cells. *J. Biol. Chem.* **285**, 40333–40341 (2010).
89. Geserick, P., Kaiser, F., Klemm, U., Kaufmann, S. H. E. & Zerrahn, J. Modulation of T cell development and activation by novel members of the Schlafen (slfn) gene family harbouring an RNA helicase-like motif. *Int. Immunol.* **16**, 1535–1548 (2004).
90. Lichtenstein, P. *et al.* 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 (2000).
91. Hofker, M. H., Fu, J. & Wijmenga, C. The genome revolution and its role in understanding complex diseases. *B. B. A. Mol. Basis Dis.* **1842**, 1889–1895 (2014).
92. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
93. Smith, A. V. Manipulating HapMap data using HaploView. *Cold Spring Harb. Protoc.* **3**, (2008).
94. Pasche, B. & Yi, N. Candidate gene association studies: Successes and failures. *Curr. Opin. Genet. Dev.* **20**, 257–261 (2010).
95. Gao, L., Nieters, A. & Brenner, H. Meta-analysis: Tumour invasion-related genetic polymorphisms and gastric cancer susceptibility. *Aliment. Pharmacol. Ther.* **28**, 565–573 (2008).

96. Loh, M. *et al.* Meta-analysis of genetic polymorphisms and gastric cancer risk: Variability in associations according to race. *Eur. J. Cancer* **45**, 2562–2568 (2009).
97. González, C. A., Sala, N. & Capellá, G. Genetic susceptibility and gastric cancer risk. *Int. J. Cancer* **100**, 249–260 (2002).
98. Lan, Q. *et al.* Glutathione S-transferase genotypes and stomach cancer in a population-based case-control study in Warsaw, Poland. *Pharmacogenetics* **11**, 655–661 (2001).
99. Figueiredo, C. *et al.* *Helicobacter pylori* and interleukin 1 genotyping: An opportunity to identify high-risk individuals for gastric carcinoma. *J. Natl. Cancer Inst.* **94**, 1680–1687 (2002).
100. Vincenzi, B. *et al.* Interleukin 1 $\beta$ -511T gene (IL1 $\beta$ ) polymorphism is correlated with gastric cancer in the Caucasian population: Results from a meta-analysis. *Oncol. Rep.* **20**, 1213–1220 (2008).
101. Camargo, M. C. *et al.* Interleukin-1 $\beta$  and interleukin-1 receptor antagonist gene polymorphisms and gastric cancer: A meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **15**, 1674–1687 (2006).
102. Persson, C., Canedo, P., Machado, J. C., El-Omar, E. M. & Forman, D. Polymorphisms in inflammatory response genes and their association with gastric cancer: A HuGE systematic review and meta-analyses. *Am. J. Epidemiol.* **173**, 259–270 (2011).
103. El-Omar, E. M. *et al.* Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* **404**, 398–402 (2000).
104. Tahara, T., Shibata, T., Hirata, I., Nakano, H. & Arisawa, T. CD14 promoter-159 polymorphism is associated with reduced risk of intestinal-type gastric cancer in a Japanese population. *Dig. Dis. Sci.* **54**, 1508–1512 (2009).
105. Hold, G. L. *et al.* CD14-159C/T and TLR9-1237T/C polymorphisms are not associated with gastric cancer risk in Caucasian populations. *Eur. J. Cancer Prev.* **18**, 117–119 (2009).
106. Wang, J. J. *et al.* Association between CD14 gene polymorphisms and cancer risk: A meta-analysis. *PLoS One* **9**, (2014).
107. Zhou, W. *et al.* The -159C/T polymorphism in the CD14 gene and cancer risk: A meta-analysis. *Onco. Targets. Ther.* **7**, 5–12 (2013).
108. Liu, J., He, C., Xu, Q., Xing, C. & Yuan, Y. NOD2 polymorphisms associated with cancer risk: A meta-analysis. *PLoS One* **9**, e89340 (2014).

109. Hold, G. L. *et al.* A functional polymorphism of toll-like receptor 4 gene increases risk of gastric carcinoma and its precursors. *Gastroenterology* **132**, 905–912 (2007).
110. Kutikhin, A. G. Impact of Toll-like receptor 4 polymorphisms on risk of cancer. *Hum. Immunol.* **72**, 193–206 (2011).
111. Zhang, K., Zhou, B., Wang, Y., Rao, L. & Zhang, L. The TLR4 gene polymorphisms and susceptibility to cancer: A systematic review and meta-analysis. *Eur. J. Cancer* **49**, 946–954 (2013).
112. Lo, S. S., Chen, J. H., Wu, C. W. & Lui, W. Y. Functional polymorphism of NFKB1 promoter may correlate to the susceptibility of gastric cancer in aged patients. *Surgery* **145**, 280–285 (2009).
113. Wang, G. Y., Lu, C. Q., Zhang, R. M., Hu, X. H. & Luo, Z. W. The E-cadherin gene polymorphism -160C->A and cancer risk: A HuGE review and meta-analysis of 26 case-control studies. *Am. J. Epidemiol.* **167**, 7–14 (2008).
114. Gao, L., Nieters, A. & Brenner, H. Cell proliferation-related genetic polymorphisms and gastric cancer risk: Systematic review and meta-analysis. *Eur. J. Hum. Genet.* **17**, 1658–1667 (2009).
115. Saadat, M. Genetic polymorphisms of glutathione S-transferase T1 (GSTT1) and susceptibility to gastric cancer: A meta-analysis. *Cancer Sci.* **97**, 505–509 (2006).
116. Chen, B., Zhou, Y., Yang, P. & Wu, X. T. ERCC2 Lys751Gln and Asp312Asn polymorphisms and gastric cancer risk: A meta-analysis. *J. Cancer Res. Clin. Oncol.* **137**, 939–946 (2011).
117. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
118. Sakamoto, H. *et al.* Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.* **40**, 730–740 (2008).
119. Abnet, C. C. *et al.* A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat. Genet.* **42**, 764–767 (2010).
120. Wang, L. D. *et al.* Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat. Genet.* **42**, 759–763 (2010).
121. Shi, Y. *et al.* A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat. Genet.* **43**, 1215–1218 (2011).



122. McNicholl, A. G. *et al.* Accuracy of GastroPanel for the diagnosis of atrophic gastritis. *Eur. J. Gastroenterol. Hepatol.* **26**, 941–948 (2014).
123. Peleteiro, B. *et al.* Association between cytokine gene polymorphisms and gastric precancerous lesions: Systematic review and meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **19**, 762–776 (2010).
124. Pereira, C. *et al.* -765G > C COX-2 polymorphism may be a susceptibility marker for gastric adenocarcinoma in patients with atrophy or intestinal metaplasia. *World J. Gastroenterol.* **12**, 5473–5478 (2006).
125. Chiurillo, M. A. *et al.* Combination of *Helicobacter pylori*-iceA2 and proinflammatory interleukin-1 polymorphisms is associated with the severity of histological changes in Venezuelan chronic gastritis patients. *FEMS Immunol. Med. Microbiol.* **59**, 170–176 (2010).
126. Zabaleta, J. *et al.* Association of interleukin-1 $\beta$  gene polymorphisms with precancerous gastric lesions in African Americans and Caucasians. *Am. J. Gastroenterol.* **101**, 163–171 (2006).
127. Togawa, S. *et al.* Interleukin-2 gene polymorphisms associated with increased risk of gastric atrophy from *Helicobacter pylori* infection. *Helicobacter* **10**, 172–178 (2005).
128. Li, Z. W. *et al.* Inflammatory cytokine gene polymorphisms increase the risk of atrophic gastritis and intestinal metaplasia. *World J. Gastroenterol.* **16**, 1788–1794 (2010).
129. Canzian, F. *et al.* Genetic polymorphisms in mediators of inflammation and gastric precancerous lesions. *Eur. J. Cancer Prev.* **17**, 178–183 (2008).
130. Kato, I. *et al.* Polymorphisms in genes related to bacterial lipopolysaccharide/peptidoglycan signaling and gastric precancerous lesions in a population at high risk for gastric cancer. *Dig. Dis. Sci.* **52**, 254–261 (2007).
131. Trejo de la O, A. *et al.* TLR4 single-nucleotide polymorphisms alter mucosal cytokine and chemokine patterns in Mexican patients with *Helicobacter pylori*-associated gastroduodenal diseases. *Clin. Immunol.* **129**, 333–340 (2008).
132. Pabalan, N., Singh, N., Pineda, M. R. & Jarjanazi, H. Meta-analysis of the association between PTPN11 G/A polymorphism at intron 3 with risk of gastric atrophy among East Asians. *J. Gastrointest. Cancer* **45**, 319–324 (2014).
133. Rosenstiel, P. *et al.* Influence of polymorphisms in the NOD1/CARD4 and NOD2/CARD15 genes on the clinical outcome of *Helicobacter pylori* infection. *Cell. Microbiol.* **8**, 1188–1198 (2006).

134. Li, W. Q. *et al.* Association between genetic polymorphisms of DNA base excision repair genes and evolution of precancerous gastric lesions in a Chinese population. *Carcinogenesis* **30**, 500–505 (2009).
135. You, W. C. *et al.* Genetic polymorphisms of CYP2E1, GSTT1, GSTP1, GSTM1, ALDH2, and ODC and the risk of advanced precancerous gastric lesions in a Chinese population. *Cancer Epidemiol. Biomarkers Prev.* **14**, 451–458 (2005).
136. Chen, S. Y. *et al.* Modification effects of GSTM1, GSTT1 and CYP2E1 polymorphisms on associations between raw salted food and incomplete intestinal metaplasia in a high-risk area of stomach cancer. *Int. J. Cancer* **108**, 606–612 (2004).
137. Carvalho, F. *et al.* MUC1 gene polymorphism and gastric cancer an epidemiological study. *Glycoconj. J.* **14**, 107–111 (1997).
138. Silva, F. *et al.* MUC1 polymorphism confers increased risk for intestinal metaplasia in a Colombian population with chronic gastritis. *Eur. J. Hum. Genet.* **11**, 380–384 (2003).
139. Silva, F. *et al.* MUC1 gene polymorphism in the gastric carcinogenesis pathway. *Eur. J. Hum. Genet.* **9**, 548–552 (2001).
140. Marín, F. *et al.* 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. *Carcinogenesis* **33**, 1072–1080 (2012).
141. Hishida, A. *et al.* Associations of a PTPN11 G/A polymorphism at intron 3 with *Helicobacter pylori* seropositivity, gastric atrophy and gastric cancer in Japanese. *BMC Gastroenterol.* **9**, 51 (2009).
142. Hishida, A. *et al.* Toll-like receptor 4 +3725 G/C polymorphism, *Helicobacter pylori* seropositivity, and the risk of gastric atrophy and gastric cancer in Japanese. *Helicobacter* **14**, 47–53 (2009).
143. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): Study populations and data collection. *Public Health Nutr.* **5**, 1113–1124 (2002).
144. Carneiro, F. *et al.* Pathology findings and validation of gastric and esophageal cancer cases in a European cohort (EPIC/EURGAST). *Scand. J. Gastroenterol.* **42**, 618–627 (2007).
145. Sala, N. *et al.* 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**, 2417–2427 (2012).

146. Palli, D. *et al.* CagA+ *Helicobacter pylori* infection and gastric cancer risk in the EPIC-EURGAST study. *Int. J. Cancer* **120**, 859–867 (2007).
147. So, H. C. & Sham, P. C. Multiple testing and power calculations in genetic association studies. *Cold Spring Harb. Protoc.* **2011**, pdb.top95 (2011).
148. González, J. R. *et al.* SNPassoc: An R package to perform whole genome association studies. *Bioinformatics* **23**, 644–645 (2007).
149. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
150. Reumers, J. *et al.* Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNPeffect and PupaSuite databases. *Nucleic Acids Res.* **36**, 825–829 (2008).
151. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
152. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
153. Gibbs, J. R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
154. Myers, A. J. *et al.* A survey of genetic human cortical gene expression. *Nat. Genet.* **39**, 1494–1499 (2007).
155. Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, (2008).
156. Zeller, T. *et al.* Genetics and beyond - the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, (2010).
157. Brown, C. D., Mangravite, L. M. & Engelhardt, B. E. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* **9**, e1003649 (2013).
158. Ausubel, F. M. *et al.* (2003). *Short Protocols in Molecular Biology*. 3<sup>era</sup> edición. New York: John Wiley & Sons.
159. Birnboim, C. H. (2004). *Composition and methods for obtaining nucleic acids from sputum*. <<http://www.google.com/na/patents/WO2003104251A9?cl=en>>
160. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).

161. Fukamachi, H., Ito, K. & Ito, Y. Runx3  $-/-$  gastric epithelial cells differentiate into intestinal type cells. *Biochem. Biophys. Res. Commun.* **321**, 58–64 (2004).
162. Zavros, Y. *et al.* Chronic gastritis in the hypochlorhydric gastrin-deficient mouse progresses to adenocarcinoma. *Oncogene* **24**, 2354–2366 (2005).
163. Lefebvre, O. *et al.* Gastric mucosa abnormalities and tumorigenesis in mice lacking the pS2 trefoil protein. *Science* **274**, 259–262 (1996).
164. Oshima, H. *et al.* Carcinogenesis in mouse stomach by simultaneous activation of the Wnt signaling and prostaglandin E2 pathway. *Gastroenterology* **131**, 1086–1095 (2006).
165. Oshima, H., Oshima, M., Inaba, K. & Taketo, M. M. Hyperplastic gastric tumors induced by activated macrophages in COX-2/mPGES-1 transgenic mice. *EMBO J.* **23**, 1669–1678 (2004).
166. Mutoh, H. *et al.* Cdx1 induced intestinal metaplasia in the transgenic mouse stomach: Comparative study with Cdx2 transgenic mice. *Gut* **53**, 1416–1423 (2004).
167. Mutoh, H. *et al.* Conversion of gastric mucosa to intestinal metaplasia in Cdx2-expressing transgenic mice. *Biochem. Biophys. Res. Commun.* **294**, 470–479 (2002).
168. Companioni, O. *et al.* Polymorphisms of *Helicobacter pylori* signaling pathway genes and gastric cancer risk in the European Prospective Investigation into Cancer-EURGAST cohort. *Int. J. Cancer* **134**, 92–101 (2014).
169. Castaño-Rodríguez, N., Kaakoush, N. O. & Mitchell, H. M. Pattern-recognition receptors and gastric cancer. *Front. Immunol.* **5**, 336 (2014).
170. Maeda, S. *et al.* *H. pylori* activates NF- $\kappa$ B through a signaling pathway involving I $\kappa$ B kinases, NF- $\kappa$ B-inducing kinase, TRAF2, and TRAF6 in gastric cancer cells. *Gastroenterology* **119**, 97–108 (2000).
171. Meyer-ter-Vehn, T., Covacci, A., Kist, M. & Pahl, H. L. *Helicobacter pylori* activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes c-fos and c-jun. *J. Biol. Chem.* **275**, 16064–16072 (2000).
172. Watanabe, T. *et al.* NOD1-mediated mucosal host defense against *Helicobacter pylori*. *Int. J. Inflam.* **2010**, 476482 (2010).
173. Jin, G. *et al.* Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in Han Chinese. *Am. J. Hum. Genet.* **91**, 928–934 (2012).

174. Saeki, N. *et al.* A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology* **140**, 892–902 (2011).
175. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.* **57**, 289–300 (1995).
176. Affymetrix, Inc. *Almac Xcel™ Array for FFPE Profiling Data Sheet*. <[http://www.carrerasresearch.org/almac-xcel-array-for-ffpe-profiling\\_38671.pdf](http://www.carrerasresearch.org/almac-xcel-array-for-ffpe-profiling_38671.pdf)>
177. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
178. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
179. Smyth, G. K., Michaud, J. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067–2075 (2005).
180. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* **1**, 409 (2011).
181. Gentleman, R. C. *et al.* Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
182. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
183. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
184. Sonachalam, M., Shen, J., Huang, H. & Wu, X. Systems biology approach to identify gene network signatures for colorectal cancer. *Front. Genet.* **3**, (2012).
185. The Broad Institute. Gene Set Enrichment Analysis GSEA User Guide. *Broad Inst.* 1–59 (2009).
186. Hanada, K. *et al.* *Helicobacter pylori* infection introduces DNA double-strand breaks in host cells. *Infect. Immun.* **82**, 4182–4189 (2014).
187. Wisnieski, F. Reference genes for quantitative RT-PCR data in gastric tissues and cell lines. *World J. Gastroenterol.* **19**, 7121 (2013).

188. Roche. *ProbeFinder Quick Reference Guide*. <[https://lifescience.roche.com/wcsstore/RASCatalogAssetStore/Articles/ProbeFinder\\_QRG.pdf](https://lifescience.roche.com/wcsstore/RASCatalogAssetStore/Articles/ProbeFinder_QRG.pdf)>
189. Clèries, R. *et al.* BootstRatio: A web-based statistical analysis of fold-change in qPCR and RT-qPCR data using resampling methods. *Comput. Biol. Med.* **42**, 438–445 (2012).
190. Neumann, B., Zhao, L., Murphy, K. & Gonda, T. J. Subcellular localization of the Schlafen protein family. *Biochem. Biophys. Res. Commun.* **370**, 62–66 (2008).
191. McCarty, K. S., Miller, L. S., Cox, E. B. & Konrath, J. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch. Pathol. Lab. Med.* **109**, 716–721 (1985).
192. Cao, L. X. *et al.* Implication of a new molecule IK in CD34+ hematopoietic progenitor cell proliferation and differentiation. *Blood* **89**, 3615–3623 (1997).
193. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
194. Durães, C. *et al.* Genetic variants in the IL1A gene region contribute to intestinal-type gastric carcinoma susceptibility in European populations. *Int. J. Cancer* **135**, 1343–1355 (2014).
195. Wee, A., Teh, M. & Kang, J. Y. Association of *Helicobacter pylori* with HLA-DR antigen expression in gastritis. *J. Clin. Pathol.* **45**, 30–33 (1992).
196. Yang, S. Gene amplifications at chromosome 7 of the human gastric cancer genome. *Int. J. Mol. Med.* **20**, 225–231 (2007).
197. Nguyen, T. V *et al.* Genetic polymorphisms in GSTA1, GSTP1, GSTT1, and GSTM1 and gastric cancer risk in a Vietnamese population. *Oncol. Res.* **18**, 349–355 (2010).
198. Bai, Z. *et al.* Proteomics-based identification of a group of apoptosis-related proteins and biomarkers in gastric cancer. *Int. J. Oncol.* **38**, 375–383 (2011).
199. Liu, J. *et al.* PI3K/Akt-dependent phosphorylation of GSK3 $\beta$  and activation of RhoA regulate Wnt5a-induced gastric cancer cell migration. *Cell. Signal.* **25**, 447–456 (2013).
200. Li, L. *et al.* Critical role of histone demethylase RBP2 in human gastric cancer angiogenesis. *Mol. Cancer* **13**, 81 (2014).
201. Aquino, P. F. *et al.* Are gastric cancer resection margin proteomic profiles more similar to those from controls or tumors? *J. Proteome Res.* **11**, 5836–5842 (2012).

202. Ikuta, K., Seno, H. & Chiba, T. Molecular changes leading to gastric cancer: A suggestion from rare-type gastric tumors with GNAS mutations. *Gastroenterology* **146**, 1417–1418 (2014).
203. Niiya, F. *et al.* Expression of SART3 tumor-rejection antigen in gastric cancers. *Jpn. J. Cancer Res.* **91**, 337–342 (2000).
204. Hu, C. *et al.* Abnormal hypermethylation of promoter region downregulates chemokine CXC ligand 14 expression in gastric cancer. *Int. J. Oncol.* **43**, 1487–1494 (2013).
205. Kalnina, Z. *et al.* Molecular characterisation and expression analysis of SEREX-defined antigen NUCB2 in gastric epithelium, gastritis and gastric cancer. *Eur. J. Histochem.* **53**, 7–18 (2009).
206. Wu, Q. *et al.* Methylation of miR-129-5p CpG island modulates multi-drug resistance in gastric cancer by targeting ABC transporters. *Oncotarget* **5**, 11552–11563 (2014).
207. Mimura, K. *et al.* The MAPK pathway is a predominant regulator of HLA-A expression in esophageal and gastric cancer. *J. Immunol.* **191**, 6261–6272 (2013).
208. Qian, Z. *et al.* Whole genome gene copy number profiling of gastric cancer identifies PAK1 and KRAS gene amplification as therapy targets. *Gene. Chromosome. Canc.* **53**, 883–894 (2014).
209. Li, S., Lu, A. P., Zhang, L. & Li, Y. D. Anti-*Helicobacter pylori* immunoglobulin G (IgG) and IgA antibody responses and the value of clinical presentations in diagnosis of *H. pylori* infection in patients with precancerous lesions. *World J. Gastroenterol.* **9**, 755–758 (2003).
210. Nam, K. H. *et al.* Caveolin 1 expression correlates with poor prognosis and focal adhesion kinase expression in gastric cancer. *Pathobiology* **80**, 87–94 (2013).
211. Watanabe, Y. *et al.* HLA-DQB1 locus and gastric cancer in *Helicobacter pylori* infection. *J. Gastroenterol. Hepatol.* **21**, 420–424 (2006).
212. Wang, Y. C. *et al.* CD24 mediates gastric carcinogenesis and promotes gastric cancer progression via STAT3 activation. *Apoptosis* **19**, 643–656 (2014).
213. Xu, H. Y. *et al.* Transfection of PDCD5 effect on the biological behavior of tumor cells and sensitized gastric cancer cells to cisplatin-induced apoptosis. *Dig. Dis. Sci.* **57**, 1847–1856 (2012).

214. Kim, K. R. *et al.* Gene expression profiling using oligonucleotide microarray in atrophic gastritis and intestinal metaplasia. *Korean J. Gastroenterol.* **49**, 209–224 (2007).
215. Xie, Ying, Guo, Z., Wang, Y. & Zhao, Y. Single-nucleotide polymorphisms of microRNA processing machinery genes are associated with risk for gastric cancer. *Onco. Targets. Ther.* **8**, 567–571 (2015).
216. Wang, J., Cui, S., Zhang, X., Wu, Y. & Tang, H. High expression of heat shock protein 90 is associated with tumor aggressiveness and poor prognosis in patients with advanced gastric cancer. *PLoS One* **8**, e62876 (2013).
217. Humphries, J. M. *et al.* Identification and validation of novel candidate protein biomarkers for the detection of human gastric cancer. *B. B. A. Proteins Proteom.* **1844**, 1051–1058 (2014).
218. Wu, W. *et al.* S100A9, GIF and AAT as potential combinatorial biomarkers in gastric cancer diagnosis and prognosis. *Proteomics. Clin. Appl.* **6**, 152–162 (2012).
219. Magnusson P. K. E. *et al.* Gastric cancer and human leukocyte antigen: Distinct DQ and DR alleles are associated with development of gastric cancer and infection by *Helicobacter pylori*. *Cancer Res.* **61**, 2684–2689 (2001).
220. Lee, J. Y., Eom, E. M., Kim, D. S., Ha-Lee, Y. M. & Lee, D. H. Analysis of gene expression profiles of gastric normal and cancer tissues by SAGE. *Genomics* **82**, 78–85 (2003).
221. Dai, J., Zhang, N., Wang, J., Chen, M. & Chen, J. Gastroke-2 is downregulated in gastric cancer and its restoration suppresses gastric tumorigenesis and cancer metastasis. *Tumour Biol.* **35**, 4199–4207 (2014).
222. Nardone, G. *et al.* Molecular expression of Gastroke 1 in normal mucosa and in *Helicobacter pylori*-related preneoplastic and neoplastic gastric lesions. *Cancer Biol. Ther.* **7**, 1890–1895 (2008).
223. Ishigami, S. *et al.* Cancerous HLA class I expression and regulatory T cell infiltration in gastric cancer. *Cancer Immunol. Immunother.* **61**, 1663–1669 (2012).
224. Ning, P. F., Liu, H. J. & Yuan, Y. Dynamic expression of pepsinogen C in gastric cancer, precancerous lesions and *Helicobacter pylori* associated gastric diseases. *World J. Gastroenterol.* **11**, 2545–2548 (2005).
225. Shi, J., Xu, S., He, P. & Xi, Z. Expression of carcinoembryonic antigen-related cell adhesion molecule 1(CEACAM1) and its correlation with angiogenesis in gastric cancer. *Pathol. Res. Pract.* **210**, 473–476 (2014).



226. Li, X. *et al.* miRNA-223 promotes gastric cancer invasion and metastasis by targeting tumor suppressor EPB41L3. *Mol. Cancer Res.* **9**, 824–833 (2011).
227. Jang, B. G., Lee, B. L. & Kim, W. H. Olfactomedin-related proteins 4 (OLFM4) expression is involved in early gastric carcinogenesis and of prognostic significance in advanced gastric cancer. *Virchows Arch.* **4**, (2015).
228. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One* **18**, e10693 (2010).
229. Corridoni, D., Arseneau, K. O., Cifone, M. G. & Cominelli, F. The dual role of nod-like receptors in mucosal innate immunity and chronic intestinal inflammation. *Front. Immunol.* **5**, (2014).
230. Helicobacter and Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* **49**, 347–353 (2001).
231. Hnatyszyn, A., Szalata, M., Stanczyk, J., Cichy, W. & Slomski, R. Association of c.802C>T polymorphism of NOD2/CARD15 gene with the chronic gastritis and predisposition to cancer in *H. pylori* infected patients. *Exp. Mol. Pathol.* **88**, 388–393 (2010).
232. Wang, P. *et al.* Association of NOD1 and NOD2 genes polymorphisms with *Helicobacter pylori* related gastric cancer in a Chinese population. *World J. Gastroenterol.* **18**, 2112–2120 (2012).
233. Angeletti, S. *et al.* NOD2/CARD15 polymorphisms impair innate immunity and increase susceptibility to gastric cancer in an Italian population. *Hum. Immunol.* **70**, 729–732 (2009).
234. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 189–192 (2003).
235. Rosenstiel, P. *et al.* A short isoform of NOD2/CARD15, NOD2-S, is an endogenous inhibitor of NOD2/receptor-interacting protein kinase 2-induced signaling pathways. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3280–3285 (2006).
236. Nicod, N., Pradas-Juni, M. & Gomis, R. Role of the single nucleotide polymorphism rs7903146 of TCF7L2 in inducing nonsense-mediated decay. *Springerplus* **3**, 41 (2014).
237. Karhukorpi, J. *et al.* Effect of CD14 promotor polymorphism and *H. pylori* infection and its clinical outcomes on circulating CD14. *Clin Exp Immunol* **128**, 326–332 (2002).

238. Smith, M. G., Hold, G. L., Tahara, E. & El-Omar, E. M. Cellular and molecular aspects of gastric cancer. *World J. Gastroenterol.* **12**, 2979–2990 (2006).
239. Li, K. *et al.* CD14 regulates gastric cancer cell epithelial-mesenchymal transition and invasion in vitro. *Oncol. Rep.* **30**, 2725–2732 (2013).
240. Baldini, M. *et al.* A Polymorphism\* in the 5' flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am. J. Respir. Cell Mol. Biol.* **20**, 976–983 (1999).
241. Wu, M. S. *et al.* Functional polymorphisms of CD14 and toll-like receptor 4 in Taiwanese Chinese with *Helicobacter pylori*-related gastric malignancies. *Hepatogastroenterology* **53**, 807–810 (2006).
242. Tahara, T., Arisawa, T., Shibata, T., Hirata, I. & Nakano, H. Association of polymorphism of TLR4 and CD14 genes with gastroduodenal diseases in Japan. *Inflammopharmacology* **15**, 124–128 (2007).
243. De Aguiar, B. B. *et al.* CD14 expression in the first 24h of sepsis: Effect of -260C>T CD14 SNP. *Immunol. Invest.* **37**, 752–769 (2008).
244. Chen, X. *et al.* The identification of trans-associations between prostate cancer GWAS SNPs and RNA expression differences in tumor-adjacent stroma. *Oncotarget* **6**, 1865–1873 (2015).
245. Loo, L. W. M. *et al.* Cis-expression QTL analysis of established colorectal cancer risk variants in colon tumors and adjacent normal tissue. *PLoS One* **7**, (2012).
246. Corvalan, A. *et al.* Epstein-Barr virus in gastric carcinoma is associated with location in the cardia and with a diffuse histology: A study in one area of Chile. *Int. J. Cancer* **94**, 527–530 (2001).
247. Yoshiyama, H., Imai, S., Shimizu, N. & Takada, K. Epstein-Barr virus infection of human gastric carcinoma cells: Implication of the existence of a new virus receptor different from CD21. *J. Virol.* **71**, 5688–5691 (1997).
248. Pugin, J. *et al.* CD14 is a pattern recognition receptor. *Immunity* **1**, 509–516 (1994).
249. Garza-González, E. *et al.* Assessment of the toll-like receptor 4 Asp299Gly, Thr399Ile and interleukin-8 -251 polymorphisms in the risk for the development of distal gastric cancer. *BMC Cancer* **7**, 70 (2007).
250. Santini, D. *et al.* Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms in gastric cancer of intestinal and diffuse histotypes. *Clin. Exp. Immunol.* **154**, 360–364 (2008).

251. Smith, M. F. *et al.* Toll-like receptor (TLR) 2 and TLR5, but not TLR4, are required for *Helicobacter pylori*-induced NF- $\kappa$ B activation and chemokine expression by epithelial cells. *J. Biol. Chem.* **278**, 32552–32560 (2003).
252. Bäckhed, F. *et al.* Gastric mucosal recognition of *Helicobacter pylori* is independent of Toll-like receptor 4. *J. Infect. Dis.* **187**, 829–836 (2003).
253. Sun, X. F. & Zhang, H. NFKB and NFKBI polymorphisms in relation to susceptibility of tumour and other diseases. *Histology and Histopathology* **22**, 1387–1398 (2007).
254. Kufe, D. W. Mucins in cancer: Function, prognosis and therapy. *Nat. Rev. Cancer* **9**, 874–885 (2009).
255. Conze, T. *et al.* MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. *Glycobiology* **20**, 199–206 (2010).
256. Boltin, D. & Niv, Y. Mucins in gastric cancer - An update. *J. Gastrointest. Dig. Syst.* **3**, 15519 (2013).
257. Moghanibashi, M., Mohamadynejad, P., Rasekhi, M., Ghaderi, A. & Mohammadianpanah, M. Polymorphism of estrogen response element in TFF1 gene promoter is associated with an increased susceptibility to gastric cancer. *Gene* **492**, 100–103 (2012).
258. Kim, E. J. *et al.* *Helicobacter pylori* infection enhances gastric mucosal inflammation in individuals carrying the 260-T allele of the CD14 gene. *Gut Liver* **7**, 317–322 (2013).
259. Renouf, B. *et al.* Cdx2 homeoprotein inhibits non-homologous end joining in colon cancer but not in leukemia cells. *Nucleic Acids Res.* **40**, 3456–3469 (2012).
260. Zang, S. *et al.* Identification of Differentially-expressed genes in intestinal gastric cancer by microarray analysis. *Genomics Proteomics Bioinformatics* **12**, 276–283 (2014).
261. Hensel, F. Ten-year follow-up of a prospective trial for the targeted therapy of gastric cancer with the human monoclonal antibody PAT-SC1. *Oncol. Rep.* **31**, 1059–1066 (2014).
262. Sasaki, T. *et al.* AKT activation and telomerase reverse transcriptase expression are concurrently associated with prognosis of gastric cancer. *Pathobiology* **81**, 36–41 (2013).

263. Cook, M. B. *et al.* Iron in relation to gastric cancer in the  $\alpha$ -tocopherol,  $\beta$ -carotene Cancer Prevention Study. *Cancer Epidemiol. Biomarkers Prev.* **21**, 2033–2042 (2012).
264. Han, Y. *et al.* Identification and validation that up-expression of HOXA13 is a novel independent prognostic marker of a worse outcome in gastric cancer based on immunohistochemistry. *Med. Oncol.* **30**, 564 (2013).
265. Liu, R. *et al.* Depletion of OLFM4 gene inhibits cell growth and increases sensitization to hydrogen peroxide and tumor necrosis factor-alpha induced-apoptosis in gastric cancer cells. *J. Biomed. Sci.* **19**, 38 (2012).
266. Vernygorodskyi, S. Immunohistochemical evaluation of mucin expression in precancerous tissue of stomach. *Exp. Oncol.* **35**, 114–117 (2013).
267. De Bolos, C., Real, F. X. & Lopez-Ferrer, A. Regulation of mucin and glycoconjugate expression: From normal epithelium to gastric tumors. *Front. Biosci.* **6**, D1256–D1263 (2001).
268. Wong, W. M., Poulosom, R. & Wright, N. A. Trefoil peptides. *Gut* **44**, 890–895 (1999).
269. Liu, T. H., Li, D. C., Gu, C. F. & Ye, S. F. Carbamyl phosphate synthetase I. A novel marker for gastric carcinoma. *Chin. Med. J. (Engl)*. **102**, 630–638 (1989).
270. Deng, Q. *et al.* Prognostic value of long non-coding RNA HOTAIR in various cancers. *PLoS One* **10**, e110059 (2014).
271. Sousa, J. F. *et al.* Proteomic profiling of paraffin-embedded samples identifies metaplasia-specific and early-stage gastric cancer biomarkers. *Am. J. Pathol.* **181**, 1560–1572 (2012).
272. Chandrasekaran, E. V *et al.* Potential tumor markers for human gastric cancer: An elevation of glycan:sulfotransferases and a concomitant loss of  $\alpha$ 1,2-fucosyltransferase activities. *J. Cancer Res. Clin. Oncol.* **133**, 599–611 (2007).
273. Pradeep, D. U. *et al.* TRIM protein mediated regulation of inflammatory and innate immune signaling and its association with antiretroviral activity. *J. Virol.* **87**, 257–272 (2013).
274. Falaleeva, M., Surface, J., Shen, M., de la Grange, P., Stamm, S. SNORD116 and SNORD115 change expression of multiple genes and modify each other's activity. *Gene.* **572**, 266-73 (2015)

275. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
276. Lerardi, E. *et al.* Effect of *Helicobacter pylori* eradication on gastric epithelial proliferation. Relationship with ras oncogene p21 expression. *Ital. J. Gastroenterol. Hepatol.* **29**, 214–219 (1997).
277. Cito, L., Pentimalli, F., Forte, I., Mattioli, E. & Giordano, A. Rb family proteins in gastric cancer (review). *Oncol. Rep.* **24**, 1411–1418 (2010).
278. Byun, E., Lim, J. W., Kim, J. M. & Kim, H.  $\alpha$ -lipoic acid inhibits *Helicobacter pylori* induced oncogene expression and hyperproliferation by suppressing the activation of NADPH oxidase in gastric epithelial cells. *Mediators Inflamm.* **2014**, 1–12 (2014).
279. De Souza, C. R. T. *et al.* MYC deregulation in gastric cancer and its clinicopathological implications. *PLoS One* **8**, e64420 (2013).
280. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on fat: The role of lipid synthesis in cancer metabolism and tumour development. *Dis. Model. Mech.* **6**, 1353–1363 (2013).
281. Bizama, C. *et al.* The low-abundance transcriptome reveals novel biomarkers, specific intracellular pathways and targetable genes associated with advanced gastric cancer. *Int. J. Cancer* **134**, 755–764 (2014).
282. Abdel-Rahman, O. Targeting vascular endothelial growth factor (VEGF) pathway in gastric cancer: Preclinical and clinical aspects. *Crit. Rev. Oncol. Hematol.* **93**, 18–27 (2015).
283. Tsang, Y. H. *et al.* *Helicobacter pylori* CagA targets gastric tumor suppressor RUNX3 for proteasome-mediated degradation. *Oncogene* **29**, 5643–5650 (2010).
284. Eguchi, H., Herschenhous, N., Kuzushita, N. & Moss, S. F. *Helicobacter pylori* increases proteasome-mediated degradation of p27kip1 in gastric epithelial cells. *Cancer Res.* **63**, 4739–4746 (2003).
285. Valenzuela, M. *et al.* *Helicobacter pylori*-induced loss of survivin and gastric cell viability is attributable to secreted bacterial gamma-glutamyl transpeptidase activity. *J. Infect. Dis.* **208**, 1131–1141 (2013).
286. Necchi, V., Sommi, P., Ricci, V. & Solcia, E. In vivo accumulation of *Helicobacter pylori* products, NOD1, ubiquitinated proteins and proteasome in a novel cytoplasmic structure. *PLoS One* **5**, (2010).

287. Shiotani, A. *et al.* Re-expression of sonic hedgehog and reduction of CDX2 after *Helicobacter pylori* eradication prior to incomplete intestinal metaplasia. *Int. J. Cancer* **121**, 1182–1189 (2007).
288. Camilo, V. *et al.* Differentiation reprogramming in gastric intestinal metaplasia and dysplasia: Role of SOX2 and CDX2. *Histopathology* **66**, 343–350 (2015).
289. Chung, I. K. *et al.* *Helicobacter pylori* and telomerase activity in intestinal metaplasia of the stomach. *Korean J. Intern. Med.* **17**, 227–233 (2002).
290. Baik, S., Youn, H. & Chung, M. Increased oxidative DNA damage in *Helicobacter pylori*-infected human gastric mucosa. *Cancer Res.* **56**, 1279–1282 (1996).
291. Yuan, X. *et al.* Activation of TLR4 signaling promotes gastric cancer progression by inducing mitochondrial ROS production. *Cell Death Dis.* **4**, e794 (2013).
292. Zhang, S. *et al.* Homeostasis of redox status derived from glucose metabolic pathway could be the key to understanding the Warburg effect. *Am. J. Cancer Res.* **5**, 928–944 (2015).
293. Barrett, M. T. *et al.* Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat. Genet.* **22**, 106–109 (1999).
294. Tamási, V., Monostory, K., Prough, R. A. & Falus, A. Role of xenobiotic metabolism in cancer: Involvement of transcriptional and miRNA regulation of P450s. *Cell. Mol. Life Sci.* **68**, 1131–1146 (2011).
295. Wallace, B. D. & Redinbo, M. R. Xenobiotic-sensing nuclear receptors involved in drug metabolism: A structural perspective. *Drug Metab. Rev.* **45**, 79–100 (2013).
296. Tahara, T. *et al.* COMT gene val158met polymorphism influences the severity of intestinal metaplasia in *Helicobacter pylori* infected older subjects. *Hepatogastroenterology.* **56**, 411–415 (2009).
297. Joncquel Chevalier Curt, M. *et al.* Alteration or adaptation, the two roads for human gastric mucin glycosylation infected by *Helicobacter pylori*. *Glycobiology* **25**, 617–631 (2015).
298. Karasawa, F. *et al.* Essential role of gastric gland mucin in preventing gastric cancer in mice. *J Clin Invest*, **122**, 923–34 (2012).
299. Gomes, C. *et al.* Glycoproteomic analysis of serum from patients with gastric precancerous lesions. *J. Proteome Res.* **12**, 1454–1466 (2013).

300. Altschuler, Y. *et al.* Clathrin-mediated endocytosis of MUC1 is modulated by its glycosylation state. *Mol. Biol. Cell* **11**, 819–831 (2000).
301. Pinho, SS., Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*, **15**, 540-55 (2015).
302. Kaebisch, R., Mejías-Luque, R., Prinz, C. & Gerhard, M. *Helicobacter pylori* cytotoxin-associated gene A impairs human dendritic cell maturation and function through IL-10-mediated activation of STAT3. *J. Immunol.* **192**, 316–323 (2014).
303. Borlace, G. N. *et al.* A role for altered phagosome maturation in the long-term persistence of *Helicobacter pylori* infection. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303**, G169–G179 (2012).
304. Henson, D. E. & Albores-Saavedra, J. An hypothesis on the presumed association between thyroid goiter and gastric cancer. *Cancer Causes Control* **24**, 609–610 (2013).
305. Liu, R. *et al.* Mechanism of cancer cell adaptation to metabolic stress: Proteomics identification of a novel thyroid hormone-mediated gastric carcinogenic signaling pathway. *Mol. Cell. Proteomics* **8**, 70–85 (2009).
306. Bubenik, G. A. Thirty four years since the discovery of gastrointestinal melatonin. *J. Physiol. Pharmacol.* **59**, 33–51 (2008).
307. Goto, T. *et al.* Enhanced expression of inducible nitric oxide synthase and nitrotyrosine in gastric mucosa of gastric cancer patients. *Clin. Cancer Res.* **5**, 1411–1415 (1999).
308. Hong, L. *et al.* The value of MG7-Ag and COX-2 for predicting malignancy in gastric precancerous lesions. *Cell Biol. Int.* **34**, 873–876 (2010).
309. Sun, Y., Tian, M. M., Zhou, L. X., You, W. C. & Li, J. Y. Value of c-Met for predicting progression of precancerous gastric lesions in rural Chinese population. *Chin. J. Cancer Res.* **24**, 18–22 (2012).
310. Lindholm, C., Quiding-Järbrink, M., Lönroth, H., Hamlet, A. & Svennerholm, A. M. Local cytokine response in *Helicobacter pylori*-infected subjects. *Infect. Immun.* **66**, 5964–5971 (1998).
311. Sassano, A. *et al.* Human Schlafen 5 (SLFN5) is a regulator of motility and invasiveness of renal cell carcinoma cells. *Mol. Cell. Biol.* **35**, 2684–2698 (2015).
312. Tian, L. *et al.* Schlafen-11 sensitizes colorectal carcinoma cells to irinotecan. *Anticancer. Drugs* **38**, 1–7 (2014).

313. Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoediting and its three component phases-elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27**, 16–25 (2014).
314. Shafaghi, A. *et al.* Serum gastrin and the pepsinogen I/II ratio as markers for diagnosis of premalignant gastric lesions. *Asian Pacific J. Cancer Prev.* **14**, 3931–3936 (2013).
315. Ng, E. K. O. *et al.* Quantitative analysis and diagnostic significance of methylated SLC19A3 DNA in the plasma of breast and gastric cancer patients. *PLoS One* **6**, (2011).
316. Joo, M., Kim, H., Kim, M. K., Yu, H. J. & Kim, J. P. Expression of Ep-CAM in intestinal metaplasia, gastric epithelial dysplasia and gastric adenocarcinoma. *J. Gastroenterol. Hepatol.* **20**, 1039–1045 (2005).
317. Fristedt, R. *et al.* Expression and prognostic significance of the polymeric immunoglobulin receptor in esophageal and gastric adenocarcinoma. *J. Transl. Med.* **12**, 83 (2014).
318. Frycz, B. A. *et al.* Expression of 17 $\beta$ -hydroxysteroid dehydrogenase type 2 is associated with some clinicopathological features in gastric cancer. *Biomed. Pharmacother.* **70**, 24–27 (2015).
319. Junnila, S., Kokkola, A., Lindsberg, M. L. K., Puolakkainen, P. & Monni, O. Genome-wide gene copy number and expression analysis of primary gastric tumors and gastric cancer cell lines. *BMC Cancer* **10**, 73 (2010).
320. Muta, H., Podack ER. CD30: from basic research to cancer therapy. *Immunol Res.* **57**,151-8 (2013).





## TABLAS ANEXAS

Tabla Anexa 1. Resultados globales del genotipado de SNPs en el estudio de asociación en el seguimiento de LPGs.											
Gen	SNP <sup>a</sup>	Cromosoma <sup>b</sup>	Posición cromosomal <sup>c</sup>	Alelos <sup>d</sup>	MAF <sup>e</sup>	EHW <sup>f</sup>	Progresión Controles <sup>g</sup>	Progresión Casos <sup>h</sup>	Regresión Controles <sup>i</sup>	Regresión Casos <sup>j</sup>	Bloque de haplotipos <sup>k</sup>
CDX1	rs6894617	5	149521614	C/T	0,259	0,397	249/187/27	56/29/10	199/128/27	106/88/10	Recomb
CDX1	rs2302275	5	149527021	C/G	0,47	0,852	128/232/100	27/43/24	102/168/82	53/107/42	1
CDX1	rs10040224	5	149527350	C/G	0,062	1,000	406/56/1	81/15/0	309/45/1	178/26/0	1
CDX1	rs887343	5	149529016	G/A	0,366	0,550	182/224/57	41/42/13	147/163/45	76/103/25	1
CDX1	rs2282812	5	149533940	G/A	0,366	1,000	184/215/61	41/39/15	142/161/50	83/93/26	2
CDX1	rs10063514	5	149546502	C/T	0,068	0,712	400/61/1	80/15/0	306/47/1	174/29/0	Recomb
CDX2	rs2481952	13	27434901	C/T	0,498	0,309	121/219/123	27/48/20	92/173/89	56/94/54	1
CDX2	rs2504212	13	27435537	C/A	0,211	0,008	297/135/30	64/30/2	231/104/19	130/61/13	1
CDX2	rs4503658	13	27442919	G/T	0,08	0,006	397/59/7	84/12/0	312/41/2	169/30/5	1
CDX2	rs3812863	13	27443268	A/G	0,398	0,441	174/213/76	40/41/15	141/164/50	73/90/41	Recomb
GAST	rs12453761	17	37120839	A/G	0,48	0,405	129/221/113	32/45/19	109/164/82	52/102/50	No es tagSNP
IL1A	rs17561	2	113253684	G/T	0,289	0,178	239/179/45	53/39/4	178/146/31	114/72/18	No es tagSNP
IL1B	rs3917366	2	113300351	G/T	0,246	0,381	258/180/24	60/33/3	203/132/19	115/81/8	1
IL1B	rs1143633	2	113306938	G/A	0,382	0,168	167/232/61	30/50/15	120/177/55	77/105/21	1
IL1B	rs1143627	2	113310858	T/C	0,338	0,409	199/215/48	43/40/12	165/152/36	77/103/24	2
PTGES	rs12001450	9	131536460	T/C	0,192	0,073	296/155/11	59/35/2	226/119/9	129/71/4	1
PTGES	rs4636306	9	131536838	G/A	0,209	0,567	275/150/18	54/33/4	206/113/15	123/70/7	1
PTGES	rs2302821	9	131541702	A/C	0,073	0,294	396/60/4	80/15/1	310/41/3	166/34/2	Recomb
PTGES	rs10760634	9	131543692	A/G	0,078	1,000	392/69/2	79/14/3	290/61/4	181/22/1	2
PTGES	rs4837404	9	131545480	A/G	0,369	1,000	183/215/63	39/45/12	138/166/50	84/94/25	2
PTGES	rs10739757	9	131547731	T/C	0,129	1,000	349/107/7	71/21/3	260/87/7	160/41/3	2
PTGES	rs7872802	9	131555551	A/G	0,134	0,841	346/107/9	72/22/2	257/90/8	161/39/3	3
PTGS2	rs2206593	1	184909052	G/A	0,052	1,000	415/47/1	82/12/1	317/35/2	180/24/0	Recomb

## Tablas Anexas

Gen	SNP <sup>a</sup>	Cromosoma <sup>b</sup>	Posición cromosomal <sup>c</sup>	Alelos <sup>d</sup>	MAF <sup>e</sup>	EHW <sup>f</sup>	Progresión Controles <sup>g</sup>	Progresión Casos <sup>h</sup>	Regresión Controles <sup>i</sup>	Regresión Casos <sup>j</sup>	Bloque de haplotipos <sup>k</sup>
PTGS2	rs5275	1	184909681	T/C	0,319	0,454	216/194/50	58/32/6	175/138/41	99/88/15	1
PTGS2	rs4648276	1	184912111	T/C	0,172	1,000	315/131/14	77/19/0	252/92/8	140/58/6	1
PTGS2	rs5277	1	184914820	G/C	0,192	0,373	306/138/19	58/32/6	225/109/21	139/61/4	1
PTGS2	rs2745557	1	184915844	G/A	0,204	0,475	294/146/22	59/33/4	228/108/18	125/71/8	1
PTGS2	rs689466	1	184917374	A/G	0,175	0,201	313/140/10	61/30/3	245/100/8	129/70/5	1
PTGS2	rs12042763	1	184918499	G/T	0,273	0,295	249/175/38	45/41/10	184/138/32	110/78/16	1
RUNX3	rs6663310	1	25094966	T/C	0,334	0,145	215/191/56	35/51/10	152/159/43	98/83/23	1
RUNX3	rs9438858	1	25096492	C/A	0,391	0,098	182/202/78	31/52/13	131/163/60	82/91/31	2
RUNX3	rs7553295	1	25096597	G/T	0,238	0,799	266/170/25	56/32/8	203/129/22	119/73/11	2
RUNX3	rs9438859	1	25096908	T/C	0,232	0,297	277/157/27	52/37/7	201/133/20	128/61/14	Recomb
RUNX3	rs7514332	1	25098262	G/C	0,054	0,384	415/46/2	86/9/0	317/36/1	184/19/1	3
RUNX3	rs2236850	1	25112928	T/C	0,408	0,503	168/215/79	28/49/19	118/168/68	78/96/30	Recomb
RUNX3	rs2282718	1	25113643	G/A	0,335	1,000	207/202/53	34/47/15	148/159/47	93/90/21	Recomb
RUNX3	rs9438876	1	25113703	A/G	0,489	0,116	131/211/120	20/48/28	91/164/100	60/95/48	Recomb
RUNX3	rs2236852	1	25116354	A/G	0,486	0,579	124/224/114	30/49/17	104/172/79	50/101/52	Recomb
RUNX3	rs7517302	1	25126904	T/C	0,362	0,616	192/204/63	35/48/13	140/164/48	87/88/28	4
TFF1	rs225352	21	42650501	C/T	0,086	1,000	384/73/3	87/9/0	304/50/0	167/32/3	Recomb
TFF1	rs178740	21	42650741	T/C	0,198	0,012	305/126/26	71/23/2	246/91/15	130/58/13	Recomb
TFF1	rs1547374	21	42651964	A/G	0,312	0,516	221/190/47	43/40/13	165/147/38	99/83/22	1
TFF1	rs4919984	21	42653915	A/G	0,222	1,000	281/159/23	54/37/5	213/125/17	122/71/11	1
TFF1	rs225355	21	42657551	T/C	0,492	0,114	127/215/116	25/41/27	93/171/86	59/85/57	1
TFF1	rs225357	21	42659382	C/T	0,431	0,638	144/234/84	38/42/16	121/179/54	61/97/46	Recomb
TFF1	rs4920094	21	42659402	G/A	0,428	0,158	146/240/77	26/47/23	97/194/64	75/93/36	Recomb
TFF1	rs225358	21	42659581	T/C	0,303	0,319	227/181/47	44/43/8	165/145/39	106/79/16	2
TFF1	rs2156310	21	42659675	G/A	0,151	1,000	332/120/10	70/23/3	247/100/8	155/43/5	2
TFF1	rs225359	21	42660505	A/G	0,327	0,833	212/199/51	39/47/9	153/157/43	98/89/17	2
TFF1	rs1788413	21	42661759	A/C	0,162	0,864	322/128/11	72/23/1	253/94/7	141/57/5	3

## Tablas Anexas

Gen	SNP <sup>a</sup>	Cromosoma <sup>b</sup>	Posición cromosomal <sup>c</sup>	Alelos <sup>d</sup>	MAF <sup>e</sup>	EHW <sup>f</sup>	Progresión Controles <sup>g</sup>	Progresión Casos <sup>h</sup>	Regresión Controles <sup>i</sup>	Regresión Casos <sup>j</sup>	Bloque de haplotipos <sup>k</sup>
TFE1	rs9976977	21	42662543	G/A	0,369	0,112	176/231/54	29/47/20	122/179/52	83/99/22	3
TFE1	rs424694	21	42662841	C/T	0,413	0,445	162/217/84	44/41/11	134/171/50	72/87/45	3
TFE1	rs13047838	21	42665938	C/T	0,257	0,183	250/189/24	49/37/10	178/152/25	121/74/9	3
TFE2	rs1079380	21	42643342	A/G	0,489	0,309	113/243/107	32/49/15	99/189/67	46/103/55	No es tagSNP
WNT1	rs4760663	12	47654900	C/T	0,429	0,571	154/221/87	32/49/15	111/171/72	75/99/30	No es tagSNP
MAP3K14	rs4247364	17	40692470	C/G	0,349	0,476	191/216/54	33/51/12	133/174/46	91/93/20	1
MAP3K14	rs1047841	17	40696414	G/A	0,055	1,000	414/48/1	88/8/0	321/34/0	181/22/1	1
MAP3K14	rs16939926	17	40699032	T/C	0,055	1,000	413/49/1	80/16/0	305/49/1	188/16/0	2
MAP3K14	rs2074293	17	40699856	T/C	0,434	0,005	133/253/73	32/44/19	104/185/62	61/112/30	2
MAP3K14	rs3785803	17	40716996	C/T	0,114	0,356	366/88/8	79/16/1	279/72/4	166/32/5	2
MAP3K14	rs2074292	17	40717274	C/T	0,495	0,009	105/257/99	25/49/20	87/196/69	43/110/50	2
MAP3K14	rs9908076	17	40731893	C/T	0,067	0,711	402/60/1	77/18/0	299/54/1	180/24/0	3
MAP3K14	rs2867316	17	40732230	C/T	0,362	0,035	177/234/50	37/45/14	138/174/41	76/105/23	3
MAPK3	rs7698	16	30033301	C/T	0,13	1,000	353/103/7	74/21/0	268/83/3	159/41/4	No hace
MAPK3	rs11865086	16	30037994	A/C	0,491	0,516	123/222/116	16/57/23	93/179/82	46/100/57	No hace
NFKB1	rs980455	4	103637989	A/G	0,391	0,173	178/206/77	32/44/19	144/153/55	66/97/41	1
NFKB1	rs13117745	4	103697738	C/T	0,122	0,085	361/91/11	70/25/0	274/74/6	157/42/5	1
NFKB1	rs1598861	4	103705733	A/C	0,365	0,003	201/184/77	46/42/8	153/137/64	94/89/21	1
NFKB1	rs1610152	4	103725412	G/C	0,046	0,614	420/42/0	89/7/0	325/29/0	184/20/0	1
NFKB1	rs11722146	4	103743667	G/A	0,268	0,906	248/179/34	45/40/10	194/133/26	99/86/18	1
NFKB1	rs4648090	4	103746106	G/A	0,152	0,858	333/117/11	63/30/3	260/85/9	136/62/5	1
NFKB1	rs4648127	4	103754951	C/T	0,033	0,402	433/29/1	89/7/0	330/24/1	192/12/0	Recomb
NFKB1	rs230547	4	103755307	C/T	0,105	0,623	370/86/6	79/15/2	280/67/7	169/34/1	Recomb
NFKB1	rs4648135	4	103755716	A/G	0,051	0,624	416/47/0	89/7/0	324/31/0	181/23/0	Recomb
NFKB1	rs7674640	4	103759828	T/C	0,485	0,228	115/217/130	30/51/14	87/164/103	58/104/41	2
NFKB1A	rs3138056	14	34938265	G/A	0,319	0,670	212/204/45	45/38/12	167/144/42	90/98/15	Recomb
NFKB1A	rs4982269	14	34939618	T/C	0,338	0,536	199/213/50	44/39/13	149/160/45	94/92/18	1

Gen	SNP <sup>a</sup>	Cromosoma <sup>b</sup>	Posición cromosomal <sup>c</sup>	Alelos <sup>d</sup>	MAF <sup>e</sup>	EHW <sup>f</sup>	Progresión Controles <sup>g</sup>	Progresión Casos <sup>h</sup>	Regresión Controles <sup>i</sup>	Regresión Casos <sup>j</sup>	Bloque de haplotipos <sup>k</sup>
NFKBIA	rs696	14	34940844	G/A	0,372	0,321	177/227/59	32/48/14	133/178/42	76/97/31	1
NFKBIA	rs3138054	14	34942058	G/A	0,154	0,859	329/122/10	72/22/0	245/100/7	156/144/3	2
NFKBIA	rs3138053	14	34944605	A/G	0,284	0,820	235/191/36	54/36/5	178/149/27	111/78/14	2
NOD1	rs11536450	7	30434089	C/G	0,13	1,000	351/103/8	69/27/0	275/78/2	145/52/6	1
NOD1	rs2907749	7	30452266	A/G	0,295	0,074	221/208/32	39/48/9	154/173/26	106/83/15	2
NOD1	rs7789045	7	30460547	T/A	0,463	0,641	131/234/96	21/52/22	88/184/80	64/102/38	3
NOD1	rs2906766	7	30466100	T/C	0,302	0,661	225/198/39	42/46/8	168/162/24	99/82/23	4
NOD1	rs3823773	7	30466486	A/G	0,087	0,768	386/72/4	81/15/0	296/56/2	171/31/2	4
NOD1	rs4272257	7	30470374	C/T	0,157	1,000	326/125/11	63/29/4	239/107/8	150/47/7	4
NOD1	rs2709803	7	30477059	C/T	0,217	1,000	285/155/21	60/31/15	218/120/15	127/66/11	4
NOD1	rs4720004	7	30478744	T/C	0,144	0,706	336/119/8	64/28/3	245/103/6	155/44/5	5
NOD1	rs2256023	7	30481516	T/C	0,426	0,924	152/225/83	26/51/19	106/180/66	72/96/36	6
NOD1	rs17770244	7	30482337	A/G	0,061	1,000	405/55/1	82/14/0	310/44/0	177/25/1	6
PTPN11	rs17822304	12	111373075	C/A	0,049	1,000	416/45/1	87/9/0	315/39/0	188/15/1	1
PTPN11	rs11614544	12	111389338	A/G	0,154	0,285	333/115/13	69/24/2	263/82/8	139/57/7	1
PTPN11	rs11066320	12	111390798	G/A	0,437	0,778	145/224/90	31/45/19	115/171/66	61/98/43	1
PTPN11	rs11066322	12	111406912	G/A	0,189	0,128	299/152/10	63/31/2	233/111/10	129/72/2	1
PTPN11	rs11066323	12	111407744	G/A	0,103	0,613	373/84/5	76/18/2	285/65/4	164/37/3	1
PTPN11	rs7958372	12	111419936	C/T	0,085	0,561	385/74/2	83/13/0	299/53/2	169/34/0	1
PTPN11	rs7953150	12	111425954	G/A	0,286	0,821	236/187/38	50/41/5	175/151/28	111/77/15	Recomb
SRC	rs6017901	20	35401507	T/C	0,043	0,583	423/38/1	84/12/0	314/39/1	193/11/0	Recomb
SRC	rs6094373	20	35402488	T/C	0,221	0,418	284/153/26	59/26/11	218/113/24	125/66/13	Recomb
SRC	rs6017916	20	35404692	A/C	0,251	0,712	260/171/32	51/32/12	192/133/29	119/70/15	1
SRC	rs12106024	20	35412762	C/G	0,203	0,115	298/140/25	61/28/7	226/108/21	133/60/11	Recomb
SRC	rs1547836	20	35414365	G/C	0,188	0,878	298/145/15	63/27/15	225/113/15	136/59/5	Recomb
SRC	rs16986606	20	35417842	A/G	0,146	0,263	340/109/13	67/26/3	261/83/11	146/52/5	2
SRC	rs6017996	20	35420067	G/A	0,144	0,851	337/115/10	65/31/0	254/94/6	148/52/4	3

## Tablas Anexas

Gen	SNP <sup>a</sup>	Cromosoma <sup>b</sup>	Posición cromosomal <sup>c</sup>	Alelos <sup>d</sup>	MAF <sup>e</sup>	EHW <sup>f</sup>	Progresión Controles <sup>g</sup>	Progresión Casos <sup>h</sup>	Regresión Controles <sup>i</sup>	Regresión Casos <sup>j</sup>	Bloque de haplotipos <sup>k</sup>
SRC	rs6018027	20	35424206	T/C	0,275	0,488	239/189/33	51/41/4	182/151/20	108/79/17	3
SRC	rs6063022	20	35426741	C/T	0,16	1,000	325/126/12	68/25/3	245/100/10	148/51/5	3
SRC	rs6018088	20	35432039	C/A	0,136	0,692	340/111/7	68/24/3	258/88/6	150/47/4	3
SRC	rs6018148	20	35439194	C/T	0,204	0,670	295/147/21	55/37/4	222/119/14	128/65/11	5
SRC	rs7269342	20	35441575	T/C	0,146	0,710	337/114/11	67/24/5	249/96/9	155/42/7	5
SRC	rs3790150	20	35448513	A/G	0,163	0,395	326/122/15	66/27/2	246/100/8	146/49/9	5
SRC	rs754626	20	35450754	A/C	0,223	0,503	274/169/20	65/28/3	217/125/13	122/72/10	Recomb
SRC	rs754625	20	35450938	T/C	0,288	0,821	234/189/40	52/38/6	185/147/23	101/80/23	Recomb
SRC	rs6018257	20	35455953	T/C	0,122	0,385	357/97/9	69/23/4	263/82/10	163/38/3	6
SRC	rs1570209	20	35462245	T/C	0,102	1,000	372/87/4	73/21/1	279/72/3	166/36/2	6
SRC	rs17785475	20	35466307	C/T	0,067	0,711	401/61/1	83/11/1	302/51/1	182/21/1	Recomb
LRFN2	rs2494938	6	40644106	A/G	0,462	0,263	141/215/106	30/51/15	105/172/77	66/94/44	No es tagSNP
MUC1	rs2070803	1	153424339	T/C	0,43	0,089	158/208/95	27/46/23	106/168/80	79/86/38	No es tagSNP
PSCA	rs2294008	8	143758933	C/T	0,437	0,349	151/219/93	26/48/22	114/166/75	63/101/40	No es tagSNP
PRKAA1	rs13361707	5	40827641	C/T	0,268	0,156	240/193/29	44/42/10	179/146/29	105/89/10	No es tagSNP
DNAH11	rs2285947	7	21550613	G/A	0,488	0,194	128/217/116	20/56/20	84/183/86	64/90/50	No es tagSNP
ZBTB20	rs9841504	3	115845454	C/G	0,059	0,209	410/48/3	85/10/0	317/34/1	178/24/2	No es tagSNP
CD14	rs11167532	5	139968081	G/A	0,45	0,075	118/211/104	42/42/11	113/163/78	77/90/37	1
CD14	rs778588	5	139987195	T/C	0,294	0,181	227/202/34	46/39/11	187/138/30	86/103/15	1
CD14	rs2569190	5	139993100	A/G	0,494	0,459	113/224/123	34/47/15	87/169/96	60/102/42	1
CD14	rs5744455	5	139993491	C/T	0,2	0,384	293/154/14	49/40/7	220/120/14	122/74/7	1
CD14 (IK)	rs1583005	5	14001721	C/T	0,418	0,390	152/235/74	21/50/25	119/176/59	54/109/40	2
CDH1	rs16260	16	67328535	C/A	0,275	0,907	239/187/34	47/39/9	173/146/33	113/80/10	1
CDH1	rs1125557	16	67367100	A/G	0,433	0,345	141/239/82	35/42/19	107/179/69	69/102/32	1
CDH1	rs12597188	16	67372327	G/A	0,317	1,000	211/201/47	42/42/11	152/159/43	101/84/15	1
CDH1	rs9929218	16	67378447	G/A	0,292	0,576	227/199/37	45/42/9	164/157/34	108/84/12	1
CDH1	rs7186053	16	67396794	G/A	0,423	0,705	154/222/85	32/44/20	113/175/66	73/91/39	2

Gen	SNP <sup>a</sup>	Cromosoma <sup>b</sup>	Posición cromosomal <sup>c</sup>	Alelos <sup>d</sup>	MAF <sup>e</sup>	EHW <sup>f</sup>	Progresión Controles <sup>g</sup>	Progresión Casos <sup>h</sup>	Regresión Controles <sup>i</sup>	Regresión Casos <sup>j</sup>	Bloque de haplotipos <sup>k</sup>
CDH1	rs4783573	16	67398089	A/G	0,356	0,419	196/205/60	42/48/5	156/161/37	82/92/28	2
IL1RN	rs2637988	2	113593250	A/G	0,389	0,041	182/201/80	37/44/15	139/165/51	80/80/44	No es tagSNP
MUC2	rs10902073	11	1050934	C/A	0,336	0,097	211/191/61	40/40/16	141/155/59	110/76/18	2
MUC2	rs10794281	11	1053149	T/C	0,397	0,176	175/206/82	33/45/18	123/158/74	85/93/26	2
MUC2	rs2071174	11	1063712	T/C	0,29	1,000	235/188/39	44/46/5	182/149/22	97/85/22	Recomb
MUC2	rs7944723	11	1083710	C/G	0,218	0,105	278/168/16	61/30/5	219/120/16	120/78/5	Recomb
MUC2	rs10794293	11	1088939	C/T	0,347	0,358	205/199/59	41/44/11	158/156/41	88/87/29	4
MUC2	rs3924453	11	1095806	G/A	0,263	0,341	247/186/29	54/39/3	188/145/22	113/80/10	5
MUC2	rs4077759	11	1095976	T/C	0,353	0,361	187/218/55	41/45/7	147/162/41	81/101/21	5
MUC6	rs6597947	11	1027029	G/T	0,079	0,005	396/58/8	83/13/0	307/46/2	172/25/6	No es tagSNP

<sup>a</sup> Los SNPs se ordenan en sentido 5'-3'. <sup>b-d</sup> Información de HapMap. <sup>d</sup> Alelo común/alelo variante. <sup>e</sup> Frecuencia del alelo menor en la muestra total estudiada. <sup>f</sup> p-valor de equilibrio de Hardy-Weinberg en los controles. <sup>g-i</sup> Número de controles y casos con cada genotipo en los grupos de progresión y regresión <sup>k</sup> Asignación de cada tagSNP a número de bloque de haplotipos o punto de recombinación. "Recomb" significa punto de recombinación.

Tabla Anexa 2. Genes sobre (Fold Change≥3) o sub-expresados (Fold Change≤0.5) en la comparación MI-No CG vs Sanos.					
Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor ajustado <sup>c</sup>	Asociados con MI <sup>d</sup>	
<i>GAST</i>	gastrin	0.110	1.85E-06	SI <sup>41</sup>	
<i>SST</i>	somatostatin	0.127	7.18E-09	SI <sup>41</sup>	
<i>PGC</i>	progastricsin (pepsinogen C)	0.130	2.20E-05	SI <sup>314</sup>	
<i>GKN1</i>	gastrokine 1	0.141	1.02E-03	SI <sup>222</sup>	
<i>GKN2</i>	gastrokine 2	0.195	1.87E-03	SI <sup>221</sup>	
<i>SLC5A5</i>	solute carrier family 5 (sodium iodide symporter), member 5	0.228	5.85E-04	Nueva molécula similar <sup>315</sup>	
<i>CLU</i>	clusterin	0.232	6.51E-06	Nueva molécula	
<i>GLUL</i>	glutamate-ammonia ligase	0.247	9.50E-09	Nueva molécula	
<i>ASS1</i>	argininosuccinate synthase 1	3.010	1.95E-06	Nueva molécula	
<i>GLS</i>	glutaminase	3.013	3.02E-07	Nueva molécula	
<i>SLC22A18A5</i>	solute carrier family 22 (organic cation transporter), member	3.025	1.97E-08	Nueva molécula similar <sup>315</sup>	
<i>TRIM40</i>	tripartite motif containing 40	3.043	2.09E-04	Nueva molécula	
<i>ATP1A1</i>	ATPase, Na+/K+ transporting, alpha 1 polypeptide	3.049	9.28E-08	Nueva molécula	
<i>IRF4</i>	interferon regulatory factor 4	3.065	2.20E-05	Nueva molécula	
<i>USP2</i>	ubiquitin specific peptidase 2	3.067	7.83E-09	Nueva molécula	
<i>CTSZ</i>	cathepsin Z	3.090	9.47E-09	Nueva molécula	
<i>HOXB6</i>	homeobox B6	3.118	4.74E-07	Nueva molécula	
<i>S100A10</i>	S100 calcium binding protein A10	3.133	1.58E-06	Nueva molécula	
<i>CDHR5</i>	cadherin-related family member 5	3.145	9.94E-09	Nueva molécula	
<i>ADH6</i>	alcohol dehydrogenase 6 (class V)	3.147	5.83E-11	Nueva molécula	
<i>EPCAM</i>	epithelial cell adhesion molecule	3.166	8.06E-08	SI <sup>316</sup>	
<i>MPP1</i>	membrane protein, palmitoylated 1, 55kDa	3.166	2.05E-05	Nueva molécula	
<i>HNF4G</i>	hepatocyte nuclear factor 4, gamma	3.169	1.37E-07	Nueva molécula similar <sup>80</sup>	
<i>SLC39A5</i>	solute carrier family 39 (metal ion transporter), member 5	3.194	2.24E-09	Nueva molécula similar <sup>315</sup>	
<i>IL2RG</i>	interleukin 2 receptor, gamma	3.217	1.11E-06	Nueva molécula	



<b>Símbolo del gen <sup>a</sup></b>	<b>Nombre del gen</b>	<b>Fold Change <sup>b</sup></b>	<b>p-valor ajustado <sup>c</sup></b>	<b>Asociados con MI <sup>d</sup></b>
UGT2A3	UDP glucuronosyltransferase 2 family, polypeptide A3	3,238	8,31E-06	Nueva molécula
GLRX	glutaredoxin (thioltransferase)	3,245	3,60E-07	Nueva molécula
CDX2	caudal type homeobox 2	3,284	1,13E-10	SI <sup>64</sup>
PDZD3	PDZ domain containing 3	3,300	2,47E-08	Nueva molécula
MS4A10	membrane-spanning 4-domains, subfamily A, member 10	3,307	7,97E-05	Nueva molécula
FAM211A	family with sequence similarity 211, member A	3,350	7,52E-08	Nueva molécula
ETHE1	ethylmalonic encephalopathy 1	3,398	4,78E-07	Nueva molécula
TM6SF2	transmembrane 6 superfamily member 2	3,411	3,54E-08	Nueva molécula
CALML4	calmodulin-like 4	3,432	9,50E-07	Nueva molécula
CEACAM18	carcinoembryonic antigen-related cell adhesion molecule 18	3,439	7,72E-07	Nueva molécula similar <sup>225</sup>
SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity)	3,471	2,07E-06	Nueva molécula
CD55	CD55 molecule, decay accelerating factor for complement	3,484	1,08E-06	SI <sup>261</sup>
TMEM150B	transmembrane protein 150B	3,497	5,63E-11	Nueva molécula
LOC100124692	maltase-glucoamylase (alpha-glucoosidase) pseudogene	3,508	9,91E-06	SI <sup>196</sup>
BTNL8	butyrophilin-like 8	3,511	2,84E-04	Nueva molécula
MSLN	mesothelin	3,519	1,17E-06	Nueva molécula
KLK1	kallikrein 1	3,524	1,64E-09	Nueva molécula
ACSL5	acyl-CoA synthetase long-chain family member 5	3,524	1,71E-07	Nueva molécula
CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	3,530	1,12E-10	SI <sup>206</sup>
FOLH1 /// FOLH1B	folate hydrolase (prostate-specific membrane antigen) 1 ///	3,540	2,67E-05	Nueva molécula
ABCG8	ATP-binding cassette, sub-family G (WHITE), member 8	3,550	6,51E-07	Nueva molécula
IGHA1/2/IGHG1/2/3/IGHM/IGHV4-31	immunoglobulin heavy locus /// immunoglobulin heavy constant	3,551	1,96E-06	SI <sup>209</sup>
ACHE	acetylcholinesterase	3,570	2,85E-10	Nueva molécula
APOBEC1	apolipoprotein B mRNA editing enzyme, catalytic	3,579	3,26E-06	Nueva molécula
VIL1	villin 1	3,602	1,58E-05	SI <sup>80</sup>

<b>Símbolo del gen <sup>a</sup></b>	<b>Nombre del gen</b>	<b>Fold Change <sup>b</sup></b>	<b>p-valor ajustado <sup>c</sup></b>	<b>Asociados con MI <sup>d</sup></b>
<i>HOXA13</i>	homeobox A13	3,624	1,98E-04	SI <sup>264</sup>
<i>FLVCR2</i>	feline leukemia virus subgroup C cellular receptor family,	3,628	6,78E-10	Nueva molécula
<i>TTL6</i>	tubulin tyrosine ligase-like family, member 6	3,628	2,11E-07	Nueva molécula
<i>UGT1A1/3-10</i>	UDP glucuronosyltransferase 1 family, polypeptide A1 ///	3,637	6,34E-06	Nueva molécula
<i>FAM84A</i>	family with sequence similarity 84, member A	3,657	2,33E-07	Nueva molécula
<i>CDA</i>	cytidine deaminase	3,695	7,51E-06	Nueva molécula
<i>ABP1</i>	amiloride binding protein 1 (amine oxidase (copper-	3,703	2,90E-08	Nueva molécula
<i>CCL14/15</i>	chemokine (C-C motif) ligand 14 ///	3,704	3,73E-06	Nueva molécula
<i>ANXA13</i>	annexin A13	3,710	1,47E-06	Nueva molécula
<i>HHLA2</i>	HERV-H LTR-associating 2	3,739	1,40E-06	Nueva molécula
<i>TMEM139</i>	transmembrane protein 139	3,762	2,65E-10	Nueva molécula
<i>LDHA</i>	lactate dehydrogenase A	3,769	1,37E-09	Nueva molécula
<i>BCMO1</i>	beta-carotene 15,15'-monooxygenase 1	3,785	5,96E-06	Nueva molécula
<i>ALPI</i>	alkaline phosphatase, intestinal	3,793	2,64E-07	Nueva molécula
<i>SLC35G1</i>	solute carrier family 35, member G1	3,804	1,12E-07	Nueva molécula similar <sup>315</sup>
<i>PIGR</i>	polymeric immunoglobulin receptor	3,809	2,98E-08	SI <sup>317</sup>
<i>GSTA1</i>	glutathione S-transferase alpha 1	3,811	5,12E-04	Nueva molécula
<i>NT5E</i>	5'-nucleotidase, ecto (CD73)	3,822	5,04E-06	Nueva molécula
<i>MME</i>	membrane metallo-endopeptidase	3,830	2,93E-05	Nueva molécula
<i>FABP1</i>	fatty acid binding protein 1, liver	3,973	1,93E-09	SI <sup>80</sup>
<i>CLDN3</i>	claudin 3	3,978	4,95E-13	SI <sup>80</sup>
<i>NR1I2</i>	nuclear receptor subfamily 1, group I, member 2	4,045	1,88E-09	Nueva molécula
<i>PEPD</i>	peptidase D	4,067	1,61E-08	Nueva molécula
<i>TMEM45B</i>	transmembrane protein 45B	4,095	2,34E-09	Nueva molécula
<i>GBA3</i>	glucosidase, beta, acid 3 (cytosolic)	4,121	1,58E-05	Nueva molécula
<i>CREB3L3</i>	cAMP responsive element binding protein 3-like 3	4,132	4,37E-08	Nueva molécula

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor ajustado <sup>c</sup>	Asociados con MI <sup>d</sup>
KRT20	keratin 20	4.252	3.41E-06	SI <sup>81</sup>
MYO7B	myosin VIIIB	4.341	1.61E-12	Nueva molécula
CCL25	chemokine (C-C motif) ligand 25	4.402	8.47E-08	Nueva molécula
MOGAT3	monoacylglycerol O-acyltransferase 3	4.426	6.19E-14	Nueva molécula
GCNT3	glucosaminyl (N-acetyl) transferase 3, mucin type	4.432	7.16E-11	Nueva molécula
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	4.457	7.18E-09	Nueva molécula
CDX1	caudal type homeobox 1	4.497	7.83E-09	SI <sup>79</sup>
OSTalpha	organic solute transporter alpha	4.500	4.04E-08	Nueva molécula
AFAP1-AS1	AFAP1 antisense RNA 1 (non-protein coding)	4.522	4.13E-08	Nueva molécula
SH3D21	SH3 domain containing 21	4.535	3.41E-13	Nueva molécula
ATP1B3	ATPase, Na+/K+ transporting, beta 3 polypeptide	4.588	4.66E-10	Nueva molécula
REEP6	receptor accessory protein 6	4.669	9.79E-11	Nueva molécula
SULT1E1	sulfotransferase family 1E, estrogen-preferring, member 1	4.727	6.74E-09	SI <sup>272</sup>
SLC5A9	solute carrier family 5 (sodium/glucose cotransporter),	4.787	1.32E-08	Nueva molécula similar <sup>315</sup>
CAMK2N1	calcium/calmodulin-dependent protein kinase II inhibitor 1	4.827	6.08E-09	Nueva molécula
A1CF	APOBEC1 complementation factor	4.837	6.82E-12	Nueva molécula
CLRN3	clarin 3	4.887	4.04E-10	Nueva molécula
VNN1	vanin 1	4.894	2.14E-11	Nueva molécula
ACY3	aspartoacylase (aminocyclase) 3	4.959	2.45E-10	Nueva molécula
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	4.969	2.65E-08	Nueva molécula
CEACAM20	Carcinoembryonic antigen-related cell adhesion molecule 20	4.990	2.65E-05	Nueva molécula similar <sup>225</sup>
HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	5.049	6.39E-09	SI <sup>318</sup>
DGKA	diacylglycerol kinase, alpha 80kDa	5.127	1.18E-11	Nueva molécula
SLC15A1	solute carrier family 15 (oligopeptide transporter), member	5.141	2.36E-05	Nueva molécula similar <sup>315</sup>
RBP2	retinol binding protein 2, cellular	5.209	7.58E-07	Nueva molécula
SLC4A7	solute carrier family4, sodium bicarbonate cotransporter, member7	5.213	2.97E-10	Nueva molécula similar <sup>315</sup>

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor ajustado <sup>c</sup>	Asociados con MI <sup>d</sup>
<i>SERPINA1</i>	serpin peptidase inhibitor, clade A (alpha-1 antitrypsin)	5.213	1.34E-08	Nueva molécula
<i>CIDEB</i>	cell death-inducing DFFA-like effector b	5.285	4.04E-11	Nueva molécula
<i>DHRS11</i>	dehydrogenase/reductase (SDR family) member 11	5.307	5.99E-13	Nueva molécula
<i>EFNA2</i>	efrin-A2	5.318	4.72E-15	Nueva molécula
<i>GK</i>	glycerol kinase	5.408	2.45E-10	Nueva molécula
<i>ITLN1</i>	intelectin 1 (galactofuranose binding)	5.583	1.65E-09	Nueva molécula
<i>GIP</i>	gastric inhibitory polypeptide	5.606	2.65E-08	Nueva molécula
<i>TM4SF20</i>	transmembrane 4 L six family member 20	5.614	1.64E-09	Nueva molécula
<i>DPP4</i>	dipeptidyl-peptidase 4	5.645	2.58E-08	Nueva molécula
<i>XPNPEP2</i>	X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-	5.756	1.99E-08	Nueva molécula
<i>SPINK4</i>	serine peptidase inhibitor, Kazal type 4	5.780	1.51E-12	Nueva molécula
<i>PLA2G2A</i>	phospholipase A2, group IIA (platelets, synovial fluid)	5.804	3.90E-05	Nueva molécula
<i>HEPH</i>	hephaestin	5.877	9.01E-11	Nueva molécula
<i>GDA</i>	guanine deaminase	5.901	1.15E-09	Nueva molécula
<i>SLC6A8</i>	solute carrier family 6 (neurotransmitter transporter,	5.926	1.37E-09	Nueva molécula similar <sup>315</sup>
<i>PRAP1</i>	proline-rich acidic protein 1	6.050	9.10E-14	Nueva molécula
<i>SLC46A3</i>	solute carrier family 46, member 3	6.143	2.76E-10	Nueva molécula similar <sup>315</sup>
<i>GPA33</i>	glycoprotein A33 (transmembrane)	6.199	4.49E-15	SI, Marcador de Barret
<i>SLC7A9</i>	solute carrier family 7 (glycoprotein-associated amino acid	6.238	3.32E-10	Nueva molécula similar <sup>315</sup>
<i>ABCC13</i>	ATP-binding cassette, sub-family C (CFTR/MRP), member 13,	6.264	3.61E-09	Nueva molécula
<i>CES2</i>	carboxylesterase 2	6.404	3.80E-12	Nueva molécula
<i>CLDN4</i>	claudin 4	6.422	2.57E-12	SI <sup>80</sup>
<i>ZG16</i>	zymogen granule protein 16 homolog (rat)	6.658	3.98E-10	Nueva molécula
<i>MUC4</i>	mucin 4, cell surface associated	6.676	1.37E-06	SI <sup>267</sup>
<i>APOC3</i>	apolipoprotein C-III	6.797	4.99E-06	Nueva molécula
<i>MUC13</i>	mucin 13, cell surface associated	6.902	5.11E-13	SI <sup>81</sup>

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor ajustado <sup>c</sup>	Asociados con MI <sup>d</sup>
CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6	6.931	4.02E-10	SI <sup>225</sup>
APOA1	apolipoprotein A-I	7.126	3.72E-06	SI <sup>80</sup>
APOA4	apolipoprotein A-IV	7.165	2.40E-07	SI <sup>80</sup>
ABCC2	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	7.356	5.15E-07	Nueva molécula
SLC17A4	solute carrier family 17 (sodium phosphate), member 4	7.418	2.41E-11	Nueva molécula similar <sup>315</sup>
LINC00483	long intergenic non-protein coding RNA 483	7.537	1.19E-10	Nueva molécula
RG52	regulator of G-protein signaling 2, 24kDa	7.558	1.61E-10	Nueva molécula
TM4SF4	transmembrane 4 L six family member 4	7.934	1.45E-11	Nueva molécula
CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	8.000	1.32E-07	SI <sup>319</sup>
DEFA5	defensin, alpha 5, Paneth cell-specific	8.219	1.67E-05	Nueva molécula
SLC5A1	solute carrier family 5 (sodium/glucose cotransporter),	8.427	1.26E-08	Nueva molécula similar <sup>315</sup>
SLC13A2	solute carrier family 13 (sodium-dependent dicarboxylate	8.586	1.44E-13	Nueva molécula similar <sup>315</sup>
ACE2	angiotensin I converting enzyme (peptidyl-dipeptidase A) 2	8.700	7.16E-11	SI <sup>81</sup>
HKDC1	hexokinase domain containing 1	8.951	2.26E-15	Nueva molécula
TFF3	trefoil factor 3 (intestinal)	9.082	7.66E-12	SI <sup>268</sup>
CHP2	calcineurin-like EF hand protein 2	9.266	4.13E-14	Nueva molécula
REG3A	regenerating islet-derived 3 alpha	10.098	2.93E-05	Nueva molécula similar <sup>214</sup>
MALL	MARVEL domain-containing protein 1-like /// mal, T-cell	10.793	1.02E-13	Nueva molécula
MOGAT2	monoacylglycerol O-acyltransferase 2	11.345	5.32E-15	Nueva molécula
ONECUT2	one cut homeobox 2	11.448	1.88E-15	Nueva molécula
BTNL3	butyrophilin-like 3	12.711	1.21E-09	Nueva molécula
SLC6A19	solute carrier family 6 (neutral amino acid transporter), member 19	13.567	9.98E-15	SI <sup>315</sup>
MEP1B	mepirin A, beta	14.867	1.44E-13	SI <sup>214</sup>
MUC3A /// MUC3B	mucin-3A-like ///mucin 3A, cell surface associated /// mucin 3B, cell surface	15.900	5.32E-16	Nueva molécula similar <sup>14</sup>
CDH17	cadherin 17, LI cadherin (liver-intestine)	18.026	1.34E-17	SI <sup>80</sup>
TMPRSS15	transmembrane protease, serine 15	18.481	2.29E-12	Nueva molécula

Símbolo del gen <sup>a</sup>	Nombre del gen	Fold Change <sup>b</sup>	p-valor ajustado <sup>c</sup>	Asociados con MI <sup>d</sup>
MUC12	mucin 12, cell surface associated	18.713	1.17E-08	Nueva molécula similar <sup>14</sup>
SLC26A3	solute carrier family 26, member 3	20.210	6.36E-11	Nueva molécula similar <sup>315</sup>
MUC17	mucin 17, cell surface associated	20.521	3.26E-15	Nueva molécula similar <sup>14</sup>
MTTP	microsomal triglyceride transfer protein	25.142	5.42E-14	SI <sup>79</sup>
CPS1	carbamoyl-phosphate synthase 1, mitochondrial	26.246	7.02E-19	SI <sup>269</sup>
MUC2	mucin 2, oligomeric mucus/gel-forming	26.300	1.26E-23	SI <sup>266</sup>
SI	sucrase-isomaltase (alpha-glucosidase)	30.086	1.05E-15	SI <sup>214</sup>
REG4	regenerating islet-derived family, member 4	32.200	1.16E-13	SI <sup>81</sup>
CLCA1	chloride channel accessory 1	33.778	6.18E-16	Nueva molécula
OLFM4	olfactomedin 4	37.040	1.10E-18	SI <sup>81, 227</sup>
APOB	apolipoprotein B (including Ag(x) antigen)	43.231	1.16E-13	SI <sup>80, 81</sup>
ANPEP	alanyl (membrane) aminopeptidase	49.970	2.86E-23	Nueva molécula
ALDOB	aldolase B, fructose-bisphosphate	58.404	7.64E-18	SI <sup>80</sup>
FABP1 /// PRDM10	fatty acid binding protein 1, liver /// PR domain containing 10	80.560	1.26E-23	SI <sup>80</sup>
DMBT1	deleted in malignant brain tumors 1	98.224	2.20E-23	SI <sup>271</sup>

<sup>a</sup> En el caso de la existencia de más de 1 gen significa que la sonda reconoce varios genes. <sup>b</sup> *Fold Change* es el promedio de la expresión de MI. No CG menos mucosa gástrica sana; los genes están ordenados de menor a mayor por dicho valor. <sup>c</sup> p-valor resultante del test de permutaciones corregido mediante FDR (tasa de falsos positivos). <sup>d</sup> Genes asociados con MI en estudios de expresión, de asociación genética, proteómicos y otros. “Nueva molécula” significa que es la primera vez que se identifica este gen como diferencialmente expresado en la MI. “Nueva molécula similar” significa que otro miembro de su familia génica ha sido previamente asociado con la MI.

Tabla Anexa 3. Conjuntos génicos sobre-expresados resultado del GSEA catálogo c2All.v.5 para la comparación MII-CG vs MII-No CG.						
Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
REACTOME_G_BETA_GAMMA_SIGNALLING_THROUGH_PI3KGAMMA	Ciclo y Proliferación celular	20	0,5932	0,0070	0,0235	895
REN_BOUND_BY_E2F	Ciclo y Proliferación celular	52	0,5709	0,0000	0,0009	3374
REACTOME_AUTODEGRADATION_OF_CDH1_BY_CDH1_APC_C	Ciclo y Proliferación celular	52	0,5589	0,0000	0,0013	5579
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_O	Ciclo y Proliferación celular	58	0,5548	0,0000	0,0010	5579
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PRO	Ciclo y Proliferación celular	59	0,5440	0,0000	0,0014	5579
REACTOME_G2_M_CHECKPOINTS	Ciclo y Proliferación celular	36	0,5457	0,0017	0,0090	4518
OLSSON_E2F3_TARGETS_DN	Ciclo y Proliferación celular	40	0,5378	0,0000	0,0080	4553
REACTOME_MAPK_TARGETS_NUCLEAR_EVENTS_MEDIATED_BY_MAP_KINASES	Ciclo y Proliferación celular	28	0,5265	0,0068	0,0239	4787
REACTOME_CELL_CYCLE_CHECKPOINTS	Ciclo y Proliferación celular	101	0,5191	0,0000	0,0007	5579
KARLSSON_TGFB1_TARGETS_UP	Ciclo y Proliferación celular	110	0,5140	0,0000	0,0006	4733
ISHIDA_E2F_TARGETS	Ciclo y Proliferación celular	45	0,5046	0,0000	0,0112	6333
REACTOME_REGULATION_OF_MITOTIC_CELL_CYCLE	Ciclo y Proliferación celular	70	0,5014	0,0000	0,0034	6112
BIOCARTA_P38MAPK_PATHWAY	Ciclo y Proliferación celular	35	0,5000	0,0050	0,0253	6659
PID_PI3KIPATHWAY	Ciclo y Proliferación celular	40	0,4723	0,0032	0,0361	6697
ASTIER_INTEGRIN_SIGNALING	Ciclo y proliferación celular	54	0,4491	0,0034	0,0312	3343
KEGG_CELL_CYCLE	Ciclo y Proliferación celular	109	0,4192	0,0000	0,0177	6872
IGLESIAS_E2F_TARGETS_UP	Ciclo y Proliferación celular	142	0,4177	0,0000	0,0117	5813
WHITFIELD_CELL_CYCLE_G2_M	Ciclo y Proliferación celular	190	0,3930	0,0000	0,0138	5781
LABBE_WNT3A_TARGETS_UP	Ciclo y Proliferación celular	99	0,3905	0,0063	0,0450	6333
CHANG_CYCLING_GENES	Ciclo y Proliferación celular	123	0,3883	0,0000	0,0324	6140
REACTOME_CELL_CYCLE_MITOTIC	Ciclo y Proliferación celular	273	0,3824	0,0000	0,0113	6864
WHITFIELD_CELL_CYCLE_M_G1	Ciclo y Proliferación celular	128	0,3820	0,0016	0,0341	4529
WHITFIELD_CELL_CYCLE_G2	Ciclo y Proliferación celular	152	0,3637	0,0000	0,0428	6121
REACTOME_CELL_CYCLE	Ciclo y Proliferación celular	329	0,3550	0,0000	0,0246	5615

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
COLLER_MYC_TARGETS_UP	Oncogenes	24	0,6308	0,0000	0,0034	4010
BERENJENO_TRANSFORMED_BY_RHOA_FOREVER_DN	Oncogenes	28	0,6268	0,0000	0,0021	1908
MENSSEN_MYC_TARGETS	Oncogenes	43	0,6151	0,0000	0,0005	5289
CROONQUIST_NRAS_SIGNALING_UP	Oncogenes	35	0,5886	0,0000	0,0026	2872
CROONQUIST_NRAS_SIGNALING_DN	Oncogenes	64	0,5429	0,0000	0,0008	5248
SEIDEN_ONCOGENESIS_BY_MET	Oncogenes	80	0,5345	0,0000	0,0007	6564
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_UP	Oncogenes	40	0,5309	0,0000	0,0087	5400
SCHLOSSER_MYC_TARGETS_REPRESSED_BY_SERUM	Oncogenes	142	0,5280	0,0000	0,0000	4330
YU_MYC_TARGETS_UP	Oncogenes	39	0,5122	0,0034	0,0157	2259
PID_MYC_ACTIVPATHWAY	Oncogenes	71	0,5053	0,0000	0,0034	5285
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_DN	Oncogenes	43	0,5040	0,0017	0,0172	4658
SCHUHMACHER_MYC_TARGETS_UP	Oncogenes	72	0,5020	0,0000	0,0027	4783
DANG_MYC_TARGETS_UP	Oncogenes	127	0,4771	0,0000	0,0018	5760
LEE_LIVER_CANCER_MYC_UP	Oncogenes	44	0,4730	0,0033	0,0304	6087
CHIARADONNA_NEOPLASTIC_TRANSFORMATION_KRAS_UP	Oncogenes	116	0,4596	0,0000	0,0039	5932
YU_MYC_TARGETS_DN	Oncogenes	48	0,4458	0,0033	0,0447	5126
BERENJENO_TRANSFORMED_BY_RHOA_UP	Oncogenes	34	0,4278	0,0000	0,0006	4980
KIM_MYC_AMPLIFICATION_TARGETS_UP	Oncogenes	166	0,4118	0,0000	0,0106	4720
SWEET_LUNG_CANCER_KRAS_DN	Oncogenes	369	0,3381	0,0000	0,0368	4938
XU_HGF_TARGETS_INDUCED_BY_AKT1_6HR	Oncogenes	17	0,6228	0,0088	0,0183	5499
SCIAN_CELL_CYCLE_TARGETS_OF_TP53_AND_TP73_DN	Supresores tumorales	21	0,5977	0,0036	0,0179	2879
EGUCHI_CELL_CYCLE_RB1_TARGETS	Supresores tumorales	23	0,5906	0,0000	0,0114	4556
REACTOME_P53_DEPENDENT_G1_DNA_DAMAGE_RESPONSE	Supresores tumorales	50	0,5398	0,0000	0,0034	6064
ONGUSAHA_TP53_TARGETS	Supresores tumorales	33	0,5336	0,0032	0,0130	4601
APRELIKOVA_BRCA1_TARGETS	Supresores tumorales	45	0,5008	0,0000	0,0161	7075
TANG_SENESCENCE_TP53_TARGETS_DN	Supresores tumorales	48	0,4756	0,0031	0,0235	4957



Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
PUJANA_BRCA_CENTERED_NETWORK	Supresores tumorales	108	0,4689	0,0000	0,0033	4165
MARKEY_RB1_ACUTE_LOF_DN	Supresores tumorales	200	0,4641	0,0000	0,0007	5105
CHICAS_RB1_TARGETS_GROWING	Supresores tumorales	203	0,4365	0,0000	0,0029	4697
WELCSH_BRCA1_TARGETS_DN	Supresores tumorales	120	0,4243	0,0000	0,0139	6946
HEDENFALK_BREAST_CANCER_BRCA1_VS_BRCA2	Supresores tumorales	145	0,4241	0,0000	0,0083	5652
MARKEY_RB1_CHRONIC_LOF_DN	Supresores tumorales	97	0,4237	0,0000	0,0177	4752
PUJANA_BRCA2_PCC_NETWORK	Supresores tumorales	361	0,4007	0,0000	0,0034	4447
WELCSH_BRCA1_TARGETS_UP	Supresores tumorales	185	0,3920	0,0000	0,0160	6907
CHICAS_RB1_TARGETS_SENESCENT	Supresores tumorales	491	0,3914	0,0000	0,0033	4715
MARKEY_RB1_ACUTE_LOF_UP	Supresores tumorales	184	0,3693	0,0000	0,0365	3592
CHICAS_RB1_TARGETS_CONFLUENT	Supresores tumorales	478	0,3608	0,0000	0,0131	4907
CHANDRAN_METASTASIS_TOP50_UP	Invasión y metástasis	36	0,5389	0,0000	0,0108	4399
AIGNER_ZEB1_TARGETS	Invasión y metástasis	32	0,5343	0,0066	0,0150	6349
PEDERSEN_METASTASIS_BY_ERBB2_ISOFORM_1	Invasión y metástasis	38	0,4895	0,0051	0,0266	2804
ALONSO_METASTASIS_UP	Invasión y metástasis	169	0,4775	0,0000	0,0007	5250
GRADE_METASTASIS_DN	Invasión y metástasis	39	0,4758	0,0048	0,0374	5765
JECHLINGER_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_UP	Invasión y metástasis	64	0,4463	0,0000	0,0280	4237
ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE	Invasión y metástasis	56	0,4412	0,0066	0,0366	5250
SARRIO_EPITHELIAL_MESENCHYMAL_TRANSITION_UP	Invasión y metástasis	150	0,4238	0,0000	0,0077	5589
WANG_TUMOR_INVASIVENESS_UP	Invasión y metástasis	332	0,4197	0,0000	0,0019	6767
ONDER_CDH1_TARGETS_1_DN	Invasión y metástasis	146	0,4191	0,0000	0,0093	5908
PEDERSEN_METASTASIS_BY_ERBB2_ISOFORM_4	Invasión y metástasis	93	0,4186	0,0000	0,0235	3368
PROVENZANI_METASTASIS_UP	Invasión y metástasis	175	0,4013	0,0000	0,0111	5733
GOTZMANN_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_DN	Invasión y metástasis	186	0,3748	0,0000	0,0261	6789
CHANDRAN_METASTASIS_UP	Invasión y metástasis	187	0,3737	0,0000	0,0273	6995
ROZANOV_MMP14_TARGETS_UP	Invasión y metástasis	217	0,3438	0,0000	0,0476	4755

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
SCHOEN_NFKB_SIGNALING	Inflamación	32	0,5170	0,0018	0,0274	2552
REACTOME_ACTIVATION_OF_NF_KAPPAB_IN_B_CELLS	Inflamación	59	0,4590	0,0050	0,0229	5579
ZHOU_TNF_SIGNALING_30MIN	Inflamación	50	0,4519	0,0016	0,0342	7334
MARZEC_IL2_SIGNALING_UP	Inflamación	96	0,4368	0,0000	0,0158	3796
HINATA_NFKB_TARGETS_FIBROBLAST_UP	Inflamación	67	0,4336	0,0083	0,0311	5623
BOSCO_INTERFERON_INDUCED_ANTIVIRAL_MODULE	Inflamación	65	0,4320	0,0016	0,0317	5632
REACTOME_ACTIVATED_TLR4_SIGNALING	Inflamación	80	0,4284	0,0000	0,0213	6722
REACTOME_NFKB_AND_MAP_KINASES_ACTIVATION_MEDIATED_BY_TLR4	Inflamación	63	0,4224	0,0082	0,0461	6722
SANA_TNF_SIGNALING_DN	Inflamación	76	0,4170	0,0064	0,0390	5458
PHONG_TNF_RESPONSE_VIA_P38_COMPLETE	Inflamación	201	0,4164	0,0000	0,0050	5985
HINATA_NFKB_TARGETS KERATINOCYTE_UP	Inflamación	81	0,4111	0,0031	0,0400	5458
THEILGAARD_NEUTROPHIL_AT_SKIN_WOUND_DN	Inflamación	205	0,3786	0,0000	0,0192	6872
PHONG_TNF_RESPONSE_VIA_P38_PARTIAL	Inflamación	137	0,3660	0,0016	0,0497	3878
BURTON_ADIPOGENESIS_PEAK_AT_16HR	Metabolismo lipídico	36	0,6194	0,0000	0,0009	4413
BURTON_ADIPOGENESIS_PEAK_AT_8HR	Metabolismo lipídico	38	0,5861	0,0000	0,0020	5841
BURTON_ADIPOGENESIS_3	Metabolismo lipídico	87	0,4981	0,0000	0,0024	5105
BURTON_ADIPOGENESIS_PEAK_AT_24HR	Metabolismo lipídico	34	0,4865	0,0083	0,0379	4269
GERHOLD_ADIPOGENESIS_DN	Metabolismo lipídico	58	0,4804	0,0000	0,0116	4460
BURTON_ADIPOGENESIS_5	Metabolismo lipídico	106	0,4593	0,0000	0,0053	5802
BURTON_ADIPOGENESIS_8	Metabolismo lipídico	78	0,4580	0,0000	0,0100	5180
BURTON_ADIPOGENESIS_2	Metabolismo lipídico	68	0,4397	0,0000	0,0228	7604
PELLICCIOтта_HDAC_IN_ANTIGEN_PRESENTATION_DN	Procesamiento antigénico	47	0,5167	0,0000	0,0080	5579
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION	Procesamiento antigénico	67	0,5005	0,0000	0,0045	5579
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGEN_S_ENDOSOMES	Procesamiento antigénico	43	0,4942	0,0017	0,0182	5579
LENAOUR_DENDRITIC_CELL_MATURATION_DN	Procesamiento antigénico	116	0,4836	0	0,0014	5840
GOLDRATH_ANTIGEN_RESPONSE	Procesamiento antigénico	308	0,4199	0,0000	0,0024	6476

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION	Procesamiento antigénico	205	0,3996	0,0000	0,0102	5579
LINDSTEDT_DENDRITIC_CELL_MATURATION_C	Procesamiento antigénico	63	0,4453	0,0000	0,0245	5360
REACTOME_REGULATION_OF_APOPTOSIS	Apoptosis	52	0,5811	0,0000	0,0007	5579
WU_APOPTOSIS_BY_CDKN1A_VIA_TP53	Apoptosis	47	0,5501	0,0000	0,0039	4506
REACTOME_APOPTOSIS	Apoptosis	131	0,4581	0,0000	0,0024	6079
ALCALA_APOPTOSIS	Apoptosis	79	0,4443	0,0000	0,0169	5851
HOLLMANN_APOPTOSIS_VIA_CD40_UP	Apoptosis	185	0,3939	0,0000	0,0146	4113
HOLLMANN_APOPTOSIS_VIA_CD40_DN	Apoptosis	238	0,3911	0,0000	0,0119	6213
HU_ANGIOGENESIS_DN	Angiogénesis	34	0,6137	0,0000	0,0014	4892
BIOCARTA_VEGF_PATHWAY	Angiogénesis	27	0,5851	0,0000	0,0076	4437
LU_TUMOR_ANGIOGENESIS_UP	Angiogénesis	21	0,5636	0,0018	0,0345	3855
WESTON_VEGFA_TARGETS	Angiogénesis	92	0,4347	0,0000	0,0174	4529
WESTON_VEGFA_TARGETS_3HR	Angiogénesis	65	0,4332	0,0016	0,0318	4529
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	Fosforilación oxidativa	59	0,5169	0,0000	0,0034	8038
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_C	Fosforilación oxidativa	74	0,5112	0,0000	0,0021	7589
KEGG_OXIDATIVE_PHOSPHORYLATION	Fosforilación oxidativa	99	0,4762	0,0000	0,0034	7229
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	Fosforilación oxidativa	109	0,4733	0,0000	0,0031	7589
WONG_MITOCHONDRIA_GENE_MODULE	Fosforilación oxidativa	194	0,4298	0,0000	0,0031	6035
PID_INTEGRIN5_PATHWAY	Adhesión celular	16	0,6307	0,0034	0,0177	3850
PID_INTEGRIN_A4B1_PATHWAY	Adhesión celular	30	0,5289	0,0084	0,0243	4395
PID_INTEGRIN1_PATHWAY	Adhesión celular	62	0,4237	0,0097	0,0447	4241
ST_INTEGRIN_SIGNALING_PATHWAY	Adhesión celular	77	0,3982	0,0064	0,0477	6779
BIOCARTA_PROTEASOME_PATHWAY	Degradación vía proteasoma	26	0,6350	0,0000	0,0027	5541
KEGG_PROTEASOME	Degradación vía proteasoma	41	0,5924	0,0000	0,0014	5579
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1	Degradación vía proteasoma	44	0,5908	0,0000	0,0009	5579
REACTOME_ANTIGEN_PROCESSING_UBIQUITINATION_PROTEASOME_DEG	Degradación vía proteasoma	171	0,4013	0,0000	0,0133	5910

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
SARTIPY_NORMAL_AT_INSULIN_RESISTANCE_UP	Genes regulados por insulina	28	0,6462	0,0000	0,0008	2898
SARTIPY_BLUNTED_BY_INSULIN_RESISTANCE_DN	Genes regulados por insulina	17	0,6163	0,0051	0,0212	6069
ROME_INSULIN_TARGETS_IN_MUSCLE_UP	Genes regulados por insulina	390	0,4170	0,0000	0,0020	6051
LUI_THYROID_CANCER_CLUSTER_3	Cáncer de tiroides	25	0,5206	0,0097	0,0336	7129
LUI_THYROID_CANCER_PAX8_PPARG_DN	Cáncer de tiroides	40	0,5084	0,0016	0,0175	7129
DELVS_THYROID_CANCER_UP	Cáncer de tiroides	390	0,3706	0,0000	0,0111	5461
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_24HR	Respuesta a daño genómico	104	0,4289	0,0000	0,0135	3976
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_6HR	Respuesta a daño genómico	72	0,4235	0,0033	0,0319	3429
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	Respuesta a daño genómico	45	0,5871	0,0000	0,0009	5579
REACTOME_ER_PHAGOSOME_PATHWAY	Endocitosis o Fagocitosis	54	0,5722	0,0000	0,0007	5579
KEGG_LYSOSOME	Endocitosis o Fagocitosis	113	0,4036	0,0000	0,0218	7368
WANG_ESOPHAGUS_CANCER_VS_NORMAL_UP	Cáncer de esófago	103	0,4286	0,0000	0,0152	5578
WANG_ESOPHAGUS_CANCER_VS_NORMAL_DN	Cáncer de esófago	89	0,4136	0,0016	0,0320	4198
REACTOME_UNFOLDED_PROTEIN_RESPONSE	<i>Unfolded protein response</i>	66	0,4667	0,0000	0,0122	4318
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_J	Degradación de ARNm NMD	88	0,4549	0,0000	0,0090	7378
VECCHI_GASTRIC_CANCER_EARLY_UP	Cáncer gástrico	357	0,3929	0,0000	0,0049	5155

<sup>a</sup> Mecanismo molecular, proceso fisiológico normal o patológico asignado subjetivamente al conjunto genómico basado en su nombre. <sup>b</sup> Número de genes que componen la vía. <sup>c</sup> Las vías de señalización están ordenadas de mayor a menor valor de *Enrichment Score* (ES). <sup>d</sup> p-valor asociado a cada vía sin ajustar por comparaciones múltiples. <sup>e</sup> p-valor asociado a cada vía resultante de la corrección para comparaciones múltiples mediante FDR (tasa de falsos positivos). <sup>f</sup> Posición en la lista de los genes en la cual se obtiene el mayor valor de ES.

<b>Tabla Anexa 4. Conjuntos génicos sobre-expresados resultado del GSEA catálogo c2All.v.5 para la comparación MI-No CG vs Sanos.</b>						
<b>Nombre</b>	<b>Procesos funcionales <sup>a</sup></b>	<b>N <sup>b</sup></b>	<b>ES <sup>c</sup></b>	<b>p-valor nominal <sup>d</sup></b>	<b>q-valor FDR <sup>e</sup></b>	<b>Rank at max <sup>f</sup></b>
KEGG_STARCH_AND_SUCROSE_METABOLISM	Efecto Warburg no tumoral	33	0,724	0,002	0,127	1852
REACTOME_AMINO_ACID_TRANSPORT_ACROSS_THE_PLASMA_MEMB	Efecto Warburg no tumoral	28	0,666	0,000	0,125	1480
MOOHTA_GLUONEOGENESIS	Efecto Warburg no tumoral	29	0,651	0,000	0,168	527
REACTOME_GLYCOLYSIS	Efecto Warburg no tumoral	24	0,640	0,002	0,124	3435
REACTOME_AMINO_ACID_AND_OLIGOPEPTIDE_SLC_TRANSPORTERS	Efecto Warburg no tumoral	44	0,618	0,000	0,142	1480
KEGG_ARGININE_AND_PROLINE_METABOLISM	Efecto Warburg no tumoral	48	0,605	0,004	0,153	1415
REACTOME_GLUCOSE_METABOLISM	Efecto Warburg no tumoral	56	0,593	0,006	0,126	3503
KEGG_ALANINE_ASPARTATE_AND_Glutamate_METABOLISM	Efecto Warburg no tumoral	25	0,563	0,018	0,242	2144
REACTOME_GLUONEOGENESIS	Efecto Warburg no tumoral	28	0,553	0,037	0,245	3118
KEGG_GALACTOSE_METABOLISM	Efecto Warburg no tumoral	23	0,546	0,030	0,222	1852
REACTOME_PURINE_METABOLISM	Efecto Warburg no tumoral	31	0,538	0,039	0,239	1793
KEGG_GLYCOLYSIS_GLUONEOGENESIS	Efecto Warburg no tumoral	55	0,511	0,016	0,201	1249
PID_HIF1_TFPATHWAY	Efecto Warburg no tumoral	62	0,510	0,006	0,168	1741
REACTOME_TRANSPORT_OF_INORGANIC_CATIONS_ANIONS_AND_AM	Efecto Warburg no tumoral	82	0,509	0,000	0,130	1480
REACTOME_TRANSPORT_OF_GLUCOSE_AND_OTHER_SUGARS_BILE_SA	Efecto Warburg no tumoral	80	0,492	0,002	0,158	1446
MOOHTA_GLYCOGEN_METABOLISM	Efecto Warburg no tumoral	19	0,473	0,039	0,244	3304
REACTOME_TRANSPORT_OF_VITAMINS_NUCLEOSIDES_AND_RELATED	Efecto Warburg no tumoral	23	0,435	0,029	0,241	3282
REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES	Efecto Warburg no tumoral	165	0,418	0,016	0,239	3285
GROSS_HYPOXIA_VIA_ELK3_AND_HIF1A_UP	Efecto Warburg no tumoral	129	0,407	0,026	0,241	3796
REACTOME_METABOLISM_OF_CARBOHYDRATES	Efecto Warburg no tumoral	200	0,379	0,002	0,234	3526
REACTOME_CHYLOMICRON_MEDIATED_LIPID_TRANSPORT	Metabolismo lipídico	15	0,843	0,000	0,157	86
REACTOME_LIPOPROTEIN_METABOLISM	Metabolismo lipídico	24	0,748	0,000	0,123	86
REACTOME_PEROXISOMAL_LIPID_METABOLISM	Metabolismo lipídico	19	0,737	0,002	0,162	2643
REACTOME_ABCA_TRANSPORTERS_IN_LIPID_HOMEOSTASIS	Metabolismo lipídico	16	0,683	0,009	0,202	167

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
REACTOME_LIPID_DIGESTION_MOBILIZATION_AND_TRANSPORT	Metabolismo lipídico	38	0,665	0,000	0,157	396
KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	Metabolismo lipídico	16	0,630	0,019	0,200	4792
REACTOME_TRIGLYCERIDE_BIOSYNTHESIS	Metabolismo lipídico	35	0,620	0,002	0,141	3623
KEGG_PPAR_SIGNALING_PATHWAY	Metabolismo lipídico	57	0,605	0,010	0,165	1723
KEGG_GLYCEROLIPID_METABOLISM	Metabolismo lipídico	43	0,543	0,008	0,171	1926
KEGG_ETHER_LIPID_METABOLISM	Metabolismo lipídico	24	0,532	0,029	0,246	1240
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	Metabolismo lipídico	62	0,517	0,004	0,167	2590
REACTOME_GLYCEROPHOSPHOLIPID_BIOSYNTHESIS	Metabolismo lipídico	74	0,499	0,004	0,128	4070
REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS	Metabolismo lipídico	407	0,458	0,000	0,133	3048
REACTOME_FATTY_ACID_TRIACYLGLYCEROL_AND_KETONE_BODY_ME	Metabolismo lipídico	154	0,439	0,022	0,238	3735
REACTOME_PHOSPHOLIPID_METABOLISM	Metabolismo lipídico	173	0,425	0,002	0,161	4273
REACTOME_AMINO_ACID_TRANSPORT_ACROSS_THE_PLASMA_MEMB	Proliferación celular	28	0,666	0,000	0,125	1480
REACTOME_AMINO_ACID_AND_OLIGOPEPTIDE_SLC_TRANSPORTERS	Proliferación celular	44	0,618	0,000	0,142	1480
KEGG_ARGININE_AND_PROLINE_METABOLISM	Proliferación celular	48	0,605	0,004	0,153	1415
MARZEC_IL2_SIGNALING_DN	Proliferación celular	33	0,565	0,016	0,228	2984
KEGG_ALANINE_ASPARTATE_AND_Glutamate_Metabolism	Proliferación celular	25	0,563	0,018	0,242	2144
REACTOME_PURINE_METABOLISM	Proliferación celular	31	0,538	0,039	0,239	1793
REACTOME_TRANSPORT_OF_INORGANIC_CATIONS_ANIONS_AND_AM	Proliferación celular	82	0,509	0,000	0,130	1480
PLASARI_TGFB1_SIGNALING_VIA_NFIC_10HR_UP	Proliferación celular	46	0,487	0,004	0,243	1414
LABBE_TARGETS_OF_TGFB1_AND_WNT3A_DN	Proliferación celular	94	0,450	0,006	0,191	2919
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	Proliferación celular	66	0,439	0,038	0,245	2682
REACTOME_TRANSPORT_OF_VITAMINS_NUCLEOSIDES_AND_RELATED	Proliferación celular	23	0,435	0,029	0,241	3282
AMIT_EGF_RESPONSE_120_HELA	Proliferación celular	61	0,425	0,023	0,241	4300
REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES	Proliferación celular	165	0,418	0,016	0,239	3285
PLASARI_TGFB1_TARGETS_10HR_UP	Proliferación celular	167	0,392	0,002	0,215	3994

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
PID_IL8CXCR1_PATHWAY	Inflamación	25	0,662	0,000	0,140	860
HINATA_NFKB_TARGETS_KERATINOCYTE_DN	Inflamación	18	0,631	0,014	0,239	1523
WUNDER_INFLAMMATORY_RESPONSE_AND_CHOLESTEROL_UP	Inflamación	49	0,623	0,014	0,165	3370
REACTOME_RIP_MEDIATED_NFKB_ACTIVATION_VIA_DAI	Inflamación	16	0,621	0,032	0,242	4586
PID_IL8CXCR2_PATHWAY	Inflamación	31	0,582	0,014	0,171	3139
REACTOME_TRAF6_MEDIATED_NFKB_ACTIVATION	Inflamación	18	0,569	0,016	0,241	4586
REACTOME_TRAF6_MEDIATED_IRF7_ACTIVATION	Inflamación	16	0,543	0,009	0,203	6196
PID_IL4_2PATHWAY	Inflamación	50	0,520	0,010	0,140	4062
NEMETH_INFLAMMATORY_RESPONSE_LPS_UP	Inflamación	82	0,513	0,032	0,204	4873
PID_AMB2_NEUTROPHILS_PATHWAY	Inflamación	36	0,498	0,014	0,241	1044
PID_IL12_2PATHWAY	Inflamación	49	0,479	0,036	0,239	5464
BROWN_MYELOID_CELL_DEVELOPMENT_UP	Inflamación	137	0,422	0,014	0,240	4270
KEGG_CHEMOKINE_SIGNALING_PATHWAY	Inflamación	149	0,389	0,027	0,241	3524
WANG_BARRETTES_ESOPHAGUS_UP	Diferenciación intestinal	47	0,744	0,000	0,144	829
LIU_CDX2_TARGETS_UP	Diferenciación intestinal	29	0,738	0,000	0,193	385
SUMI_HNF4A_TARGETS	Diferenciación intestinal	29	0,655	0,004	0,169	1467
OHGUCHI_LIVER_HNF4A_TARGETS_DN	Diferenciación intestinal	117	0,605	0,000	0,129	1356
LUCAS_HNF4A_TARGETS_UP	Diferenciación intestinal	49	0,601	0,031	0,241	1688
SERVITJA_ISLET_HNF1A_TARGETS_DN	Diferenciación intestinal	93	0,515	0,002	0,153	1707
PID_HNF3BPATHWAY	Diferenciación intestinal	39	0,494	0,036	0,226	1550
SERVITJA_LIVER_HNF1A_TARGETS_DN	Diferenciación intestinal	127	0,454	0,004	0,207	2032
PID_HNF3APATHWAY	Diferenciación intestinal	40	0,421	0,019	0,241	1890
BAUS_TFF2_TARGETS_UP	Procesamiento antigénico	27	0,789	0,000	0,167	952
SHAFFER_IRF4_MULTIPLE_MYELOMA_PROGRAM	Procesamiento antigénico	33	0,625	0,000	0,155	1758
SHAFFER_IRF4_TARGETS_IN_ACTIVATED_DENDRITIC_CELL	Procesamiento antigénico	50	0,579	0,004	0,172	1793
SHAFFER_IRF4_TARGETS_IN_PLASMA_CELL_VS_MATURE_B_LYMPHOC	Procesamiento antigénico	54	0,572	0,017	0,170	3680

Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
SHAFFER_IRF4_TARGETS_IN_MYELOMA_VS_MATURE_B_LYMPHOCYTE	Procesamiento antigénico	88	0,537	0,002	0,167	3401
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	Procesamiento antigénico	67	0,515	0,004	0,169	4873
SHAFFER_IRF4_TARGETS_IN_ACTIVATED_B_LYMPHOCYTE	Procesamiento antigénico	68	0,508	0,025	0,240	3401
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	Procesamiento antigénico	95	0,436	0,004	0,159	5026
BIOCARTA_CASPASE_PATHWAY	Apoptosis	21	0,718	0,008	0,189	3416
BIOCARTA_MITOCHONDRIA_PATHWAY	Apoptosis	20	0,682	0,002	0,167	4809
SA_CASPASE_CASCADE	Apoptosis	18	0,630	0,013	0,240	3416
REACTOME_INTRINSIC_PATHWAY_FOR_APOPTOSIS	Apoptosis	29	0,613	0,006	0,172	5120
PID_CASPASE_PATHWAY	Apoptosis	47	0,566	0,006	0,169	3416
ALCALA_APOPTOSIS	Apoptosis	79	0,524	0,002	0,138	3103
KEGG_APOPTOSIS	Apoptosis	81	0,492	0,000	0,138	4882
CROONQUIST_NRAS_SIGNALING_UP	Oncogenes	35	0,588	0,010	0,242	3054
WANG_NEOPLASTIC_TRANSFORMATION_BY_CCND1_MYC	Oncogenes	20	0,570	0,002	0,191	284
DANG_MYC_TARGETS_DN	Oncogenes	28	0,519	0,019	0,204	7011
SWEET_KRAS_ONCOGENIC_SIGNATURE	Oncogenes	78	0,489	0,013	0,187	2171
PID_RAS_PATHWAY	Oncogenes	28	0,430	0,028	0,239	2515
ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_DN	Oncogenes	95	0,395	0,020	0,199	5253
BILD_HRAS_ONCOGENIC_SIGNATURE	Oncogenes	224	0,393	0,032	0,242	2504
MARZEC_IL2_SIGNALING_DN	Proliferación celular	33	0,565	0,016	0,228	2984
PLASARI_TGFB1_SIGNALING_VIA_NFIC_10HR_UP	Proliferación celular	46	0,487	0,004	0,243	1414
LABBE_TARGETS_OF_TGFB1_AND_WNT3A_DN	Proliferación celular	94	0,450	0,006	0,191	2919
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	Proliferación celular	66	0,439	0,038	0,245	2682
AMIT_EGF_RESPONSE_120_HELA	Proliferación celular	61	0,425	0,023	0,241	4300
PLASARI_TGFB1_TARGETS_10HR_UP	Proliferación celular	167	0,392	0,002	0,215	3994
GENTILE_UV_LOW_DOSE_UP	Respuesta a daño genómico	20	0,682	0,000	0,138	4283
DAZARD_UV_RESPONSE_CLUSTER_G24	Respuesta a daño genómico	23	0,671	0,000	0,152	312



Nombre	Procesos funcionales <sup>a</sup>	N <sup>b</sup>	ES <sup>c</sup>	p-valor nominal <sup>d</sup>	q-valor FDR <sup>e</sup>	Rank at max <sup>f</sup>
GENTILE_UV_HIGH_DOSE_UP	Respuesta a daño genómico	17	0,605	0,021	0,170	661
DAZARD_UV_RESPONSE_CLUSTER_G2	Respuesta a daño genómico	25	0,520	0,021	0,239	1286
WOTTON_RUNX_TARGETS_DN	Supresores tumorales	26	0,624	0,008	0,201	437
MCMURRAY_TP53_HRAS_COOPERATION_RESPONSE_UP	Supresores tumorales	22	0,533	0,022	0,241	633
KANNAN_TP53_TARGETS_UP	Supresores tumorales	48	0,439	0,019	0,242	1165
PID_P53DOWNSTREAMPATHWAY	Supresores tumorales	125	0,429	0,004	0,158	4936
CHEN_PDGF_TARGETS	Angiogénesis	15	0,585	0,013	0,175	1382
KEGG_VEGF_SIGNALING_PATHWAY	Angiogénesis	63	0,432	0,012	0,171	3718
PID_PDGFRRBPATWAY	Angiogénesis	121	0,418	0,030	0,235	5748
VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_DN	Cáncer Gástrico	129	0,614	0,004	0,171	2735
AUNG_GASTRIC_CANCER	Cáncer Gástrico	42	0,566	0,000	0,147	2619
VECCHI_GASTRIC_CANCER_EARLY_UP	Cáncer Gástrico	357	0,509	0,033	0,208	4205
PEDERSEN_METASTASIS_BY_ERBB2_ISOFORM_1	Invasión y metástasis	38	0,518	0,033	0,241	2276
ZUCCHI_METASTASIS_DN	Invasión y metástasis	39	0,505	0,039	0,236	4466
RICKMAN_METASTASIS_DN	Invasión y metástasis	227	0,379	0,021	0,169	3992
KEGG_DRUG_METABOLISM_OTHER_ENZYMES	Metabolismo de xenobióticos	30	0,673	0,007	0,162	1348
REACTOME_PHASE_II_CONJUGATION	Metabolismo de xenobióticos	45	0,647	0,015	0,202	2050
REACTOME_PHASE1_FUNCTIONALIZATION_OF_COMPOUNDS	Metabolismo de xenobióticos	53	0,515	0,012	0,237	3035
REACTOME_TERMINATION_OF_O_GLYCAN_BIOSYNTHESIS	Glicosilación aberrante	19	0,716	0,002	0,144	210
REACTOME_O_LINKED_GLYCOSYLATION_OF_MUCINS	Glicosilación aberrante	48	0,566	0,002	0,168	2594
REACTOME_SYNTHESIS_OF_BILE_ACIDS_AND_BILE_SALTS	Reflujo gastroesofágico	15	0,641	0,032	0,238	2114
REACTOME_BILE_ACID_AND_BILE_SALT_METABOLISM	Reflujo gastroesofágico	21	0,582	0,024	0,236	2296
MURATA_VIRULENCE_OF_H_PILORI	H.pylori	22	0,669	0,002	0,137	1081
GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_YELLOW_UP	Unfolded protein response	26	0,667	0,002	0,158	3074
REACTOME_PYRUVATE_METABOLISM	Fosforilación oxidativa	18	0,591	0,018	0,241	2867

<sup>a</sup> Mecanismo molecular, proceso fisiológico normal o patológico asignado subjetivamente al conjunto génico basado en su nombre. <sup>b</sup> Número de genes que componen la vía. <sup>c</sup> Las vías de señalización están ordenadas de mayor a menor valor de *Enrichment Score* (ES). <sup>d</sup> p-valor asociado a cada vía sin corregir por comparaciones múltiples. <sup>e</sup> p-valor asociado a cada vía resultante de la corrección para comparaciones múltiples FDR (tasa de falsos positivos). <sup>f</sup> Posición en la lista ranqueada en la cual se obtiene el mayor valor de ES, los conjuntos génicos más interesantes poseen *Rank at max* muy pequeño o muy grande.

## PUBLICACIÓN

En este apéndice se incluye la publicación obtenida como consecuencia directa del trabajo realizado en esta tesis:

**“Polymorphisms of *Helicobacter pylori* signaling pathway genes and gastric cancer risk in the European Prospective Investigation into Cancer-Eurgast cohort”.** *Int. J. Cancer*. 2014 Jan 1; 134(1):92-101. doi: 10.1002/ijc.28357.

**Autores:** Osmel Companioni, Catalina Bonet, Xavier Muñoz, Elisabete Weiderpass, Salvatore Panico, Rosario Tumino, Domenico Palli, Claudia Agnoli, Paolo Vineis, Marie-Christine Boutron-Ruault, Antoine Racine, Françoise Clavel-Chapelon, Ruth C. Travis, Kay-Tee Khaw, Elio Riboli, Neil Murphy, Anne-Claire Vergnaud, Antonia Trichopoulou, Vassiliki Benetou, Dimitrios Trichopoulos, Eiliv Lund, Dorthe Johansen, Bjorn Lindkvist, Mattias Johansson, Malin Sund, Eva Ardanaz, Emilio Sánchez-Cantalejo, Jose M. Huerta, Miren Dorronsoro, José Ramón Quiros, Anne Tjønneland, Lotte Maxild Mortensen, Kim Overvad, Jenny Chang-Claude, Cosmeri Rizzato, Heiner Boeing, Bueno-de-Mesquita H. Bas, Peter Siersema, Petra H.M. Peeters, Mattijs E. Numans, Fatima Carneiro, Ildir Licaj, Heinz Freisling, Nuria Sala, Carlos A. González.



## Polymorphisms of *Helicobacter pylori* signaling pathway genes and gastric cancer risk in the European Prospective Investigation into Cancer-Eurgast cohort

Osmel Compañoni<sup>1</sup>, Catalina Bonet<sup>1</sup>, Xavier Muñoz<sup>1,2</sup>, Elisabete Weiderpass<sup>3,4,5,6</sup>, Salvatore Panico<sup>7</sup>, Rosario Tumino<sup>8</sup>, Domenico Palli<sup>9</sup>, Claudia Agnoli<sup>10</sup>, Paolo Vineis<sup>11,12</sup>, Marie-Christine Boutron-Ruault<sup>13,14,15</sup>, Antoine Racine<sup>13,14,15</sup>, Françoise Clavel-Chapelon<sup>13,14,15</sup>, Ruth C. Travis<sup>16</sup>, Kay-Tee Khaw<sup>17</sup>, Elio Riboli<sup>18</sup>, Neil Murphy<sup>18</sup>, Anne-Claire Vergnaud<sup>19</sup>, Antonia Trichopoulos<sup>20,21</sup>, Vassiliki Benetou<sup>20</sup>, Dimitrios Trichopoulos<sup>21,22,23</sup>, Eiliv Lund<sup>24</sup>, Dorthe Johansen<sup>25</sup>, Björn Lindkvist<sup>26</sup>, Mattias Johansson<sup>27,28</sup>, Malin Sund<sup>29</sup>, Eva Ardanaz<sup>30,31</sup>, Emilio Sánchez-Cantalejo<sup>30,32</sup>, Jose M. Huerta<sup>33</sup>, Miren Dorronsoro<sup>30,34</sup>, José Ramón Quirós<sup>35</sup>, Anne Tjønneland<sup>36</sup>, Lotte Maxild Mortensen<sup>37</sup>, Kim Overvad<sup>37</sup>, Jenny Chang-Claude<sup>38</sup>, Cosmeri Rizzato<sup>39</sup>, Heiner Boeing<sup>40</sup>, H. Bas Bueno de Mesquita<sup>41,42</sup>, Peter Siersema<sup>42</sup>, Petra H.M. Peeters<sup>43</sup>, Mattijs E. Numans<sup>43</sup>, Fatima Carneiro<sup>44</sup>, Ildir Licaj<sup>27</sup>, Heinz Freisling<sup>27</sup>, Núria Sala<sup>1,2\*</sup> and Carlos A. González<sup>1</sup>

<sup>1</sup> Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, Barcelona, 08908, Spain

<sup>2</sup> Molecular Epidemiology Group, Translational Research Laboratory, Catalan Institute of Oncology-IDIBELL, Barcelona, 08908, Spain

<sup>3</sup> Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway

<sup>4</sup> Department of Research, Cancer Registry of Norway, Oslo, Norway

<sup>5</sup> Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

<sup>6</sup> Samfundet Folkhälsan, Helsinki, Finland

<sup>7</sup> Department of Clinical and Experimental Medicine, Federico II University Naples, Italy

<sup>8</sup> Cancer Registry and Histopathology Unit, "Civile - M.P. Arezzo" Hospital, ASP Ragusa, Italy

<sup>9</sup> Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute-ISP, Florence, Italy

<sup>10</sup> Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

<sup>11</sup> Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom

<sup>12</sup> Molecular and Genetic Epidemiology Unit, HuGeF, Human Genetics Foundation, Torino, Italy

<sup>13</sup> Insem, Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health Team, Villejuif, France

<sup>14</sup> Univ Paris Sud, UMR5 1018, F-94805, Villejuif, France

<sup>15</sup> IGR, F-94805, Villejuif, France

<sup>16</sup> Cancer Epidemiology Unit, University of Oxford, Oxford, UK

<sup>17</sup> Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

<sup>18</sup> Unit School of Public Health, Imperial College London, London, United Kingdom

**Key words:** gastric cancer, genetic susceptibility, *Helicobacter pylori*, NOD2, CD14

\*N.S. and C.A.G. contributed equally to this work

Grant sponsor: Fundació "LaCaixa"; Grant number: BM06-130-0; Grant sponsor: Health Research Fund of the Spanish Ministry of Health; Grant numbers: PI070130 and PI081420; Grant sponsor: European Commission FP5, QLGI-CT-2001-01049 ECNIS Network of Excellence of the 6th EU Framework Programme; Grant number: FOOD-CT-2005-513 943; Grant sponsor: Spanish Ministry of Health network RTICCC; Grant numbers: ISCIII RD06/0020/0091 and RD12/0036/0018; Grant sponsor: AGAUR, Generalitat de Catalunya; Grant number: 2009SGR939; Grant sponsor: Instituto de Salud Carlos III of the National Health System; Grant number: CA06/0200; Grant sponsors: "Europe Against Cancer" Programme of the European Commission (SANCO), Ligue contre le Cancer (France), Société 3M (France), Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM), German Cancer Aid, German Cancer Research Center, German Federal Ministry of Education and Research, Danish Cancer Society, the participating regional governments and institutions of Spain, Cancer Research UK, Medical Research Council, UK, the Stroke Association, UK, British Heart Foundation, Department of Health, UK, Food Standards Agency, UK, the Wellcome Trust, UK, Hellenic Ministry of Health and Social Solidarity, the Stavros Niarchos Foundation and the Hellenic Health; Italian Association for Research on Cancer, Italian National Research Council, Dutch Ministry of Public Health, Welfare and Sports, Dutch Ministry of Health, Dutch Prevention Funds, LK Research Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Swedish Cancer Society, Swedish Scientific Council, Regional Government of Skane, Sweden, the Norwegian Cancer Society, the Research Institute of Bellvitge, IDIBELL.

DOI: 10.1002/ijc.28357

History: Received 28 Feb 2013; Accepted 16 Apr 2013; Online 3 Jul 2013

Correspondence to: Carlos A. González, Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, Barcelona 08908, Spain, Tel.: 34-932607401, Fax: 34-932607787, E-mail: cagonzalez@iconcologia.net or Núria Sala, Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, Barcelona 08908, Spain, Tel.: 34-932607464, Fax: 34-932607466, E-mail: nsala@iconcologia.net

- <sup>19</sup>Department of Epidemiology and Biostatistics Medical Building, School of Public Health, Imperial College London, London, UK
- <sup>20</sup>Department of Hygiene, Epidemiology and Medical Statistics, WHO Collaborating Center for Food and Nutrition Policies, University of Athens Medical School, Athens, Greece
- <sup>21</sup>Hellenic Health Foundation, GR-115 27, Athens, Greece
- <sup>22</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA
- <sup>23</sup>Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece
- <sup>24</sup>Department of Community Medicine, University of Tromsø, Norway
- <sup>25</sup>Department of Surgery, Skåne University Hospital, SUS, Malmö
- <sup>26</sup>Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
- <sup>27</sup>International Agency for Research on Cancer (IARC/WHO), Lyon, France
- <sup>28</sup>Department of Biobank Research, Umeå University, Umeå, Sweden
- <sup>29</sup>Department of Surgery, Umeå University, 90185 Umeå, Sweden
- <sup>30</sup>CIBER Epidemiology and Public Health CIBERESP, Spain
- <sup>31</sup>Navarre Public Health Institute, Pamplona, Spain
- <sup>32</sup>Andalusian School of Public Health, Granada, Spain
- <sup>33</sup>Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain
- <sup>34</sup>Public Health Division of Gipuzkoa, BioDonostia Research Institute, Basque Regional Health Department, San Sebastian, Spain
- <sup>35</sup>Public Health Directorate, Asturias, Spain
- <sup>36</sup>Department of Cardiology, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark
- <sup>37</sup>Department of Public Health, Institute of Epidemiology and Social Medicine, Aarhus University, Aarhus, Denmark
- <sup>38</sup>Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany
- <sup>39</sup>Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
- <sup>40</sup>Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbrücke, Germany
- <sup>41</sup>National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
- <sup>42</sup>Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands
- <sup>43</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
- <sup>44</sup>Department of Pathology, IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, Medical Faculty of the University of Porto, Centro Hospitalar de Sao João, Porto, Portugal

*Helicobacter pylori* is a recognized causal factor of noncardia gastric cancer (GC). Lipopolysaccharide and peptidoglycan of this bacterium are recognized by CD14, TLR4 and NOD2 human proteins, while NFKB1 activates the transcription of pro-inflammatory cytokines to elicit an immune response. Single nucleotide polymorphisms (SNPs) in these genes have been associated with GC in different populations. We genotyped 30 SNPs of these genes, in 365 gastric adenocarcinomas and 1,284 matched controls from the European Prospective Investigation into Cancer cohort. The association with GC and its histological and anatomical subtypes was analyzed by logistic regression and corrected for multiple comparisons. Using a log-additive model, we found a significant association between SNPs in *CD14*, *NOD2* and *TLR4* with GC risk. However, after applying the multiple comparisons tests only the *NOD2* region remained significant ( $p = 0.009$ ). Analysis according to anatomical subtypes revealed *NOD2* and *NFKB1* SNPs associated with noncardia GC and *CD14* SNPs associated with cardia GC, while analysis according to histological subtypes showed that *CD14* was associated with intestinal but not diffuse GC. The multiple comparisons tests confirmed the association of *NOD2* with noncardia GC ( $p = 0.0003$ ) and *CD14* with cardia GC ( $p = 0.01$ ). Haplotype analysis was in agreement with single SNP results for *NOD2* and *CD14* genes. From these results, we conclude that genetic variation in *NOD2* associates with noncardia GC while variation in *CD14* is associated with cardia GC.

#### What's new?

Variations in immune genes appear to play an important role in determining susceptibility to gastric cancer linked to *Helicobacter pylori* colonization of gastric mucosa. However, little is known about the influence of variation on anatomical localization and histological subtype of this malignancy. The results of this study first confirm that *NOD2* and *CD14*, which encode proteins that recognize *H. pylori* lipopolysaccharide and peptidoglycan, are significantly associated with gastric cancer risk and second indicate that *NOD2* associates with noncardia and *CD14* with cardia gastric cancer. The differential effects of variation on the anatomical localization of disease warrant further investigation.

Despite its decreasing incidence, gastric cancer (GC) remains the fourth most common malignancy and second most common cause of cancer-related death in the world.<sup>1</sup> *Helicobacter pylori* (*H. pylori*) infection is the best established risk factor which induces chronic atrophic gastritis and hypochlorhydria states that can progress to noncardia GC.<sup>2</sup> However, only

between 0.1% and 4% of infected individuals develop this condition, underlining the relevance of host genetic and environmental predisposing cofactors that contribute to disease development.<sup>3</sup>

The innate immune system uses a variety of receptors to recognize bacterial antigens such as lipopolysaccharides (LPS) and peptidoglycans, allowing the induction of an immune response to the infection. CD14 (Monocyte Differentiation Antigen CD14) and TLR4 (Toll-Like Receptor 4) are involved in the recognition of LPS from Gram negative bacteria such as *H. pylori*.<sup>4</sup> LPS binding induces signal transduction through MyD88 and TRAF6 to activate the transcription factor NFκB1 (Nuclear factor kappa-b, Subunit 1). Moreover, NOD2 (Nucleotide-binding oligomerization domain Protein 2) recognizes the peptidoglycan of *H. pylori*, and this also induces signal transduction to activate NFκB1. This transcription factor translocates into the nucleus to induce the expression of pro-inflammatory cytokines, chemokines and cell adhesion molecules, responsible for inflammation as a component of the immune response against *H. pylori*.<sup>3-5</sup>

Single nucleotide polymorphisms (SNPs) in the above genes have been associated with GC in several populations across the world. *CD14* rs2569190 TT genotype was associated with GC in three Asian studies but not in two Caucasian populations.<sup>6-9</sup> An association with *NOD2* rs2066842 (P268S) was described in Polish and German populations.<sup>10,11</sup> *NOD2* rs2066844 (R702W) and rs5743293 have been associated with GC in Italians and with several gastrointestinal diseases (including GC) in a Danish study.<sup>12,13</sup> *TLR4* rs4986790 was associated with noncardia GC in Polish, U.S. and Brazilian populations, and rs4986791 was associated with the intestinal subtype in Italians but neither of these *TLR4* SNPs were found associated with GC in Mexicans and Japanese.<sup>14-18</sup> With respect to *NFKB1*, an insertion/deletion variant (-94 ins/del ATTG) was associated with GC in a Chinese cohort.<sup>19</sup> In general, these studies show discrepant results and, as far as we know, none of them analyzes the association of common polymorphisms in these genes with both different anatomical locations as well as histological subtypes of gastric adenocarcinoma, in the same sample groups. The aim of our study was to analyze the association of selected SNPs in the *CD14*, *TLR4*, *NOD2* and *NFKB1* genes with GC and its subtypes in the European Prospective Investigation into Cancer (EPIC-Eurgast) population.

## Material and Methods

### Subjects

The study design is a nested case control study in the EPIC cohort, which was described previously.<sup>20</sup> Briefly, this cohort includes healthy subjects recruited between 1992 and 1998 in 10 European countries. GC cases were all subjects having blood collected and diagnosed during the follow-up with adenocarcinoma of stomach, defined by code C16 of the International Classification of Diseases, 10th Revision. An

independent panel of pathologists confirmed and validated the diagnosis.<sup>21</sup>

For each case, up to four controls were randomly selected among cohort members alive and free of cancer at the time of diagnosis of the cases, matched by center, sex, age and date of blood collection. Signed informed consent was available from all individuals. The study was approved by ethical committees at IARC and in each of the EPIC recruitment centers. Samples were excluded from the analyses if DNA was not amplified or had a genotyping call rate < 80%. The final sample size was 365 GC cases and 1,284 controls (Table 1).

### SNP selection criteria

For each candidate gene, we analyzed the HapMap information for Caucasians (Phase II CEU population, Releases 23a or 24) and compiled a list of SNPs located from a minimum of 10 kb upstream to at least 3 kb downstream of each gene. Haplotype blocks were defined according to Gabriel *et al.*<sup>22</sup> using the default parameters in Haploview v4.0 software.<sup>23</sup> Haplotype Tagging SNPs (tagSNPs) were selected by use of the Tagger algorithm, as implemented in Haploview. Parameters used for tagSNP selection were minor allele frequency (MAF)  $\geq 5\%$  in Caucasians,  $r^2 \geq 0.8$  between each pair of tagged and tagSNPs, and SNPs tagging haplotypes with frequency  $\geq 0.05$ . Following these criteria, we selected six tagSNPs that defined two haplotype blocks in the *CD14* gene region, nine tagSNPs (two blocks) in *NOD2*, six tagSNPs (one block) in *TLR4* and 14 tagSNPs (one block) in *NFKB1*. Some selected tagSNPs were also potentially functional SNPs retrieved from the SNP500 database and literature searches. Two functional SNPs in *TLR4*, not in complete linkage disequilibrium (LD) with the haploview selected tagSNPs, were also included. One SNP selected in *TLR4* (rs1252037) and three in *NOD2* (rs2066842, rs5743291 and rs718226) were excluded from analysis due to poor cluster separation in the genotyping process. *NOD2* rs12324931 and *NFKB1* (rs4647992 and rs4648144) were also excluded from the analysis due to their low MAFs in controls (0.036, 0.050 and 0.001). Therefore, 30 SNPs in four gene regions were finally genotyped and analyzed (Table 2).

### DNA extraction and genotyping

Genotyping was described in details elsewhere.<sup>24</sup> Briefly, genomic DNA was extracted from buffy coat and genotyped at the Spanish National Genotyping Center (CEGEN) by Illumina BeadStation Platform and GoldenGate technology (Illumina, San Diego, CA), according to the manufacturer's protocols. Internal and duplicate samples (5%) were included as controls to assess genotyping error rate.

### *H. pylori* assays

Quantification of *H. pylori* and CagA antibodies in plasma was done by ELISA<sup>25</sup> or immunoblot using the HELICOBLT

Table 1. Baseline main characteristics of the analyzed population

Variables		Controls (N = 1,284), N (%)	Cases (N = 365), N (%)
Sex	Male	759 (59.1)	214 (58.6)
	Female	525 (40.9)	151 (41.4)
Age at recruitment (years)	Mean $\pm$ SD	58.4 $\pm$ 7.69	58.4 $\pm$ 7.93
Country	France	3 (0.23)	2 (0.55)
	Italy	206 (16.0)	56 (15.3)
	Spain	134 (10.4)	41 (11.2)
	United Kingdom	135 (10.5)	41 (11.2)
	The Netherlands	99 (7.71)	26 (7.12)
	Greece	88 (6.85)	24 (6.58)
	Germany	186 (14.5)	48 (13.2)
	Sweden	220 (17.1)	64 (17.5)
	Denmark	205 (16.0)	61 (16.7)
	Norway	8 (0.62)	2 (0.55)
	<i>H. pylori</i> infection	No	492 (38.3)
Yes		779 (60.7)	294 (80.5)
Missings		13 (1)	3 (0.9)
cagA in <i>H. pylori</i> infected controls and noncardia cases <sup>1</sup>	+	584 (74.97)	151 (92.64)
	-	195 (25.03)	12 (7.36)
Anatomical subtype of GC	Cardia	-	107 (29.3)
	Noncardia	-	181 (49.6)
	Mixed	-	6 (1.64)
	Unknown	-	71 (19.5)
Histological subtype of GC	Intestinal	-	126 (34.5)
	Diffuse	-	128 (35.1)
	Mixed	-	8 (2.19)
	Unknown	-	103 (28.2)

<sup>1</sup>The (%) of cagA status (+ or -) was calculated with respect to 163 noncardia cases and 779 controls *H. pylori* infected.

\*Indicates *p*-value < 0.05 for the comparison between cases and controls.

2.1.<sup>26</sup> Infection status was considered positive when one of them was positive.

#### Statistical and bioinformatic analyses

To test genetic association between the selected SNPs and GC in general and with its anatomical (cardia/noncardia) and histological (intestinal/diffuse) subtypes, adjusted odds ratios (OR, 95% CI) were estimated using unconditional logistic regression, adjusting by the matching variables age, sex and country. In a first analysis, ORs were estimated under the log-additive (per allele) model. In subsequent analysis, other genetic models (dominant, recessive and codominant) were estimated. As a rule, all results presented in the text and tables are based in the log-additive model. When other models were used, it is indicated. Due to their low MAF and absence of variant homozygotes, *TLR4* rs4986790 and rs4986791, were analyzed under the codominant model. In all analyses, the most frequent allele/genotype was set as the

reference category. A heterogeneity test was used to estimate a significant differential effect between GC subtypes under the log-additive model. To correct for multiple testing, a gene-based permutation test (1,000 permutations) was performed.<sup>27</sup> The permutation *p*-value of each gene was then corrected by the number of analyzed genes (*n* = 4 in our study) by Bonferroni. Therefore, significance after Bonferroni correction is set at 0.0125.

As noncardia GC is the only anatomical localization associated with *H. pylori* infection,<sup>28</sup> we also analyzed whether the presence of cagA antibodies (cagA+/cagA-) modifies the genotype effect. To do this, in those cases and controls *H. pylori* infected, we calculate ORs for noncardia GC, comparing cagA+ vs. those cagA-.

Associations between GC, its anatomical location and histological subtype with haplotypes (frequency > 0.05) were assessed using the most frequent haplotype among the whole population as the reference group. Haplotype frequencies

Table 2. Characteristics of analyzed SNPs and association with gastric cancer risk

Gene	dbSNP number <sup>1</sup>	SNP location <sup>2</sup>	Tag/functional SNP	Genotypes <sup>3</sup>	Genotype counts (controls, 1,284)	Genotype counts (cases, 365)	OR (95% CI) <sup>4</sup>	p-value <sup>4</sup>	Permutation test p-value <sup>5</sup>
<b>CD14</b> <sup>6</sup>	rs11167532	Flanking 5' UTR	tagSNP	GG/GA/AA	382/582/198	112/158/54	0.97 (0.81–1.16)	0.720	0.188
	rs7711117	Flanking 5' UTR	tagSNP	CC/CT/TT	770/439/67	203/142/19	1.14 (0.94–1.38)	0.202	
	rs778588	Flanking 3' UTR	tagSNP	TT/TC/CC	667/526/89	189/143/32	1.05 (0.87–1.26)	0.614	
	rs4914	Coding, L367L	tagSNP and Functional	CC/CG/GG	957/261/13	283/62/7	0.91 (0.69–1.19)	0.479	
	rs2569190 (-159C/T)	5' UTR	tagSNP and Functional	GG/GA/AA	307/621/264	103/173/76	0.92 (0.77–1.09)	0.341	
	rs1583005	IK, intron	tagSNP	CC/CT/TT	476/600/205	115/179/71	1.20 (1.02–1.42)	0.032	
<b>NOD2</b>	rs7202124	SNX20, intron	tagSNP	AA/AG/GG	675/492/113	220/126/18	0.74 (0.61–0.89)	0.002	0.009
	rs2111235	Intron	tagSNP and Functional	CC/CT/TT	631/507/134	205/129/26	0.77 (0.64–0.93)	0.005	
	rs5743289	Intron	tagSNP and Functional	CC/CT/TT	895/348/41	251/101/13	1.05 (0.85–1.31)	0.634	
	rs3135500	3' UTR	tagSNP	GG/GA/AA	472/598/214	121/174/70	1.13 (0.96–1.33)	0.150	
	rs3785142	CYLD, Intron	tagSNP	TT/TC/CC	343/614/317	115/156/92	0.92 (0.78–1.08)	0.324	
<b>TLR4</b>	rs1329061	Flanking 5' UTR	tagSNP	TT/TC/CC	572/580/129	150/166/49	1.16 (0.98–1.39)	0.089	0.240
	rs1329060	Flanking 5' UTR	tagSNP	GG/GA/AA	948/309/26	262/91/12	1.13 (0.90–1.42)	0.302	
	rs1329057	Flanking 5' UTR	tagSNP	AA/AG/GG	825/410/48	222/126/17	1.13 (0.92–1.39)	0.234	
	rs10491851	Flanking 5' UTR	tagSNP	GG/GT/TT	678/516/87	217/126/21	0.81 (0.67–0.99)	0.037	
	rs4986790 (Asp299Gly)	Coding, D299G	Functional	AA/AG/GG	1,134/133/3	316/45/0	1.22 (0.85–1.75)	0.273	
	rs4986791 (Thr399Ile)	Coding, T399I	Functional	CC/CT/TT	1,124/134/5	309/45/0	1.23 (0.86–1.77)	0.156	
	rs11536889	3' UTR	tagSNP and Functional	GG/GC/CC	940/308/35	258/98/9	1.09 (0.87–1.36)	0.457	
<b>NFKB1</b>	rs2085549	Flanking 5' UTR	tagSNP	TT/TC/CC	754/450/79	212/133/19	1.00 (0.82–1.21)	0.965	0.577
	rs2085548	Flanking 5' UTR	tagSNP	GG/GA/AA	642/531/111	182/153/30	1.00 (0.83–1.20)	1.000	
	rs747559	Flanking 5' UTR	tagSNP	AA/AG/GG	471/610/177	128/178/51	1.04 (0.87–1.24)	0.669	
	rs3774934	Intron	tagSNP and Functional	GG/GA/AA	1,045/225/14	296/65/3	0.99 (0.75–1.31)	0.972	
	rs1585214	Intron	tagSNP	GG/GA/AA	387/635/250	119/182/61	0.89 (0.75–1.05)	0.178	
	rs4648022	Intron	tagSNP and Functional	CC/CT/TT	1,068/206/8	298/59/8	1.23 (0.93–1.62)	0.146	
	rs4648055	Intron	tagSNP	GG/GA/AA	656/511/117	183/155/27	0.99 (0.82–1.18)	0.896	
	rs9790601	Intron	tagSNP	AA/AG/GG	624/532/125	170/161/32	1.03 (0.86–1.23)	0.747	
	rs1609798	Intron	tagSNP	CC/CT/TT	628/528/124	177/153/35	1.02 (0.85–1.22)	0.830	
	rs7677509	Flanking 3' UTR	tagSNP	CC/CT/TT	323/627/327	105/174/83	0.88 (0.74–1.03)	0.113	
	rs765789	MANBA, intron	tagSNP	TT/TA/AA	845/395/42	226/117/21	1.22 (0.99–1.49)	0.062	
rs228611	MANBA, intron	tagSNP	GG/GA/AA	341/639/297	114/175/75	0.86 (0.73–1.01)	0.073		

<sup>1</sup>All SNPs in each genetic region are oriented in the 5'–3' direction of the sense strand.

<sup>2</sup>Due to tagSNP selection criteria, some SNPs are located in genes flanking the gene region being analyzed.

<sup>3</sup>Genotypes are reported as common allele homozygotes/heterozygotes/minor allele homozygotes.

<sup>4</sup>OR (95% CI) and p-values are reported according to the log-additive model adjusted for country, sex and age. In the case of rs4986790 and rs4986791, the codominant model was estimated due to low risk allele frequency.

<sup>5</sup>p-values obtained from the gene-based permutation test for gastric cancer. Significance after Bonferroni correction is set at 0.0125.

<sup>6</sup>CD14 gene runs in the antisense DNA strand.

were inferred using the Expectation Maximization Algorithm as implemented in the haplo.stats R package. The statistical analyses were performed using SNPpass library of R software<sup>29</sup> and considered significant when  $p < 0.05$ .

We tested possible functional consequences of SNPs (intronic, 5'/3' UTR, synonymous) associated with GC by use of the "Variant Effect Predictor"<sup>30</sup> web tool at Ensembl Genome Browser (<http://www.ensembl.org>). We also used the PupaSuite 3.1 software<sup>31</sup> (<http://pupasuite.bioinfo.cipf.es/>) which searches for SNPs located within coding sequences (nonsynonymous SNPs), transcription factor binding sites (TFBS), Exonic Splicing Enhancer (ESE), Exonic Splicing Silencer (ESS), Splice sites, miRNAs and Mouse/Human conserved regions.

## Results

Baseline main characteristics of GC cases and matched controls are described in Table 1. Description of analyzed SNPs is presented in Table 2. All genotyped SNPs were in Hardy-Weinberg equilibrium in controls. We found a positive association between GC risk and *CD14* (rs1583005) and an inverse associations with *NOD2* (rs7202124 and rs2111235) and *TLR4* (rs10491851) (Table 2). The associations with *NOD2* rs7202124 (OR 0.74; 95% CI 0.61-0.89,  $p = 0.002$ ) and rs2111235 (OR 0.77; 95% CI 0.64-0.93,  $p = 0.005$ ) were the ones showing the highest significance and after correcting for multiple comparisons only the *NOD2* region remained significant (Table 2).

When cases were grouped according to tumor localization (Table 3), four SNPs of *CD14* were associated with cardia GC, with the heterogeneity test being significant for rs11167532 and rs2569190. *NFKB1* rs2085549 was additionally inversely associated with cardia GC under the recessive model (data not shown). After correcting for multiple comparisons, only the *CD14* region remained significantly associated with cardia GC (Table 3). The inverse association of *NOD2* SNPs rs7202124 and rs2111235 with GC remained highly significant only for noncardia GC, with a significant heterogeneity test for rs2111235. In addition, rs765789 and rs228611 of *NFKB1* were also associated with noncardia GC under the log-additive model and *TLR4* rs1329061 was positively associated under the recessive model (data not shown). *NOD2* was the only genetic region that remained significantly associated with noncardia GC after the multiple comparisons test was applied (Table 3). Because *H. pylori* infection is a risk factor only for noncardia GC, we stratified noncardia GC cases and their controls by presence of *cagA* antibodies (Table 4). *NOD2* rs2111235 remained significantly associated with noncardia GC only in the *cagA*+ individuals. Association with rs7202124 was significant in both *cagA*+ and *cagA*- individuals and rs5743289 was significant only in the *cagA*- subgroup, but the sample size of cases in this group was very low. Other SNPs associated in *cagA*+ were rs228611 (dominant) and rs1329061 (recessive).

Regarding the association with histological subtypes (Table 5), *NOD2* rs7202124 was negatively and significantly associated

with both histological subtypes, while rs2111235 was only associated with the intestinal subtype but the heterogeneity test was not significant. By contrast, *CD14* rs778588 was also found only associated with the intestinal subtype with a statistically significant heterogeneity test. Several SNPs of *NFKB1* were inversely (rs1585214, rs228611 and rs7677509) and positively (rs765789) associated with the diffuse subtype, but none showed a significantly different effect between both subtypes, as assessed by the heterogeneity test. Other associations observed were *CD14* rs2569190, which was found inversely associated with the intestinal-type of GC under the dominant genetic model and *NFKB1* rs228611, which was inversely associated with the diffuse subtype under the recessive model (data not shown). None of these associations was confirmed after adjustment for multiple comparisons (Table 5).

Haplotype analysis also revealed several significant associations (Table 6). The most significant result was observed with haplotype GCCCGT of *CD14* as a risk factor for the cardia and intestinal subtypes. This haplotype carries the risk alleles of rs778588 (cardia and intestinal-types) and rs1583005 (cardia), as well as the common alleles of rs11167532 and rs2569190, whose variant alleles were found inversely associated with cardia GC. The other remarkable finding was the *NOD2* haplotype GTCGC, which appeared as a protective factor for GC and its noncardia and intestinal subtypes. This haplotype carries the G allele of rs7202124 and the T allele of rs2111235, which in the single gene analysis were also found inversely associated with GC and its noncardia and intestinal subtypes.

When we tested the phenotypic effect of the associated SNPs using the "Variant Effect Predictor" tool, we found that the most significant intronic *NOD2* SNP rs2111235 was placed in a regulatory region of the gene. Also, rs2569190 at 5'-UTR of *CD14*, is located in a regulatory region of the gene. We also explored functional consequences of associated SNPs using the PupaSuite 3.1 software. The *CD14* rs2569190 (underlined in the 971-981 nucleotides reverse strand sequence, TCCCCCTCCCT) and *NFKB1* rs1585214 (underlined in the 22,640-22,652 nucleotides, forward strand sequence, CCCTTTCCTTTT) were predicted as located in DNA triplex sequences larger than 10 polypurines or polypirimidines, which can disturb the normal regulation of a particular gene and affect its normal expression.<sup>31</sup> Pupasuite also showed that *NOD2* (rs2111235 and rs5743289), *TLR4* rs10491851 and *NFKB1* (rs765789 and rs228611) were localized in Mouse/Human conserved regions.

## Discussion

In our study, we analyzed the association between genes belonging to the *H. pylori* signaling pathway and the risk of GC and its subtypes in the EPIC-Eurgast study. This analysis indicates that *NOD2* was associated with noncardia GC while *CD14* was associated with cardia GC.

The most significant observed associations were with *NOD2* rs7202124 and rs2111235 variants, which were found



Table 3. Odds ratios (95% CI) adjusted by country, sex and age, for SNPs significantly associated with cardia or noncardia gastric cancer

Gene	SNP	Cardia (N = 107)				Noncardia (N = 181)				
		Genotype counts in cases <sup>1</sup>	OR (95% CI) <sup>2</sup>	p-value <sup>2</sup>	Permutation test p-value <sup>3</sup>	Genotype counts in cases <sup>1</sup>	OR (95% CI) <sup>2</sup>	p-value <sup>2</sup>	Permutation test p-value <sup>3</sup>	p-value of heterogeneity
CD14	rs11167532	40/46/10	0.69 (0.49–0.95)	0.023	0.010	51/74/33	1.13 (0.89–1.44)	0.328	0.907	0.007
	rs778588	44/52/11	1.50 (1.09–2.05)	0.013		94/70/16	1.02 (0.79–1.31)	0.881		0.060
	rs2569190	36/55/15	0.69 (0.51–0.93)	0.013		51/79/42	0.99 (0.78–1.25)	0.939		0.025
	rs1583005	25/55/27	1.61 (1.20–2.14)	0.001		65/83/33	1.05 (0.84–1.32)	0.645		0.056
NOD2	rs7202124	58/40/9	0.92 (0.67–1.26)	0.595	0.786	117/58/5	0.60 (0.46–0.79)	0.00019	0.0003	0.090
	rs2111235	53/39/13	0.99 (0.74–1.34)	0.961		115/56/8	0.58 (0.44–0.76)	3.664e–05		0.021
NFKB1	rs765789	71/31/5	1.09 (0.76–1.56)	0.655	0.617	103/66/11	1.42 (1.09–1.85)	0.012	0.131	0.154
	rs228611	34/47/26	0.93 (0.70–1.24)	0.635		61/85/35	0.79 (0.63–0.99)	0.041		0.243

<sup>1</sup>Genotype counts are reported as common allele homozygotes/heterozygotes/minor allele homozygotes, as stated in Table 2.

<sup>2</sup>OR (95% CI) and p-values are reported according to the log-additive model adjusted for country, sex and age.

<sup>3</sup>p-values from the gene-based permutation test for cardia and noncardia gastric cancer localizations. Significance after Bonferroni correction is set at 0.0125.

Table 4. Odds ratios (95% confidence intervals) adjusted by country, sex and age, for SNPs significantly associated with noncardia gastric cancer in *H. pylori* infected individuals, according to *cagA* status<sup>1</sup>

Gene	SNP	<i>H. pylori</i> infected and <i>cagA</i> + (controls N = 584, noncardia cases N = 151)				<i>H. pylori</i> infected and <i>cagA</i> - (controls N = 195, noncardia cases N = 12)			
		Genotype counts controls <sup>2</sup>	Genotype counts cases <sup>2</sup>	OR (95% CI) <sup>3</sup>	p-value <sup>3</sup>	Genotype counts controls <sup>2</sup>	Genotype counts cases <sup>2</sup>	OR (95% CI) <sup>3</sup>	p-value <sup>3</sup>
NOD2	rs7202124	313/225/45	97/48/5	0.66 (0.47–0.91)	0.009	92/77/25	9/3/0	0.30 (0.08–1.08)	0.035
	rs2111235	295/224/60	99/43/7	0.57 (0.42–0.79)	0.0003	86/83/23	7/4/1	0.68 (0.25–1.87)	0.446
	rs5743289	409/54/21	99/44/8	1.23 (0.90–1.70)	0.204	148/43/4	4/7/1	3.76 (1.33–10.63)	0.012

<sup>1</sup>The *cagA* status (+ or –) was calculated with respect to all 163 noncardia GC cases and 779 controls *H. pylori* infected.

<sup>2</sup>Genotype counts are reported as common allele homozygotes/heterozygotes/minor allele homozygotes, as stated in Table 2.

<sup>3</sup>OR (95%CI) and p-values are reported according to the log-additive model adjusted for country, sex and age.

inversely associated with noncardia GC and both histological subtypes. The haplotype analysis corroborates the single variants results, being the haplotype that carries these two variants also inversely associated with GC, noncardia and intestinal types. As far as we know, there are not previous studies on *NOD2* rs7202124 and rs2111235 and GC, maybe because previous studies were not based in a haplotype tagging SNP strategy. However, other *NOD2* SNPs have been associated with GC. For example, rs2066842 (P268S) has been reported to be a risk variant for GC in Polish and German populations, but we could not analyze this SNP due to poor cluster separation of its genotypes.<sup>10,11</sup> Moreover,

rs3135500 has been reported to increase the risk for GC in the Chinese population, a finding not confirmed by us.<sup>32</sup> Nevertheless, according to the haplotype analysis of CEU HapMap population, the haplotype inversely associated with GC in our study carries the wild-type or major variants of these two polymorphisms. In an Italian population, rs2066844 and rs5743293 were also reported to increase susceptibility to GC, but these SNPs are not included in the HapMap version we used.<sup>12</sup>

*NOD2* is a member of the nucleotide-binding oligomerization domain-like receptor family and exists both as a cytosolic and an extracellular receptor, mostly expressed in

Table 5. Odds ratios (95% CI) adjusted by country, sex and age, for SNPs significantly associated with intestinal or diffuse histological subtypes of gastric cancer

Gene	SNP	Intestinal (N = 126)			Diffuse (N = 128)			p-value of heterogeneity		
		Genotype counts in cases <sup>1</sup>	OR (95% CI) <sup>2</sup>	p-value <sup>2</sup>	Permutation test p-value <sup>3</sup>	Genotype counts in cases <sup>1</sup>	OR (95% CI) <sup>2</sup>		p-value <sup>2</sup>	Permutation test p-value <sup>3</sup>
CD14	rs778588	52/59/15	1.49 (1.12–1.97)	0.006	0.041	76/41/10	0.80 (0.59–1.09)	0.151	0.652	0.010
NOD2	rs7202124	76/43/7	0.73 (0.54–0.99)	0.040	0.077	78/44/5	0.72 (0.53–0.99)	0.037	0.204	0.922
	rs2111235	74/44/7	0.68 (0.50–0.93)	0.011		69/50/8	0.83 (0.62–1.10)	0.191		0.232
NFKB1	rs1585214	38/66/21	0.94 (0.72–1.23)	0.665	0.727	46/65/16	0.76 (0.58–0.99)	0.044	0.213	0.129
	rs228611	40/53/32	0.93 (0.72–1.21)	0.605		39/74/15	0.74 (0.56–0.96)	0.024		0.144
	rs765789	75/44/7	1.31 (0.96–1.80)	0.096		73/43/11	1.46 (1.07–1.99)	0.018		0.280
	rs7677509	36/54/35	0.97 (0.75–1.26)	0.819		36/72/19	0.77 (0.59–1.00)	0.047		0.107

<sup>1</sup>Genotype counts are reported as common allele homozygotes/heterozygotes/minor allele homozygotes, as stated in table 2.

<sup>2</sup>OR (95% CI) and p-values are reported according to the log-additive model adjusted for country, sex and age.

<sup>3</sup>Significant p-values from the gene-based permutation test for intestinal and diffuse gastric cancer histological subtypes. Significance after Bonferroni correction is set at 0.0125.

Table 6. Haplotype analysis

Gene	Haplotype	Haplotype frequency in controls	Haplotype frequency in cases	OR (CI 95%) <sup>1</sup>	p-value	Anatomical location or histological subtypes
CD14	ACTCAC	0.413	–	1.00 (Reference <sup>2</sup> )	–	–
	GCCCGT	0.128	0.212	2.15 (1.41–3.27)	0.0004	Cardia
	GCCCGT	0.128	0.193	1.71 (1.17–2.49)	0.006	Intestinal
NOD2	ACCAT	0.355	–	1.00 (Reference <sup>2</sup> )	–	–
	GTCGC	0.198	0.154	0.72 (0.56–0.91)	0.007	Gastric cancer
	GTCGC	0.198	0.135	0.60 (0.43–0.86)	0.005	Noncardia
	GTCGC	0.198	0.141	0.64 (0.43–0.96)	0.031	Intestinal
TLR4	TGATACG	0.267	–	1.00 (Reference <sup>2</sup> )	–	–
	CAGGACG	0.135	0.153	1.33 (1.01–1.75)	0.040	Gastric cancer
	CGAGACG	0.074	0.112	1.98 (1.14–3.47)	0.016	Cardia
NFKB1	TGAGACGACTTA	0.328	–	1.00 (Reference <sup>2</sup> )	–	–
	TGAGGTGACCAG	0.066	0.095	1.78 (1.15–2.73)	0.009	Noncardia

<sup>1</sup>Odds ratios (95% CI) adjusted by country, sex and age, for haplotypes significantly associated with gastric cancer and/or any of its anatomical locations or histological subtypes. Estimates are based on the log-additive model.

<sup>2</sup>Reference is the most common haplotype.

The SNP order in CD14 haplotypes is rs11167532, rs7711117, rs778588, rs4914, rs2569190 and rs1583005. For NOD2 is rs7202124, rs2111235, rs5743289, rs3135500 and rs3785142. For TLR4 is rs1329061, rs1329060, rs1329057, rs10491851, rs4986790, rs4986791 and rs11536889. For NFKB1 is rs2085549, rs2085548, rs747559, rs3774934, rs1585214, rs4648022, rs4648055, rs9790601, rs1609798, rs7677509, rs765789 and rs228611.

monocytes/macrophages. NOD2 is over-expressed in the gastric mucosa of *H. pylori* infected patients.<sup>33</sup> It plays a role in the immune response to *H. pylori* by recognizing muramyl dipeptide, a degradation product of peptidoglycan from bac-

terial wall. This role of NOD2 as a sensor of *H. pylori* infection is in agreement with NOD2 association with noncardia GC which, in contrast to cardia GC, is strongly associated with *H. pylori* infection.<sup>28</sup> After ligand sensing, NOD2

activates the transcription factor NFKB1 through the RIP2/IKK pathway leading to expression of proinflammatory cytokines.<sup>34</sup>

With respect to *CD14*, we found that four SNPs (rs11167532, rs778588, rs2569190 and rs1583005) and a haplotype that carries the risk allele of each of these SNPs are significantly associated with GC of the cardia. The multiple comparisons tests confirmed the association of *CD14* with cardia GC. Interestingly, rs2569190 was located in a DNA triplex region that might affect gene expression. One of these SNPs (rs778588) and the same haplotype were also associated to the intestinal subtype of GC, however, this result was not significant after correcting for multiple comparisons tests.

The published studies of *CD14* have been focused on rs2569190, which is placed near the Sp1 transcription factor binding site and relevant for *CD14* expression.<sup>35</sup> In concordance with our study, one study in Japan<sup>8</sup> found an inverse association with intestinal subtype and studies in Poland,<sup>9</sup> USA<sup>9</sup> and Taiwan populations<sup>36</sup> did not find an association with noncardia GC. Unlike our findings, a Chinese study reported a positive association with GC.<sup>7</sup> As far as we know, no study has yet been performed regarding the association of *CD14* with cardia GC.

*CD14* exists both as a membrane bound (mCD14) and a soluble protein, the former is mostly expressed in monocytes/macrophages, which are accumulated in *H. pylori* infected mucosa.<sup>37</sup> mCD14 interacts with LPS, a cell wall component of *H. pylori*, and it is transferred to TLR4-MD2 complex. This signal is transduced through MyD88, IRAK1, TRAF6 and IKK to activate NFKB1 and transcription of inflammatory cytokines.<sup>34</sup> The persistent expression of these cytokines and environmental influences causes a chronic inflammation that leads to preneoplastic changes of gastric mucosa and, finally, to GC.<sup>38</sup> In the *CD14* LD region, rs1583005 (associated with GC and cardia GC in our study) is localized at IK gene. Interestingly, this gene is a downregulator of HLA-DR II expression on hematopoietic cells,<sup>39</sup> and it is well recognized that loss of HLA expression is an escape mechanism of immune recognition employed by several cancers.<sup>40</sup> As *H. pylori* infection is not associated with cardia GC,<sup>28</sup> it is plausible that other mechanisms different from *H. pylori* recognition could explain the association of *CD14* with cardia GC. In this sense, *CD14* could act as receptor for Epstein-Barr virus,<sup>41,42</sup> a known carcinogenic infection associated with GC of cardia localization.<sup>43,44</sup> Unfortunately, we do not have data regarding Epstein-Barr virus infection in our population.

We found associations with some *TLR4* and *NFKB1* SNPs and haplotypes; however, these genes were no longer significantly associated with GC or with any of its subtypes after the multiple correction tests were applied (Tables 3 and 5). Most of the previous studies on *TLR4* and GC are focused on two nonsynonymous SNPs, rs4986790 and rs4986791, whose low frequency variant alleles (MAF 0.033 for both in CEU HapMap) are hyporesponsive to LPS and as a consequence increase the risk of developing chronic inflammation

and cancer.<sup>14</sup> rs4986790 has been found to be associated with GC in Brazilians<sup>15</sup> and its G allele was positively associated with noncardia GC in Poland and U.S. populations,<sup>14</sup> but the association with this localization was not confirmed in Mexicans.<sup>45</sup> The rs4986791 SNP has also been positively associated with intestinal type GC in Italy.<sup>16</sup> The frequencies of risk alleles were very small (0.057 and 0.061 in controls) in our population, and we could only estimate the association under the codominant model, without any significance. This finding is in agreement with studies performed in Japan<sup>18</sup> and Mexican populations.<sup>17,45</sup>

The *NFKB1* rs228611 SNP was found to be inversely associated with noncardia GC and with the diffuse subtype under the recessive model, but none of these associations was significant after the multiple comparisons tests. Two previous studies have reported a positive<sup>19</sup> and a lack of association<sup>46</sup> of polymorphism (-94 ins/del ATTTG) in the promoter of *NFKB1* with GC. Also, rs4648068 has been associated with GC risk in Chinese.<sup>47</sup>

When we stratified noncardia GC by *cagA* status, the *NOD2* SNPs (rs7202124 and rs2111235) remained significantly associated with noncardia GC in the *cagA*+ individuals. However, the number of cases *cagA*- was too low to draw any conclusion.

Our study has several advantages. We explored the SNPs association with both anatomical localizations and histological types of GC, one aspect not always considered by previous studies. The EPIC cohort is constituted by a large sample size of subjects who have been extensively followed for the onset of cancers and characterized for life-style risk factors and *H. pylori* infection. The diagnosis of GC cases was confirmed by a panel of pathologists.<sup>21</sup> Limitations of our study are that the number of GC cases is relatively small when subgrouping by histological subtypes and anatomical localizations of GC with the consequent decrease in statistical power to detect an effect on these subtypes. Also, we did not genotype SNPs in other host genes (*MD2*, *TLR2/5*) that may interact with *H. pylori* during the infection and the genetic coverage of *TLR4*, *NOD2* and *NFKB1* was decreased due to failure in the genotyping in some selected tagSNPs.

In conclusion, our results indicate that genetic variation in *NOD2* may play a role in GC of noncardia localization, while *CD14* was associated with cardia GC.

#### Acknowledgements

The authors acknowledge technical contribution from Mrs. Nadia García from the Catalan Institute of Oncology, Dr. Magda Montfort and Dr. Sebastián Morán, from the Spanish National Genotyping Center (CEGEN), Dr. Francis Megraud from the INSERM U853, Bordeaux, France, for his contribution to the laboratory analysis of *H. pylori* and to the pathology panel members Drs. Fatima Carneiro, Drs. Hendrik Blaker, Drs. Claus Fenger, Drs. Laszlo Igali, Drs. Gabriella Nesi and Drs. Roger Stenling for their contribution to the collection and review of paraffin tumor blocks, slides and pathology reports. Osmel Companioni is supported by a PhD fellowship from the Research Institute of Bellvitge, IDIBELL.

## References

1. Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2008;127:2893-917.
2. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Schistosomes, liver flukes and *Helicobacter pylori*: views and expert opinion of an IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, which met in Lyon, 7-14 June 1994. ISBN 92 382 1261 4. ISSN 0250-9555.
3. Smith G, Hold GL, Taham E, et al. Cellular and molecular aspects of gastric cancer. *World J Gastroenterol* 2006;12:2979-90.
4. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. *Int Rev Immunol* 2011;30:16-34.
5. Armant MA, Fenton MJ. Toll-like receptors: a family of pattern-recognition receptors in mammals. *Genome Biol* 2002;29:3reviews3011.
6. Tahara T, Shibata T, Hirata I, et al. CD14 promoter-159 polymorphism is associated with reduced risk of intestinal-type gastric cancer in a Japanese population. *Dig Dis Sci* 2009;54:1508-12.
7. Zhao D, Sun T, Zhang X, et al. Role of CD14 promoter polymorphisms in *Helicobacter pylori* infection related gastric carcinoma. *Clin Cancer Res* 2007;13:2362-68.
8. Tahara T, Arisawa T, Shibata T, et al. Association of polymorphism of TLR4 and CD14 genes with gastroduodenal diseases in Japan. *Inflammopharmacology* 2007;15:124-28.
9. Hold GL, Rabkin CS, Gammon MD, et al. CD14-159C/T and TLR9-1237T/C polymorphisms are not associated with gastric cancer risk in Caucasian populations. *Eur J Cancer* 2009;18:117-19.
10. Hnatyszyn A, Szalata M, Stanczyk J, et al. Association of c.802C/T polymorphism of NOD2/CARD15 gene with the chronic gastritis and predisposition to cancer in H. *pylori* infected patients. *Exp Mol Pathol* 2010;88:388-93.
11. Wex T, Ebert MP, Kropf S, et al. Gene polymorphisms of the NOD-2/CARD-15 gene and the risk of gastric cancer in Germany. *Anticancer Res* 2008;28:757-62.
12. Angeletti S, Galluzzo S, Santini D, et al. NOD2/CARD15 polymorphisms impair innate immunity and increase susceptibility to gastric cancer in an Italian population. *Hum Immunol* 2009;70:729-32.
13. Yazdanyar S, Nordestgaard BG. NOD2/CARD15 genotype and common gastrointestinal diseases in 43 600 individuals. *J Intern Med* 2010;267:228-36.
14. Hold GL, Rabkin CS, Chow WH, et al. Polymorphism of toll-like receptor 4 gene increases risk of gastric carcinoma and its precursors. *Gastroenterology* 2007;132:905-12.
15. de Oliveira JG, Silva AE. Polymorphisms of the TLR2 and TLR4 genes are associated with risk of gastric cancer in a Brazilian population. *World J Gastroenterol* 2012;18:1235-42.
16. Santini D, Angeletti S, Ruzzo A, et al. Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms in gastric cancer of intestinal and diffuse histotypes. *Clin Exp Immunol* 2008;154:360-64.
17. Trejo-de la OA, Torres J, Pérez-Rodríguez M, et al. TLR4 single-nucleotide polymorphisms alter mucosal cytokine and chemokine patterns in Mexican patients with *Helicobacter pylori*-associated gastroduodenal diseases. *Clin Immunol* 2008;129:333-40.
18. Hishida A, Matsuo K, Goto Y, et al. Toll-like receptor 4 + 3725 G/C polymorphism *Helicobacter pylori* seropositivity, and the risk of gastric atrophy and gastric cancer in Japanese. *Helicobacter* 2009;14:47-53.
19. Lo SS, Chen JH, Wu CW, et al. Functional polymorphism of NFKB1 promoter may correlate to the susceptibility of gastric cancer in aged patients. *Surgery* 2009;145:280-85.
20. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol* 1992;3:783-91.
21. Carneiro F, Moutinho C, Pera G, et al. Pathology findings and validation of gastric and esophageal cancer cases in a European cohort (EPIC/EUR-GAST). *Scand J Gastroenterol* 2007;42:618-27.
22. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.
23. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-65.
24. Sala N, Muñoz X, Travier N, et al. Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: results from the EPIC-EURGAST study. *Int J Cancer* 2012;130:2417-27.
25. Palli D, Masala G, Del Giudice G, et al. CagA+ *Helicobacter pylori* infection and gastric cancer risk in the EPIC-EURGAST study. *Int J Cancer* 2007;120:859-67.
26. González CA, Megraud F, Buissonniere A, et al. *Helicobacter pylori* infection assessed by ELISA and by immunoblot and noncardia gastric cancer risk in a prospective study: the Eurgast-EPIC project. *Ann Oncol* 2012;23:1320-24.
27. So HC, Sham PC. Multiple testing and power calculations in genetic association studies. *Cold Spring Harb Protoc* 2011;2011.pdb.top95. doi: 10.1101/pdb.top95.
28. Helicobacter and Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* 2001;49:347-53.
29. González JR, Armengol L, Solé X, et al. SNPpass: an R package to perform whole genome association studies. *Bioinformatics* 2007;23:644-5.
30. McLaren W, Pritchard B, Rios D, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* 2010;26:2069-70.
31. Reumers J, Pritchard B, Rios D, et al. Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNP effect and PupaSuite databases. *Nucleic Acid Res* 2008;36:D825-9.
32. Wang P, Zhang L, Jiang JM, et al. Association of NOD1 and NOD2 genes polymorphisms with *Helicobacter pylori* related gastric cancer in a Chinese population. *World J Gastroenterol* 2012;18:2112-20.
33. Rosenstiel P, Hellmig S, Hampe J, et al. Influence of polymorphisms in the NOD1/CARD4 and NOD2/CARD15 genes on the clinical outcome of *Helicobacter pylori* infection. *Cell Microbiol* 2006;8:1188-98.
34. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599-603.
35. Baldini M, Lohman IC, Halonen M, et al. A polymorphism in the 5' flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am J Respir Cell Mol Biol* 1999;20:976-83.
36. Wu MS, Cheng TY, Shun CT, et al. Functional polymorphisms of CD14 and toll like receptor 4 in Taiwanese Chinese with *Helicobacter pylori* related gastric malignancies. *Hepatogastroenterology* 2006;53:807-10.
37. Karhukorpi J, Yan Y, Niemela S, et al. Effect of CD14 promoter polymorphism and H. *pylori* infection and its clinical outcomes on circulating CD14. *Clin Exp Immunol* 2002;128:326-32.
38. Correa P. A human model of gastric carcinogenesis. *Cancer Res* 1988;48:3554-60.
39. Cao LX, Le Bousse-Kerdiles MC, Clay D, et al. Implication of a new molecule IK in CD34 hematopoietic progenitor cell proliferation and differentiation. *Blood* 1997;89:3615-23.
40. Baba T, Hanagiri T, Takenoyama M, et al. Identification of a lung cancer antigen evading CTL attack due to loss of human leukocyte antigen (HLA) class I expression. *Cancer Sci* 2010;101:2115-20.
41. Yoshiyama H, Imai S, Shimizu N, et al. Epstein-Barr virus infection of human gastric carcinoma cells: implication of the existence of a new virus receptor different from CD21. *J Virol* 1997;71:5688-91.
42. Kurt-Jones EA, Popova L, Kwinn L, et al. Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. *Nat Immunol* 2000;1:398-401.
43. Corvalan A, Koriyama C, Akiba S, et al. Epstein-Barr virus in gastric carcinoma is associated with location in the cardia and with a diffuse histology: a study in one area of Chile. *Int J Cancer* 2001;94:527-30.
44. Lee JH, Kim SH, Han SH, et al. Clinicopathological and molecular characteristics of Epstein-Barr virus-associated gastric carcinoma: a meta-analysis. *J Gastroenterol Hepatol* 2009;24:354-65.
45. Garza-Gonzalez G, Bosques-Padilla FJ, Mendoza-Ibarra SI, et al. Assessment of the toll-like receptor 4 Asp299Gly, Thr399Ile and interleukin-8 -251 polymorphisms in the risk for the development of distal gastric cancer. *BMC Cancer* 2007;7:70.
46. Gwang Kim J, Sohn SK, Chae YS, et al. No association of the NFKB1 insertion/deletion promoter polymorphism with survival in patients with gastric cancer. *Jpn J Clin Oncol* 2009;39:497-501.
47. Lu R, Gao X, Chen Y, et al. Association of an NFKB1 intron SNP (rs4648068) with gastric cancer patients in the Han Chinese population. *BMC Gastroenterol* 2012;12:87.