

## **TREBALL DE FINAL DE GRAU**

***SUSCEPTIBILITAT GENÈTICA AL CÀNCER GÀSTRIC:  
CARACTERITZACIÓ DE LES INTERACCIONS DE SEGON ORDRE ENTRE GENS I VARIANTS  
IDENTIFICADES EN L'ESTUDI EPIC-EURGAST***

NÚRIA MONTSERRAT ROS I ROCHER



**GRAU EN BIOTECNOLOGIA**

**COTUTORES:** DRA. NÚRIA SALA SERRA (ICO – IDIBELL) I DRA. MALU CALLE ROSINGANA (UVIC)

VIC, JUNY DEL 2013

## ÍNDEX

1. RESUM	
1.1. RESUM TREBALL FINAL DE GRAU	PÀGINA 04
1.2. FINAL DEGREE PROJECT SUMMARY	PÀGINA 05
2. INTRODUCCIÓ	PÀGINA 06
3. ANTECEDENTS	
3.1. ANÀLISI D'ASSOCIACIÓ AMB LES DADES DEL PROJECTE EPIC-EURGAST	PÀGINA 09
3.2. ANÀLISI PER <i>GENE-BASED PERMUTATION TEST</i>	PÀGINA 10
3.3. RELLEVÀNCIA DE L'ANÀLISI D'INTERACCIONS	PÀGINA 10
4. OBJECTIU	PÀGINA 12
5. MATERIALS	PÀGINA 12
6. METODOLOGIA	
6.1. ANÀLISI D'INTERACCIONS APLICANT EL MÈTODE <i>MODEL-BASED MULTIFACTOR DIMENSIONALITY REDUCTION</i>	PÀGINA 13
6.2. MÈTODE 1: FILTRATGE D'INTERACCIONS AMB UN SNP SIGNIFICATIU	PÀGINA 15
6.3. MÈTODE 2: FILTRATGE D'INTERACCIONS SEGONS LA SINÈRGIA	PÀGINA 15
6.4. ANÀLISI GLOBAL DELS RESULTATS	PÀGINA 17
7. RESULTATS	
7.1. MÈTODE 1: FILTRATGE D'INTERACCIONS AMB UN SNP SIGNIFICATIU	PÀGINA 17
7.2. MÈTODE 2: FILTRATGE D'INTERACCIONS SEGONS LA SINÈRGIA	PÀGINA 20
7.3. RESULTAT FINAL DE L'ANÀLISI D'INTERACCIONS	PÀGINA 22
8. DISCUSSIÓ DE RESULTATS	
8.1. COMPARACIÓ AMB ANÀLISIS PRÈVIES	PÀGINA 23
8.2. PLAUSIBILITAT BIOLÒGICA DE LES INTERACCIONS IDENTIFICADES	PÀGINA 23

9. CONCLUSIONS	PÀGINA 27
10. EL FUTUR	PÀGINA 28
11. REFERÈNCIES	PÀGINA 29
12. ANNEXES	
12.1. <i>MODEL-BASED MULTIFACTOR DIMENSIONALITY REDUCTION METHOD</i>	PÀGINA 34
12.2. SCRIPTS APLICATS PER A L'ANÀLISI DEL MÈTODE 1 AMB R	PÀGINA 35
12.3. MÈTODE 1 – RESULTAT STEP 1 I 2 <i>MB-MDR</i>	
12.3.1. MÈTODE 1 – RESULTAT STEP 1 <i>MB-MDR</i>	PÀGINA 37
12.3.2. MÈTODE 1 – RESULTAT STEP 2 <i>MB-MDR</i>	PÀGINA 40
12.4. PAQUET ESTADÍSTIC SINÈRGIA	PÀGINA 41
12.5. SCRIPTS APLICATS PER A L'ANÀLISI DEL MÈTODE 2 AMB R	PÀGINA 42
12.6. MÈTODE 2 – RESULTATS STEP 1 I 2 <i>MB-MDR</i>	
12.6.1. MÈTODE 2 – RESULTATS STEP 1 <i>MB-MDR</i>	PÀGINA 43
12.6.2. MÈTODE 2 – RESULTATS STEP 2 <i>MB-MDR</i>	PÀGINA 45

## 1. RESUM

### 1.1 RESUM TREBALL FINAL DE GRAU – GRAU EN BIOTECNOLOGIA

**Títol:** *Susceptibilitat genètica al càncer gàstric: caracterització de les interaccions de segon ordre entre gens i variants identificades en l'estudi EPIC-EURGAST*

**Paraules clau:** *MB-MDR, interaccions genètiques, SNP, Sinèrgia*

**Autora:** Núria Montserrat Ros i Rocher

**Tutors:** Dra. Núria Sala Serra (ICO – IDIBELL) i Dra. Malu Calle Rosingana (UVic)

**Data:** Juny de 2013

En aquest *Treball de Final de Grau* s'exposen els resultats de l'anàlisi de les dades genètiques del projecte **EurGast2** "*Genetic susceptibility, environmental exposure and gastric cancer risk in an European population*", estudi cas-control niat a la cohort europea **EPIC** "*European Prospective Investigation into Cancer and Nutrition*", que té per objectiu l'estudi dels factors genètics i ambientals associats amb el risc de desenvolupar càncer gàstric (CG). A partir de les dades resultants de l'estudi **EurGast2**, en el què es van analitzar 1.294 SNPs en 365 casos de càncer gàstric i 1.284 controls en l'anàlisi *Single SNP* previ, la hipòtesi de partida del present *Treball de Final de Grau* és que algunes variants amb un efecte marginal molt feble, però que conjuntament amb altres variants estarien associades al risc de CG, podrien no haver-se detectat.

Així doncs, l'objectiu principal del projecte és la identificació d'interaccions de segon ordre entre variants genètiques de gens candidats implicades en la carcinogènesi de càncer gàstric. L'anàlisi de les interaccions s'ha dut a terme aplicant el mètode estadístic *Model-based Multifactor Dimensionality Reduction Method (MB-MDR)*, desenvolupat per *Calle et al.* l'any 2008 i s'han aplicat dues metodologies de filtratge per seleccionar les interaccions que s'exploraran: 1) filtratge d'interaccions amb un SNP significatiu en el *Single SNP analysis* i 2) filtratge d'interaccions segons la mesura *Sinèrgia*.

Els resultats del projecte han identificat 5 interaccions de segon ordre entre SNPs associades significativament amb un major risc de desenvolupar *càncer gàstric*, amb *p-valor* inferior a  $10^{-4}$ . Les interaccions identificades corresponen a interaccions entre els gens *MPO* i *CDH1*, *XRCC1* i *GAS6*, *ADH1B* i *NR5A2* i *IL4R* i *IL1RN* (que s'ha validat en les dues metodologies de filtratge). Excepte *CDH1*, cap altre d'aquests gens s'havia associat significativament amb el CG o prioritzat en les anàlisis prèvies, el que confirma l'interès d'analitzar les interaccions genètiques de segon ordre. Aquestes poden ser un punt de partida per altres anàlisis destinades a confirmar gens putatius i a estudiar a nivell biològic i molecular els mecanismes de carcinogènesi, i orientades a la recerca de noves dianes terapèutiques i mètodes de diagnosi i pronòstic més eficients.

## 1.2. FINAL DEGREE PROJECT SUMMARY – DEGREE IN BIOTECHNOLOGY

**Title:** *Genetic susceptibility to gastric cancer: characterization of second order interactions between genes and variants identified in the EPIC-EURGAST study*

**Keywords:** *MB-MDR, genetic interactions, SNP, Synergy*

**Autor:** Núria Montserrat Ros i Rocher

**Tutors:** Dra. Núria Sala Serra (ICO – IDIBELL) and Dra. Malu Calle Rosingana (UVic)

**Date:** June of 2013

In this *Final Degree Project* the results of the genetic analysis of the nested-case/control **EurGast2** project "*Genetic susceptibility, environmental exposure and gastric cancer risk in an European population*" of the European **EPIC** cohort "*European Prospective Investigation into Cancer and Nutrition*" are exposed. Its aim is to study the genetic and environmental factors associated to the risk of developing gastric cancer (GC). In base of the **EurGast2** study which analyzed 1.294 SNPs in 365 gastric cancer cases and 1.284 controls in a previous *Single SNP* analysis, the present hypothesis of this *Final Degree Project* defends that some variants with a weak marginal effect which may be associated to GC together with other variants could have not been detected.

The aim of this project is to identify second order interactions between genetic variants from candidate genes implicated in gastric cancer carcinogenesis. The interaction's analysis has been carried out applying the statistical *Model-based Multifactor Dimensionality Reduction Method MB-MDR* developed by *Calle et al. in 2008* and two different methodologies have been applied in order to select the best second order interactions to explore: 1) interactions' filtering with one significant SNP in the *Single SNP analysis* and 2) interactions' filtering using the *Synergy* measure.

The results of the project have identified that **5** second order interactions between different SNPs are statistically associated with a major risk of developing *gastric cancer* at a  $p\text{-value} < 10^{-4}$ . These interactions identified correspond to interactions between SNPs from *MPO* and *CDH1* genes, *XRCC1* and *GAS6* genes, *ADH1B* and *NR5A2* genes and *IL4R* and *IL1RN* genes (which has been validated in both independent filter methods). Except *CDH1*, any other of these genes have been associated statistically to gastric cancer or has been prioritized in the previous univariant analysis. This fact confirms the interest of analyzing genetic second order interactions which can represent a start point for other analysis to confirm putative genes and study at a biological and molecular level the mechanisms underpinning gastric cancer carcinogenesis which can represent a step closer to the research of new therapeutic targets and the definition of more efficient diagnosis methods.

## SUSCEPTIBILITAT GENÈTICA AL CÀNCER GÀSTRIC: CARACTERITZACIÓ DE LES INTERACCIONS DE SEGON ORDRE ENTRE GENS I VARIANTS IDENTIFICADES EN LA COHORT EUROPEA EPIC

### 2. INTRODUCCIÓ

Actualment, una de les principals causes de mort a nivell mundial després de les malalties cardiovasculars són les neoplàsies malignes, la majoria de les quals estan determinades per factors genètics i alteracions epigenètiques.

Entre els molts estudis sobre càncer que es porten a terme a diferents nivells, un dels més importants és l'estudi prospectiu europeu sobre dieta, càncer i salut, l'**estudi EPIC**- *European Prospective Investigation into Cancer and Nutrition* (2, 27), en el que s'avalua l'efecte de la dieta, factors ambientals, d'estil de vida i genètics, sobre la incidència i la mortalitat global deguda al càncer i la seva relació quant a les principals localitzacions tumorals.

Els estudis prospectius, com l'**estudi EPIC**, permeten superar la possible inconsistència de les associacions resultants d'estudis retrospectius o estudis cas-control convencionals basats en individus malalts de càncer, ja que permeten l'avaluació de les dades dels pacients en el passat. L'avaluació d'aquestes dades, com ara la dieta, és de gran complexitat i variabilitat durant el temps si no s'usen els mètodes i dissenys adequats que també poden presentar les seves pròpies limitacions i moltes vegades requereixen validacions posteriors.

Table 2 | **EPIC-study cancers**

Cancer type	Number
Upper gastrointestinal tract	372
Stomach	368
Colorectal	1,913
Pancreatic	336
Lung	1,309
Breast	6,169
Cervical	676
Ovarian	659
Uterine	785
Prostate	1,515
Kidney	368
Bladder	700
Other	9,025

EPIC, European Prospective Investigation into Cancer and Nutrition.

FIGURA 1: *Tipus de càncers inclosos en l'estudi EPIC.* BINGHAM, S. and RIBOLI, E. "Diet and Cancer – The European Prospective Investigation into Cancer and Nutrition". *Nature Reviews | Cancer*, 4 (2004): 206-215, doi:10.1038/nrc1298.

L'**estudi EPIC** inclou més de mig milió d'individus sans de 23 centres participants de diferents regions geogràfiques en 10 països europeus, reclutats i enquestats per recopilar dades sobre la seva dieta, salut, estil de vida i mesures antropomètriques entre el 1992 i el 2000. De la majoria també es va obtenir mostra biològica com plasma, sèrum i limfòcits, que s'ha conservat congelada en nitrogen líquid fins al seu ús.

Periòdicament es va fer un seguiment de la cohort per registrar-ne dades clíniques, especialment les relacionades amb el desenvolupament de càncer (2).

El seguiment dels individus analitzats en aquest estudi es va completar entre el 2003 i 2006 depenent dels centres d'estudi. Pel que fa al càncer, en aquest projecte (**Figura 1**) es van identificar gairebé 20.000 nous casos de càncer, incloent 6.169 casos de càncer de mama, 1.913 de càncer colorectal, 1.515 casos per càncer de pròstata, 1.309 casos de càncer de pulmó i 368 pel càncer gàstric o estómac.

Cal destacar que el càncer gàstric (CG) tant en homes com en dones és el segon càncer que causa més mortalitat a nivell mundial, després del càncer de pulmó en homes i el de mama en dones. A nivell europeu és el quart càncer més prevalent i presenta les seves taxes d'incidència més elevades a Corea, Japó i altres països de l'est d'Àsia. A l'estat Espanyol és un dels càncers que causa més mortalitat.

Depenent de la seva histologia, el CG es classifica en tipus *difús*, quan les cèl·lules cancerígenes són infiltrades a través de la mucosa i no presenten un patró ben diferenciat, i en tipus *intestinal*, quan les cèl·lules estan ben diferenciades en estructures glandulars. Depenent de la seva localització, el CG es classifica en adenocarcinoma proximal o *càrdies* que presenta un pitjor pronòstic i distal o *no càrdies*, el més comú (**Figura 2**) (25, 30, 31, 34, 37).

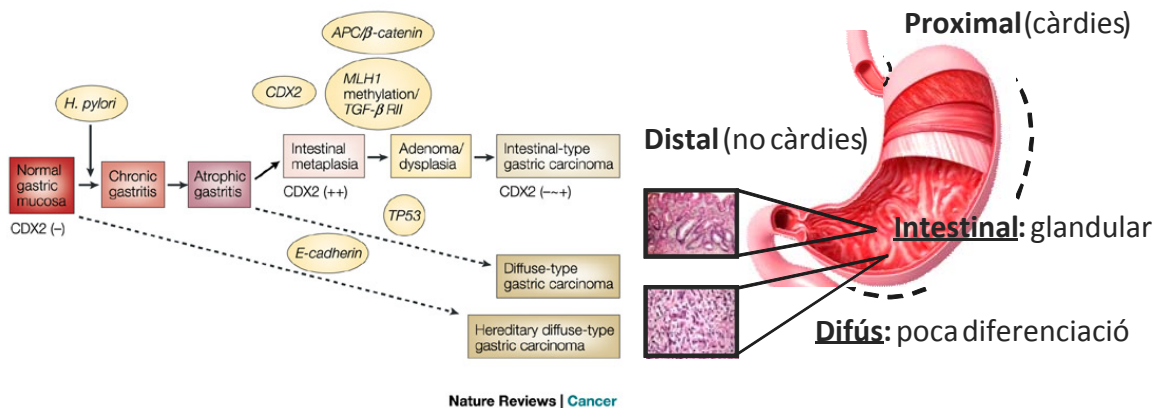


FIGURA 2: Desenvolupament del càncer gàstric i els seus subtipus histològics i anatòmics. YUSASA, Y. "Control of gut differentiation and intestinal-type gastric carcinogenesis". *Nature Reviews | Cancer*, 3 (2004): 592-600, doi: 10.1038/nrc1141.

La seva incidència a la població se sap que pot estar condicionada per diversos factors tant genètics com ambientals; de fet, s'observen diferències en la seva incidència a la població espanyola que habita en diferents regions geogràfiques.

Un dels principals factors de risc pel CG més comú, que és el distal o *no càrdies*, és la infecció per *Helicobacter pylori*. Aquest risc es veu incrementat segons la virulència de la soca bacteriana, condicionada per la pròpia susceptibilitat de l'hoste i la presència o absència d'una illa de patogenicitat al seu genoma que confereix a l'hoste més risc a desenvolupar un càncer gàstric (3, 10, 24).

El procés de desenvolupament del carcinoma s'inicia amb una inflamació localitzada en el focus de la infecció que progressa a una gastritis crònica atròfica multifocal en el cos amb hipoclorhídria i acaba amb una metaplàsia intestinal, que pot arribar a desenvolupar displàsia i finalment el carcinoma. Aquest procés arriba a desenvolupar-se només en un 1% dels individus infectats per *Helicobacter pylori* que, combinat amb altres factors ambientals com l'alimentació i estil de vida, el consum d'alcohol i tabac entre d'altres, pot també condicionar el desenvolupament del càncer gàstric (16, 23, 25).



IMATGE 1: *Helicobacter pylori*<sup>1</sup>.

Paral·lelament als estudis basats en les condicions ambientals de risc, una altra de les línies d'investigació sobre el càncer gàstric és l'estudi dels factors genètics que poden estar associats amb el risc de desenvolupar la malaltia amb l'objectiu d'identificar biomarcadors òptims per a la seva detecció precoç. Se sap que el CG hereditari representa un 10% de tots els tumors gàstrics i és majoritàriament de tipus *difús*. A més, s'ha observat que un terç dels casos amb CG hereditari són causats per mutacions del gen de la *E-cadherina (CDH1)* (18). A part del sever efecte de les mutacions en aquest gen, l'efecte més lleu de variants polimòrfiques en altres gens (gens de susceptibilitat) també pot determinar l'aparició de mutacions somàtiques i/o amplificacions o sobreexpressió dels oncògens *c-K-ras*, *c-erbB-2*, *K-sam* i *c-met*, la inactivació de gens supressors de tumors com *p53*, *APC* i *DCC*, i la transcripció anormal de *CD44*. Es creu que tenen un paper clau en la carcinogènesi gàstrica, però varien en freqüència entre els dos subtipus histològics *intestinal* i *difús* (14, 30, 31). Per aquest motiu, actualment els mecanismes moleculars que poden determinar el desenvolupament d'un subtipus o altre són encara poc coneguts.

Per altra banda, encara que avui en dia la incidència d'aquesta patologia està disminuint progressivament gràcies a la millora en els hàbits sanitaris, higiènics, la qualitat dels productes que consumim i els tractaments i mètodes de diagnòstic, el càncer gàstric és una malaltia greu amb un baix índex de supervivència (menys del 20% dels pacients sobreviuen als 5 anys), principalment perquè es diagnostica en estadis avançats. La gastroscòpia és actualment la mesura més habitual pel control i prevenció de la malaltia. La determinació dels factors de susceptibilitat genètica de l'hoste i la validació i caracterització de les variants que poden estar associades significativament amb el CG pot permetre confirmar gens putatius relacionats amb aquesta patologia i estudiar a nivell biològic i molecular els mecanismes de carcinogènesi. El futur d'aquests estudis podrà orientar-se a la recerca de noves dianes terapèutiques i mètodes de diagnòstic i pronòstic pels pacients més eficients i així poder aconseguir una millora en la seva qualitat de vida, augmentar la supervivència i arribar potser a la cura del càncer gàstric.

<sup>1</sup> Font IMATGE 1: [<http://www.cosasdesalud.es/infeccion-por-helicobacter-pylori/>]



### 3. ANTECEDENTS

#### 3.1. ANÀLISI D'ASSOCIACIÓ AMB LES DADES DEL PROJECTE EPIC-EURGAST

El **Projecte EurGast** "*Genetic susceptibility, environmental exposure and gastric cancer risk in an European population*", és un estudi cas-control niat a la cohort europea EPIC (2, 27, 29). Entre l'any 2002 i 2004 es va dur a terme l'estudi **EurGast1** amb l'objectiu d'analitzar 93 polimorfismes de 47 gens candidats en 243 casos i 1.141 controls<sup>2</sup>. Tres anys després de la seva finalització, es va iniciar l'estudi **EurGast2** (entre el 2007 i el 2010) amb l'objectiu de reanalitzar gens/polimorfismes amb resultats inconcluents en l'estudi previ **EurGast1** i analitzar nous candidats que poden estar relacionats amb el CG en una mostra ampliada de 365 casos i 1.284 controls<sup>1</sup>.

Per a l'estudi **EurGast2** es van genotipar 1.536 SNPs seleccionats per *haplotype tagging* a partir de la informació del *Projecte HapMap* i altres SNPs funcionals anotats en diferents Bases de Dades. Dels 1.536 SNPs seleccionats, 1.270 SNPs marcaven o es localitzaven en 136 gens candidats a estar associats al CG, 67 s'havien associat a càncers epitelials en diferents *GWAS (Genome Wide Association Study)* i 199 marcaven 52 regions de microRNAs associades a càncer i/o que tenien com a *targets* un mínim de 20 dels gens candidats analitzats. En total es van analitzar aproximadament 250 regions genòmiques diferents.

Després del genotipat a la plataforma d'*Illumina Golden Gate*, el filtratge de les dades i el control de qualitat van obligar a excloure els SNPs que presentaven formes monomòrfiques, baixa senyal en els al·lels, una mala separació de *clústers* (diferents genotips) o baix percentatge en la genotipació, de tal manera que es va procedir finalment a l'anàlisi estadístic *Single SNP* de 1.294 SNPs (27, 29).

En una primera anàlisi del total de les dades, es va analitzar l'associació per regressió logística amb les variables resposta CG global i els seus subtipus histològics (*intestinal* i *difús*) i anatòmics (*càrdies* i *no càrdies*) segons els diferents models d'herència *Dominant*, *Recessiu*, *Codominant*, *Overdominant* i *Additiu*.

El nombre de casos, controls i *missings* de les variables resposta de CG global i els seus subtipus anatòmics i histològics es mostren a la **Taula 1**. En aquestes anàlisis es va aplicar el paquet estadístic *SNPassoc* del programa R, també utilitzat per analitzar l'*Equilibri de Hardy-Weinberg HWE* sobre els controls (27, 29).

---

<sup>2</sup>*Matched controls*: per a cada cas es van seleccionar entre 3 i 4 controls del mateix centre, edat, sexe i data d'extracció de sang en els estudis **EurGast1** i **EurGast2**.

**Taula 1:** Distribució del nombre de casos, controls i *missings* segons les diferents variables resposta per CG.

VARIABLE RESPOSTA					
	CG GLOBAL	CÀRDIES	NO CÀRDIES	DIFÚS	INTESTINAL
<b>CASOS</b>	<b>365</b>	107	181	128	126
<b>CONTROLS</b>	<b>1284</b>	1284	1284	1284	1284
<b>MISSINGS</b>	-	258	184	237	239

Dels 1.294 SNPs avaluats, 315 SNPs van mostrar associació amb les variables resposta CG global o algun dels seus subtipus histològics i/o anatòmics, amb  $p\text{-valor} < 0.05$ , en algun dels diferents models d'herència avaluat. D'aquests SNPs, 135 s'associaven amb el CG en general i la resta s'associaven amb algun dels seus subtipus.

### 3.2. ANÀLISI GENE-BASED PERMUTATION TEST SOBRE LES DADES DEL PROJECTE EPIC-EURGAST

Un cop realitzat el *Single SNP analysis* de cadascun dels polimorfismes es va dur a terme una altra metodologia per tal de poder determinar l'associació estadística a nivell de gens (*Gene-based Permutation Test*) (1, 11, 19, 32). Aquesta metodologia ajusta els resultats obtinguts en l'anàlisi prèvia pel nombre de polimorfismes o SNPs analitzats per gen i proporciona un  $p\text{-valor}$  a nivell de gen. Els resultats van indicar que 10 gens o regions genòmiques diferents s'associaven significativament amb el CG en general i uns altres 42 gens s'associaven a algun dels seus subtipus histològics o anatòmics. L'únic gen que va superar el *Gene-based Permutation Test* i també la correcció posterior pel nombre de regions analitzades mitjançant *Bonferroni* va ser *PSCA* (27, 29).

### 3.3. RELLEVÀNCIA DE L'ANÀLISI D'INTERACCIONS

A partir de la informació d'aquestes anàlisis d'associació prèvies sobre les dades de l'estudi **EPIC-EurGast** es va obtenir una llista de gens candidats per a realitzar validacions independents. Tot i això, és probable que algunes variants associades amb CG no s'hagin detectat en aquesta anàlisi inicial. Per exemple, aquelles variants que tenen un efecte marginal molt feble però que conjuntament amb altres variants tenen un impacte important en el risc de CG.

Per aquest motiu, una anàlisi independent d'associació entre les interaccions SNP a SNP i el CG podria ser molt interessant ja que podria aportar nous candidats que no van ser detectats en aquests estudis previs.

Una interacció o *epístasi*, terme més utilitzat a la genètica de poblacions, defineix l'efecte que es produeix quan l'efecte d'un gen és canviat o modulad per un o més gens; és a dir, l'expressió fenotípica d'un gen depèn de la presència d'un o més gens moduladors tot i que no necessàriament implica una interacció bioquímica entre ells o els seus productes (25). En aquest *Treball de Final de Grau*, una interacció entre dos SNPs correspon a un efecte conjunt de dues variants genètiques superior a la suma dels efectes individuals de cadascuna de les variants.

Així doncs, l'anàlisi de les interaccions genètiques pot revelar noves funcions de gens i també interaccions entre proteïnes; de fet, la combinació de diverses proteïnes en complexos proteics és responsable de moltes funcions i processos biològics. Per tant, l'anàlisi de les interaccions genètiques pot ser una eina important per aportar una nova visió integrada d'interaccions funcionals entre gens que podria ser clau en la carcinogènesi gàstrica (9).

#### 4. OBJECTIU

- Analitzar les interaccions de segon ordre (SNP a SNP) entre polimorfismes de gens inclosos a l'estudi **EPIC-Eurgast2** aplicant el mètode *Model-based Multifactor Dimensionality Reduction Method (MB-MDR)* (6, 8) (*Annexes, apartat 12.1*) per a identificar nous possibles candidats i confirmar altres gens ja seleccionats en les anàlisis prèvies que poden tenir un paper important en la carcinogènesi gàstrica.

#### 5. MATERIALS

L'estudi de les interaccions entre SNPs s'ha fet a partir de la informació de la **Base de Dades del Projecte EurGast2**, especialment la referida als genotips dels casos i controls pels diferents polimorfismes analitzats.

La Base de Dades conté els resultats del genotipat de 1.536 SNPs en un total de 1.649 individus (365 casos de CG i 1.284 controls), amb informació addicional dels identificadors **EPIC** de les mostres, del seu estatus (cas/control) i, pels casos, subtipus anatòmic i histològic del CG. També hi ha altra informació addicional com la data d'extracció de sang, país, centre, sexe i edat de cadascuna de les mostres.

A tenir en compte que el nombre total d'SNPs vàlids per a l'anàlisi, un cop filtrats segons els diferents motius d'exclusió (*Antecedents, apartat 3.1*), és de 1.294 SNPs.

La variable resposta que s'ha considerat en aquest projecte per a l'anàlisi és la variable CG global, ja que per subtipus histològics i anatòmics, al tractar-se de subgrups de dades sobre la mateixa variable resposta global, no hi ha prou grandària de mostra - N (valors perduts i pocs casos) - i la potència estadística no és tan bona (veure **Taula 1**).

L'anàlisi de les dades es durà a terme amb el programari R (*Annexes, apartat 12.2, 12.5*) utilitzant els següents paquets estadístics:

- **Paquet *SNP assoc* (17).**
- **Paquet *mbmdr* (6).**
- **Paquet *utils* i *base* per aplicar la funció *Sinèrgia*.**

## 6. METODOLOGIA

### 6.1. ANÀLISI D'INTERACCIONS APLICANT EL MÈTODE *MODEL-BASED MULTIFACTOR DIMENSIONALITY REDUCTION*

El mètode *Model-based Multifactor Dimensionality Reduction Method (MB-MDR)*, desenvolupat per *Calle et al.* el 2008 (6, 8) i implementat en el paquet estadístic d'R *mbmdr* proposa una nova metodologia per explorar les interaccions *gen-gen* o *SNP-SNP* en estudis d'associació cas-control.

És una extensió del clàssic *Multifactor Dimensionality Reduction Method (MDR)* proposat per *Ritchie et al.* (28). Permet analitzar tant trets binaris com quantitius usant un model de regressió logística que defineix tres categories de risc (alt risc, baix risc i una categoria per aquelles interaccions que no mostren evidència d'associació), podent ajustar les dades per covariants i per efectes principals en tres passos:

#### ***MB-MDR STEP 1: Classificació dels genotips en les categories de risc.***

La combinació de dos SNPs bial·lèlics, cadascun amb 3 genotips possibles, dona lloc a 9 genotips conjunts. El primer pas de l'anàlisi classifica aquests 9 genotips conjunts en tres categories diferents segons la relació entre el nombre de casos i controls (*Odds Ratio - OR*) amb un llindar de significació per defecte de  $p\text{-valor}=0.10$ .

La categoria **H: High risk**, agrupa les combinacions de genotips que presenten un  $OR>1$  i un  $p\text{-valor}<0.10$  (remarcats a la **Taula 2** en vermell).

La categoria **L: Low risk**, agrupa les combinacions de genotips que presenten un  $OR<1$  i un  $p\text{-valor}<0.10$  (remarcats a la **Taula 2** en verd). Aquestes podrien considerar-se com a factors de protecció davant la variable resposta conferint menys risc a desenvolupar la malaltia.

Per últim, la categoria **0: No evidence of association**, agrupa aquelles combinacions de genotips que presenten un  $p\text{-valor}\geq 0.10$  i també aquelles que presenten un OR pràcticament d'1 i pocs individus (destacats a la **Taula 2** en negreta).

De fet, els *OR* i *p-valors* s'obtenen a partir d'una regressió logística i això permet ajustar les anàlisis per altres factors o pels efectes marginals dels dos SNPs.

**Taula 2:** Exemple de classificació de genotips en un estudi de Càncer de Bufeta.

SNP1-SNP2	0 (bb)	1 (Bb)	2 (BB)
0 (aa)	88/77 (OR=1.03)	50/59 (OR=0.72)	12/6 (OR=1.81)
1 (Aa)	102/114 (OR=0.73)	96/37 (OR=2.70)	14/18 (OR=0.68)
2 (AA)	38/34 (OR=0.99)	18/28 (OR=0.55)	6/6 (OR=0.89)
TOTAL	424/379		

**MB-MDR STEP 2: Agrupació de Genotips.**

En aquest segon pas s'agrupen tots els individus amb genotips d'alt risc en una categoria **H** i es fa una regressió logística aplicant el **test d'associació (Wald Test Statistic)** d'aquest grup contra la resta de dades, que dona un *p*-valor **p<sub>H</sub>**. Es fa el mateix pels individus amb genotips de baix risc (categoria **L**) i s'obté un *p*-valor **p<sub>L</sub>**. Finalment, es retorna el *p*-valor mínim: **p<sub>min</sub>=min(p<sub>H</sub>, p<sub>L</sub>)** (destacat a la Taula 3 en blau).

**Taula 3:** Agrupació de genotips segons l'exemple anterior de la Taula 2.

Categoria	Casos	Controls	OR	<i>p</i> -valor
<b>H</b>	96	37	2.68	<b>p<sub>H</sub>=0.00015</b>
<b>L</b>	120	142	0.65	<b>p<sub>L</sub>=0.37&gt;0.05</b>
<b>0</b>	208	200	NA	<b>NA</b>

**MB-MDR STEP 3: Test de Permutacions.**

Les agrupacions de genotips en alt risc i baix risc s'han fet segons les dades disponibles i això fa que les associacions estiguin sobreajustades; per tant, és necessari ajustar el *p*-valor mitjançant un *Test de Permutacions*. Aquest pas és costós computacionalment i només es realitza sobre aquelles interaccions que surten significatives en l'*Step 2*. El nivell de significació establert per a les diferents anàlisis dels 2 últims passos de l'*MB-MDR* és *p*-valor=10<sup>-4</sup>, amb 10.000 permutacions fixades en el *Test de Permutacions*.

A partir de les dades del genotipat en l'estudi **EurGast2** s'exploren les possibles interaccions genètiques de segon ordre (entre dos SNPs).

Tenint en compte la dimensió de la Base de Dades que inclou els 1.294 SNPs, l'anàlisi exhaustiu de totes les interaccions possibles d'ordre dos (SNP a SNP) equivaldria a gairebé dos milions d'interaccions, càlcul no factible amb els recursos computacionals disponibles.

Per tant, s'ha dut a terme una selecció d'aquests 1.294 SNPs seguint dues metodologies diferents exposades a continuació que permetran obtenir un rànquing d'interaccions on es podran destriar les més significatives. En el primer mètode s'exploren aquelles interaccions SNP a SNP on almenys un dels dos SNPs s'associa significativament amb el CG en el *Single SNP analysis* i tenint en compte tots els models d'herència. En el segon mètode, s'exploren aquelles interaccions que conjuntament mostren certa associació amb el risc de CG segons una mesura anomenada **Sinèrgia** (5).

## 6.2. MÈTODE 1: FILTRATGE D'INTERACCIONS AMB UN SNP SIGNIFICATIU

A partir dels resultats previs del *Single SNP analysis* en els diferents models d'herència es van seleccionar 135 SNPs que s'associaven significativament ( $p\text{-valor}\leq 0.05$ ) amb la variable resposta CG global. Amb aquests SNPs, s'aplicà el paquet *MB-MDR*, fixant-los i avaluant-los sobre el total de la mostra, ajustant pels efectes principals per estudiar si tenen interaccions amb algun dels 1.293 SNPs restants genotipats (SNPs totals de la Base de Dades excepte ells mateixos). Per tant, amb com a mínim un SNP significatiu, s'obtindrà un rànquing de les interaccions més interessants per a ser replicades, on un dels dos SNPs a comparar és significatiu en l'anàlisi d'associació.

Tot i això, s'han de tenir en compte les limitacions d'aquesta primera anàlisi: és molt possible que no s'avaluïn interaccions en què els dos SNPs no són significatius en l'anàlisi d'associació però sí tenen una significació en la seva interacció.

## 6.3. MÈTODE 2: FILTRATGE D'INTERACCIONS SEGONS LA SINÈRGIA

La funció **Sinèrgia** va ser introduïda per *Varadan et al.* l'any 2006 (33) i va ser desenvolupada per *Calle et al.* el 2008 (5) com a mesura de la informació addicional que un conjunt de factors aporta a un fenotip després d'eliminar la informació individual que aporta cada factor (veure **Figura 3**).

Estadísticament, és especialment adequada per la detecció d'interaccions genètiques ja que captura el concepte de *l'epístasi* estadística, que pot definir-se com a una desviació d'un model matemàtic additiu, resumint la relació entre genotips multi-locus i la variació fenotípica en una població.

Utilitzar la mesura de *Sinèrgia* pot ser molt útil com a tècnica de selecció en conjunts on el número de possibles interaccions és massa elevat i una anàlisi exhaustiva de totes les interaccions no és factible. Permet identificar aquelles interaccions genètiques interessants per a anàlisis posteriors. Així, com que la *Sinèrgia* elimina la informació individual de cada factor, el seu ús és una forma indirecta d'ajust per efectes genètics principals.

En aquest cas, el cost computacional de calcular la *Sinèrgia* entre dos SNPs és molt baix i per aquest motiu és possible analitzar totes les interaccions entre els 1.294 SNPs de la Base de Dades (s'analitzaran gairebé dos milions d'interaccions). La mesura de *Sinèrgia*, però, té dues limitacions importants: no permet ajustar per altres variables i no proporciona una mesura de significació (*p-valor*).

Per això, després del càlcul de la *Sinèrgia* que proporciona un rànquing de les interaccions amb més potencialitat s'aplicarà el mètode *MB-MDR*. Concretament, es seleccionaran les 500 millors interaccions resultants de l'anàlisi per la *Sinèrgia*. D'aquesta manera s'identificaran aquelles interaccions en les que la suma de l'efecte conjunt dels SNPs és diferent a la suma dels seus efectes individuals.

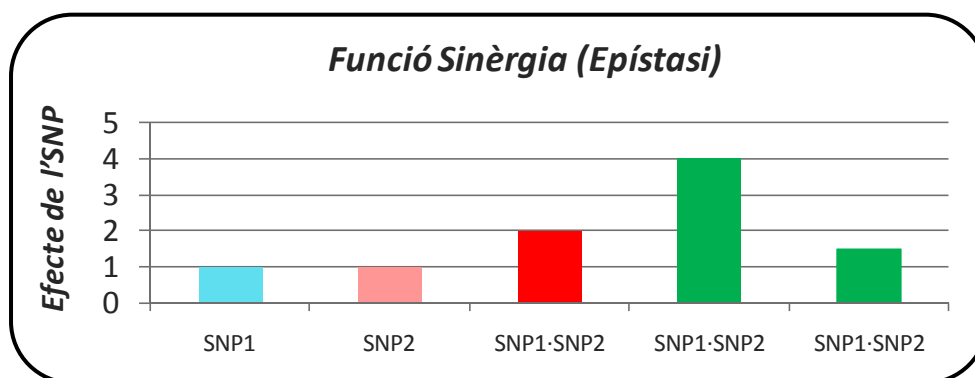


FIGURA 3: Relació de la funció Sinèrgia amb l'efecte individual i conjunt de dos SNPs. En vermell es destaca la interacció quan presenta *Sinèrgia*=0 i en verd quan presenta *Sinèrgia*≠0.

Així doncs, aplicant aquest filtratge es poden trobar interaccions on cap dels dos SNPs havien estat identificats anteriorment. Sobre aquest rànquing s'aplicà l'*MB-MDR*, seguint els mateixos criteris que en l'anàlisi pel **Mètode 1** per tal d'analitzar amb més detall i obtenir la significació de les interaccions prioritzades per la *Sinèrgia*. S'han de tenir en compte les limitacions del mètode: al haver seleccionat el conjunt d'interaccions d'SNPs amb més bon resultat de la funció *Sinèrgia* moltes interaccions no seran avaluades.



#### 6.4. ANÀLISI GLOBAL DELS RESULTATS

Les dades resultants de les dues anàlisis independents, es van posar en comú i es dugué a terme un nou rànquing de les interaccions més significatives segons el *p-valor*, per a avaluar-ne la plausibilitat biològica i el seu paper en el desenvolupament de la carcinogènesi gàstrica.

### 7. RESULTATS

#### 7.1. MÈTODE 1: FILTRATGE D'INTERACCIONS AMB UN SNP SIGNIFICATIU

A partir de la selecció dels 135 SNPs significatius obtinguts prèviament en l'anàlisi univariant d'associació per a la variable resposta GC global, es van explorar **174.555 interaccions entre SNPs** (135 SNPs contra 1.294 menys ells mateixos) amb el paquet *MB-MDR*. Les anàlisis d'interaccions es van fer ajustant per efectes principals per tal d'assegurar que el senyal d'associació que s'obté és degut a una interacció i no pas a l'efecte individual dels SNPs. Només 5 interaccions van presentar un *p-valor* inferior o igual a 0.0001, tal i com es presenta a la **Taula 4**.

**Taula 4:** Resultat de l'anàlisi *MB-MDR* indicant les interaccions amb *p-valor* inferior o igual a  $10^{-4}$ .

NÚM.	SNP1	SNP2	N <sub>H</sub>	OR <sub>H</sub>	<i>p-valor</i>
1	rs2243828	rs2274240	2	1.679170632	<1,00E-04
2	rs2107545	rs2274240	2	1.665124674	<1,00E-04
3	rs2030404	rs6602902	3	1.698252872	<1,00E-04
4	rs1229982	rs3790843	3	1.638203148	<1,00E-04
5	rs8056488	rs315947	3	1.614136276	1,00E-04

Per exemple, a la interacció 1 entre l'SNP *rs2243828* i el *rs2274240*, dues combinacions de genotips es van classificar a la categoria d'alt risc, expressat als paràmetres N<sub>H</sub> (número de combinacions entre els genotips de l'SNP1 i SNP2 a la categoria *High Risk*). Mirant el detall dels resultats de l'*Step 1* del mètode *MB-MDR* es poden conèixer quins genotips conjunts van ser classificats com alt risc i quins com baix risc. En el cas concret de la interacció 1 els resultats es presenten a la **Taula 3.1** (*Annexes, apartat 12.3.1*). La combinació dels genotips A/A i C/C per a l'SNP1 i SNP2 respectivament (homozigots *wildtype*) i la combinació de genotips A/G i C/T per a l'SNP1 i SNP2 (heterozigots), van ser agrupades a la categoria d'alt risc de forma significativa, essent la segona la que va presentar un *p-valor* més significatiu.

Els resultats de l'*Step 2* del mètode *MB-MDR* es troben als *Annexes*, **Taula 4**. Per a la interacció 1, els genotips d'alt risc van resultar els més significatius. L'OR que presenta aquesta categoria d'alt risc equival a 1.68 aproximadament (**OR<sub>H</sub>**, *High risk category*), fet que indica que aquesta combinació de genotips confereix gairebé el doble de risc per al CG global de forma significativa. Les altres interaccions també van ser classificades a la categoria d'alt risc amb OR molt semblants i totes 5 presenten un *p-valor* final inferior o igual a  $10^{-4}$ .

La interacció 2 entre els SNPs *rs2107545* i *rs2274240* va classificar dues combinacions genotípiques com a *High risk* T/T i C/C (homozigot *wildtype*) i T/C i C/T (heterozigots) respectivament, presentats a la **Taula 3.2** (*Annexes, apartat 12.3.1*), que va sortir més significativa amb un **OR<sub>H</sub>** de 1.67, molt semblant a la interacció anterior.

Destacar que la interacció 3 entre *rs2030404* i *rs6602902* va presentar com a significatives les següents combinacions de genotips: A/A i G/G (homozigot *wildtype*), A/G i G/A (heterozigots), **G/G** i **A/A** (homozigots variants) tal i com es pot observar a la **Taula 3.3** (*Annexes, apartat 12.3.1*). Totes les interaccions van ser classificades a la categoria *High risk* amb un **OR<sub>H</sub>** equivalent a 1.70 aproximadament. La interacció entre els homozigots variants va mostrar un *p-valor* inferior.

La interacció entre *rs1229982* i *rs3790843* va presentar les següents combinacions de genotips en la categoria d'alt risc: G/G i G/G (homozigot *wildtype*), G/T i G/A (heterozigots), **T/T** i **A/A** (homozigots variants), respectivament tal i com es mostra a la **Taula 3.4** (*Annexes, apartat 12.3.1*). Van presentar un **OR<sub>H</sub>** molt semblant a les anteriors, d'aproximadament 1.64.

Finalment, la interacció entre *rs8056488* i *rs315947* va classificar les combinacions T/G i C/C (entre l'heterozigot de l'**SNP1** i l'homozigot *wildtype* de l'**SNP2**), T/T i C/T (entre l'homozigot *wildtype* de l'**SNP1** i l'heterozigot de l'**SNP2**) i T/G i T/T (entre l'heterozigot per l'**SNP1** i l'homozigot variant per l'**SNP2**), presentades a la **Taula 3.5** (*Annexes, apartat 12.3.1*).

Pel que fa a l'anàlisi dels gens corresponents als SNPs que apareixen a les interaccions analitzades segons el **Mètode 1** (**Taula 5**), cal destacar que hi ha una interacció entre SNPs de gens que pertanyen a la mateixa via metabòlica (*IL4R* i *IL1RN*) que ha presentat una significació estadística amb *p-valor* inferior a  $10^{-4}$ . A més a més, dues interaccions entre els SNPs dels gens *MPO* i *CDH1* han mostrat significació.

Considerant el rànquing d'interaccions exposat a la **Taula 4** (*Annexes, apartat 12.3.2*) altres interaccions que es troben molt al llindeu del nivell de significació fixat podrien considerar-se per anàlisis posteriors. Un exemple és un SNP pertanyent al gen *PSCA* que apareix com a variant significativa interaccionant amb el miRNA del *Clúster 13\_2* amb un *p-valor* inferior a  $10^{-4}$  en l'*Step 2* de l'*MB-MDR* (**Taula 4**). Tot i això, el seu *p-valor* un cop aplicat l'*MB-MDR* equival a 0.0002.

**Taula 5:** Gens i vies metabòliques on es localitzen els SNPs de les interaccions presentades a la **Taula 3**.

NÚM.	SNP1	SNP2	GEN SNP1	GEN SNP2	VIA METABÒLICA SNP1	VIA METABÒLICA SNP2	p-valor
1	rs2243828	rs2274240	MPO	CDH1	Estrés oxidatiu	Adhesió Cel·lular Resposta a <i>H. pylori</i>	< 1E-04
2	rs2107545	rs2274240	MPO	CDH1	Estrés oxidatiu	Adhesió Cel·lular Resposta a <i>H. pylori</i>	<1E-04
3	rs2030404	rs6602902	XRCC1	GAS6	Processos de reparació del DNA	Transducció de senyal	<1E-04
4	rs1229982	rs3790843	ADH1B	NR5A2 (1q32.1)	Metabolisme de l'Alcohol	Transcripció	< 1E-04
5	rs8056488	rs315947	IL4R	IL1RN	Inflamació Resposta immunitària	Inflamació Resposta immunitària	1E-04

Per altra banda, en les interaccions identificades els SNPs no coincideixen en els subtipus de CG histològics i anatòmics de l'anàlisi previ univariante (**Taula 6**). Cal destacar que les interaccions 1 i 2 es produeixen entre SNPs que van associar-se significativament amb el CG global en el *Single SNP analysis* previ; en canvi, en les interaccions entre SNPs 3-5 només un dels SNPs va associar-se al CG global.

**Taula 6:** Interaccions significatives en l'anàlisi MB-MDR pel **Mètode 1** de filtratge i relació amb l'estudi previ d'associació. Relació de l'associació amb les variables resposta de subtipus histològics i anatòmics.

NÚM.	SNP1	SNP2	CG GLOBAL	CG GLOBAL	SUBTIPUS CG	SUBTIPUS CG	p-valor
			SNP1	SNP2	SNP1	SNP2	
1	rs2243828	rs2274240	Significatiu (C,D,O,A)*	Significatiu (C,R)*	Difús (D,A)*	Intestinal (C,R,A)*	< 1E-04
2	rs2107545	rs2274240	Significatiu (C,D,A)*	Significatiu (C,R)*	Difús (A)* i càrdies (D,A)*	Intestinal (C,R,A)*	<1E-04
3	rs2030404	rs6602902	Significatiu (R)*	No evidència d'associació	No càrdies (C,R,A)*	Intestinal (C,R)* i Difús (R)*	<1E-04
4	rs1229982	rs3790843	Significatiu (O)*	No evidència d'associació	Càrdies (C,D,O,A)*	Difús (O)*	< 1E-04
5	rs8056488	rs315947	Significatiu (C,R,O)*	No evidència d'associació	Intestinal (C,R)*, Difús (C,O)*, No càrdies (C,R)* i Càrdies (C,R)*	No evidència d'associació	1E-04

**Taula 6:** \*Models d'Herència - C: Codominant, D: Dominant, R: Recessiu, O: Overdominant, A: Additiu.

## 7.2. MÈTODE 2: FILTRATGE D'INTERACCIONS SEGONS LA SINÈRGIA

Es van seleccionar les 500 interaccions que van donar una puntuació més alta en l'anàlisi per la *Sinèrgia* i es va procedir a realitzar l'anàlisi per *MB-MDR*.

D'aquestes 500 anàlisis per *MB-MDR*, 19 van donar un  $p\text{-valor} \leq 0.05$ , presentades a la **Taula 6** (*Annexes, apartat 12.6*), i només 4 un  $p\text{-valor}$  inferior a 0.01 (**Taula 7**). Segons els criteris per a seleccionar les possibles interaccions candidates amb un nivell de significació fixat a  $10^{-4}$ , cap dels resultats obtinguts amb el **Mètode 2** supera aquest llindar i, per tant, no es consideraran en el rànquing final d'interaccions, tot i que es discuteixen a continuació.

**Taula 7:** Resultat de l'anàlisi per *MB-MDR* de les 500 interaccions seleccionades aplicant el filtratge pel **Mètode 2** (*Sinèrgia*). Només es mostren les interaccions significatives amb  $p\text{-valor} \leq 0.01$ .

NÚM.	SNP1	SNP2	NH	OR <sub>H</sub>	NL	OR <sub>L</sub>	<i>p</i> -valor
1	rs8050048	rs315947	2	1.46067696	-	-	0.0011
2	rs4239495	rs1443623	2	1.42675124	-	-	0.0021
3	rs2555111	rs10277279	-	-	2	0.55582599	0.0032
4	rs2758320	rs2470144	1	1.71995822	-	-	0.0044

Tal i com es mostra a la **Taula 7**, en la interacció 1 entre els SNPs *rs8050048* i *rs315947* va presentar dues combinacions de genotips classificades a la categoria d'alt risc. Les interaccions entre els genotips G/C i C/C per a l'**SNP1** i **SNP2** respectivament (heterozigot i homozigot *wildtype*) i G/G i C/T (homozigot *wildtype* i heterozigot), presentades a la **Taula 6.1** (*Annexes, apartat, 12.6.1*) van mostrar un OR<sub>H</sub> d'1.46.

La interacció 2 entre l'SNP *rs4239495* i el *rs1443623*, dues combinacions entre els genotips T/C i C/C per a l'**SNP1** i **SNP2** (heterozigot i homozigot *wildtype*) i T/T i C/T per a l'**SNP1** i **SNP2** respectivament (homozigot *wildtype* i heterozigot), tal i com es presenta a la **Taula 6.2** (*Annexes, apartat 12.6.1*), es van classificar a la categoria d'alt risc amb un valor de l'OR<sub>H</sub> equivalent a 1.43.

En la interacció 3 entre l'SNP *rs2555111* i el *rs10277279*, dues combinacions entre els genotips A/A i G/A per a l'**SNP1** i **SNP2** (homozigot *wildtype* i heterozigot) i A/G i A/A per a l'**SNP1** i **SNP2** (heterozigot i homozigot variant) respectivament, es van classificar a la categoria de baix risc, tal i com es presenta a la **Taula 6.3** (*Annexes, apartat 12.6.1*). L'OR<sub>L</sub> que presenta equival a 0.56 aproximadament, fet que indica que confereix gairebé la meitat de risc per al CG global de forma significativa, podent actuar com a factor protector per la malaltia.

Finalment, la interacció 4 entre l'SNP *rs2758320* i el *rs2470144* va presentar una única interacció classificada a la categoria d'alt risc entre els genotips G/T i G/G respectivament (heterozigot i homozigot variant) amb un valor  $OR_H$  de 1.72, indicant que podria conferir el doble de risc a desenvolupar CG, tal i com es presenta a la **Taula 6.4** (Annexes, apartat 12.6.1).

**Taula 8:** Gens i vies metabòliques on es localitzen els SNPs de les interaccions significatives.

NÚM.	SNP1	SNP2	GEN SNP1	GEN SNP2	VIA METABÒLICA SNP1	VIA METABÒLICA SNP2	p-valor
1	<i>rs8050048</i>	<i>rs315947</i>	<i>ILR4</i>	<i>IL1RN</i>	Inflamació Resposta immunitària	Inflamació Resposta immunitària	0.0011
2	<i>rs4239495</i>	<i>rs1443623</i>	clúster 19_3 (miRNA)	<i>MBD1</i>	Regulació miRNA	Expressió gènica metil·lació	0.0021
3	<i>rs2555111</i>	<i>rs10277279</i>	<i>miR-10a</i>	desert-hCG_20237 76 (GWAs)	Regulació miRNA	Desconeguda	0.0032
4	<i>rs2758320</i>	<i>rs2470144</i>	<i>SOD2</i>	<i>CYP19A1</i>	Oxidació - Reducció	Metabolisme de Carcinògens	0.0044

Cal destacar que segons la correspondència amb els gens i les vies metabòliques (**Taula 8**) que la primera interacció entre SNPs dels gens de *ILR4* i *IL1RN* (*rs8050048-rs315947*) també és significativa en la primera anàlisi pel **Mètode 1**.

**Taula 9:** Interaccions significatives en l'anàlisi MB-MDR pel **Mètode 2** de filtratge (*Sinèrgia*) i relació amb l'estudi previ d'associació. Relació de l'associació amb les variables resposta de subtipus histològics i anatòmics.

NÚM.	SNP1	SNP2	CG GLOBAL SNP1	CG GLOBAL SNP2	SUBTIPUS CG SNP1	SUBTIPUS CG SNP2	p-valor
1	<i>rs8050048</i>	<i>rs315947</i>	Significatiu (C,R,O)*	No evidència d'associació	Intestinal (C,R)*, Difús (C,O)*, No càrdies (C,R)* i Càrdies (C,R)*	No evidència d'associació	0.0011
2	<i>rs4239495</i>	<i>rs1443623</i>	No evidència d'associació	No evidència d'associació	No evidència d'associació	No evidència d'associació	0.0021
3	<i>rs2555111</i>	<i>rs10277279</i>	Significatiu (C,O)*	No evidència d'associació	Difús (O)* i No càrdies (O)*	No evidència d'associació	0.0032
4	<i>rs2758320</i>	<i>rs2470144</i>	No evidència d'associació	No evidència d'associació	Càrdies (D,A)*	No evidència d'associació	0.0044

**Taula 9:** \*Models d'Herència - C: Codominant, D: Dominant, R: Recessiu, O: Overdominant, A: Additiu.

En les interaccions identificades els SNPs no coincideixen en els subtipus de CG histològics i anatòmics del *Single SNP analysis* (Taula 9). 5 dels SNPs que mostren interacció no s'associaven amb el CG en general ni amb cap dels seus subtipus histològics i anatòmics.

### 7.3. RESULTAT FINAL DE L'ANÀLISI D'INTERACCIONS

Finalment, el rànquing de les interaccions més significatives ( $p\text{-valor}\leq 10^{-4}$ ) entre gens es presenten a la **Taula 10**.

**Taula 10:** Gens de les interaccions més significatives en l'anàlisi MB-MDR i el *Permutation Test* del **Mètode 1**.

<b><i>MPO x CDH1</i></b>
<b><i>XRCC1 x GAS6</i></b>
<b><i>ADH1B x NR5A2</i></b>
<b><i>IL4R x IL1RN</i></b>

Aquestes 4 interaccions, obtingudes a partir de l'anàlisi aplicant el **Mètode 1**, són les més bones candidates per a validar-les en anàlisis independents.

Tenint en compte els resultats del **Mètode 2** aplicant el filtre segons la mesura *Sinèrgia*, encara que cap interacció superés el llindar de significació establert cal mencionar que amb un  $p\text{-valor}$  entre 0.0040 i 0.0011 s'han obtingut 4 interaccions entre els gens *IL4R* i *IL1RN*, *clúster 19\_3 (miRNA)* i *MBD1*, *miR-10a* i *desert-hCG\_202377*, i *SOD2* i *CYP19A1*. La interacció entre el gen *IL4R* i *IL1RN* apareix dins les més significatives, fet que reforça la hipòtesi que podria tenir un paper important en la carcinogènesi gàstrica.

Per tal de ser més exhaustius amb els resultats obtinguts a la **Taula 4 (Annexes, apartat 12.3.2)** es presenta un rànquing de 47 interaccions que a l'*Step 2* de l'anàlisi *MB-MDR* aplicat al **Mètode 1** van presentar un  $p\text{-valor}$  inferior a  $10^{-3}$  que no van ser avaluades amb el *Test de Permutacions (Step 3 MB-MDR)* però que en anàlisis posteriors podrien estudiar-se ja que també podrien sortir resultats interessants.

## 8. DISCUSSIÓ DELS RESULTATS

Amb les dues metodologies de filtratge per tal de seleccionar les interaccions entre SNPs més bons candidats, s'han obtingut 5 interaccions entre SNPs diferents (4 interaccions entre gens diferents) amb *p*-valor inferior a  $10^{-4}$ , entre SNPs dels gens *MPO* i *CDH1*, *XRCC1* i *GAS6*, *ADH1B* i *NR5A2* i *IL4R* i *IL1RN*, totes elles obtingudes amb l'anàlisi realitzat pel **Mètode 1** on es consideraven interaccions on com a mínim un dels dos SNPs que forma la interacció era significatiu a nivell del *Single SNP analysis*.

La interacció entre *IL4R* i *IL1RN* ha estat també significativa en l'anàlisi independent de filtratge segons la *Sinèrgia* tot i que amb un *p*-valor superior al nivell de significació fixat.

### 8.1. COMPARACIÓ AMB ANÀLISIS PRÈVIES

Dels gens candidats obtinguts amb el mètode *MB-MDR* i el filtratge segons el **Mètode 1** en l'anàlisi d'interaccions, tots els gens excepte *CDH1* no van ser prioritzats en l'anàlisi univariant previ ja que cap d'ells va sortir significatiu en el *Gene-based Permutation Test*.

En altres estudis com els que surten a la base de dades *Gene Prospector* apareixen associats tots els gens candidats proposats anteriorment excepte *GAS6* i *NR5A2* (gen seleccionat per contenir un SNP associat a càncer en un GWAS previ). Per tant, també seria interessant poder validar-los ja que tampoc havien estat prioritzats en l'anàlisi univariant previ.

El gen *PSCA* que, tal i com s'ha comentat a l'apartat **3. 2 d'Antecedents**, va ser el que va donar la millor significació en les anàlisis estadístiques prèvies, també apareix associat en aquesta anàlisi d'interaccions, si bé es troba al llindar del nivell de significació de l'anàlisi.

Per altra banda, cal mencionar que la interacció entre variants dels gens *IL4R-IL1RN* es manté significativa en les dues anàlisis independents fetes, segons diferents mètodes de filtratge dels gens candidats.

### 8.2. PLAUSIBILITAT BIOLÒGICA DE LES INTERACCIONS IDENTIFICADES

De les 5 interaccions entre SNPs diferents obtingudes un cop aplicat l'*MB-MDR* a les dades filtrades segons la primera metodologia aplicada descrita anteriorment, es pot considerar la següent valoració respecte la seva possible plausibilitat biològica, segons informació obtinguda en diferents Bases de Dades i a la literatura:

### INTERACCIÓ *MPO* - *CDH1*

La interacció entre SNPs dels gens *MPO* (*Myeloperoxidase*) i *CDH1* (*Cadherina-1*), inclosos a l'estudi com a gens candidats, va presentar un *p*-valor inferior a  $10^{-4}$ , suggerint que poden tenir un paper important en la carcinogènesi gàstrica. No obstant, les interaccions es produeixen entre dos SNPs diferents del gen *MPO* (*rs2243828* i *rs2107545*) i un SNP de *CDH1* (*rs2274240*). Si s'examina el *Desequilibri de Lligament* (*Linkage Disequilibrium – LD*) entre els dos SNPs del gen *MPO* s'observa que efectivament es troben en fort desequilibri de lligament ( $D'=1$ ,  $r^2=0.727$  i  $LOD=20$ ) i, per tant, en realitat podrien reflectir la mateixa interacció.

El gen *MPO* forma part dels sistema de defensa de l'hoste en leucòcits polimorfonucleats i és responsable de l'activitat microbicida contra els microorganismes infecciosos en processos catabòlics del peròxid d'hidrogen. *MPO* catalitza la producció d'àcids hipohàlics, principalment l'àcid hipoclorós en determinades situacions fisiològiques, i també altres intermediaris tòxics que augmenten l'activitat microbicida dels leucòcits polimorfonucleats. També participa en processos de regulació negativa de processos apoptòtics i en l'eliminació de radicals superòxids.

Per altra banda, el gen *CDH1* forma part de la família de proteïnes d'adhesió cel·lular calci-depenents. Interactuen preferiblement entre elles homofílicament, però també poden contribuir en l'ordenament de diferents tipus cel·lulars heterogenis. Està involucrat en diferents mecanismes de regulació de l'adhesió cèl·lula-cèl·lula, mobilitat i proliferació de cèl·lules epitelials. Té un paper de supressió d'invasió molt important i és un lligand de les integrines *alpha-E/beta-7*. Se sap que defectes en *CDH1* són una de les causes principals del càncer gàstric hereditari tipus difús, caracteritzat per lesions infiltrades pobrament diferenciades (16).

Els dos gens indirectament tenen un paper molt important en la resposta del procés inflamatori que pot desencadenar la infecció per *Helicobacter pylori* ja que per una banda *MPO* està involucrat en el procés de defensa de la immunitat innata de microorganismes i *CDH1* participa en els mecanismes de regulació d'adhesió, mobilitat i proliferació de cèl·lules epitelials.

### INTERACCIÓ *XRCC1* - *GAS6*

*XRCC1* (*X-ray repair cross-complementing protein 1*) forma part del complex reparador de DNA per talls d'una sola cadena (*Base Excision Repair - BER*) en cèl·lules de mamífer junt amb la beta polimerasa ( $\beta$ -Pol), la *Lligasa III* (n'és el cofactor) i *PARP* (*Poli ADP Ribosa Polimerasa*). Aquesta via s'ha relacionat amb diferents tipus de neoplàsies a la literatura amb evidència a nivell experimental i, a més, s'ha demostrat que juga un paper molt important en la via de reparació de danys al DNA provocats en el procés inflamatori induït per la infecció per *Helicobacter pylori* (7, 21).



Per altra banda, *GAS6* (*Growth arrest-specific gene 6*) és un factor potenciador del creixement dependent de la vitamina K. S'ha vist que el bloqueig de la seva activitat frena la proliferació cel·lular, suggerint que la seva activitat pot tenir un paper important en el CG (13,22, 36).

La interacció d'aquests dos gens, inclosos a l'estudi com a gens candidats, podria basar-se en el fet que una desregulació en *XRCC1* que forma part de la maquinària de reparació de DNA de la cèl·lula podria condicionar un mal funcionament de *GAS6* i, per tant, augmentar el risc a desenvolupar tumorigènesi.

### **INTERACCIÓ *ADH1B* - *NR5A2***

Pel que fa al polimorfisme que pertany al gen del metabolisme de l'alcohol (*Alcohol deshidrogenases: subtipus B1*), a la literatura i a diferents bases de dades<sup>3,4</sup> ja s'ha relacionat amb càncers i també els altres subtipus A i C, i s'ha reconegut que formen part de mecanismes de la carcinogènesi química del càncer; de fet, van ser inclosos a l'estudi EurGast com a gens candidats. La seva funció biològica consisteix en catalitzar l'última etapa del metabolisme de l'etanol, la oxidació de l'alcohol a acetaldehid.

Per altra banda, el gen *NR5A2* (*Nuclear Receptor Subfamily 5, group A – member 2*) també conegut com *Liver Receptor Homolog-1* (*LRH-1*) pertany a la família de receptors nuclears de factors de transcripció intracel·lulars. Aquest gen va ser inclòs a l'estudi pel fet que SNPs del mateix s'havien associat a altres càncers epitelials, com el de pàncrees, en un GWAS preliminar (19, 32). La seva activitat és en la regulació del desenvolupament, transport de colesterol, homeòstasi de les sals biliars i esteroidogènesi, tot i que se sap que també juga un paper molt important en el manteniment de la pluripotència de les cèl·lules mare durant el desenvolupament embrionari (19, 32). La desregulació d'aquest tipus de vies relacionades amb el manteniment i regulació del desenvolupament podria estar relacionada amb la carcinogènesi, tal i com s'ha observat en diferents estudis que relacionen *LRH-1* amb diferents tipus de càncer, com el de mama, pancreàtic, i també càncer gàstric.

### **INTERACCIÓ *IL4R* - *IL1RN***

En la interacció entre gens de la *IL4R* (*Interleukin 4 Receptor*) i *IL1RN* (*Interleukin 1 Receptor agonist*) ambdós gens van ser inclosos a l'estudi com a gens candidats i, de fet, a diverses bases de dades com la *Gene Prospector*, i en molts estudis han associat amb CG (14, 34). En concret, l'associació d'un VNTR de *IL1RN* amb el CG va ser una de les primeres en ser identificades (12, 20). El receptor de les Interleuquines 4 i 13 (*IL4R*) participa en la via de transducció de senyal *Jak-STAT* i *PI3K-Akt*.

<sup>3</sup> Kyoto Encyclopedia of Genes and Genomes (KEGG) Database: [www.genome.jp/kegg/](http://www.genome.jp/kegg/)

<sup>4</sup> Online Mendelian Inheritance in Man (OMIM) Database: [www.omim.org/](http://www.omim.org/) - #613659, Gastric Cancer (ERBB2).

Les respostes immunitàries mediades per la unió amb els seus lligands, *IL4* i *IL13*, regulen la producció d'IgE, citoquines, mucus i la diferenciació de la subpoblació de Limfòcits T a *TH2*. S'ha identificat que la *IL4R* soluble inhibeix la proliferació cel·lular mediada per *IL4* i participa en la sobreexpressió d'*IL5* per LT (*GeneCards*).

L'antagonista del receptor de la Interleuquina 1 (*IL1RN/IL1a*) se sap que inhibeix l'activitat de la *IL1* unint-se al *IL1R* per competició i prevé la seva associació amb el co-receptor *IL1RAP* per senyalització. Està predominantment expressat en cèl·lules epitelials. S'ha relacionat amb la via de resposta de fase aguda (*acute phase response*) en la que es produeix una resposta inflamatòria molt acusada davant infeccions microbianes. Se sap que mutacions en el gen *IL1RN* donen lloc a la producció d'una proteïna truncada i no secretada donant lloc a una inflamació severa (15, 38).

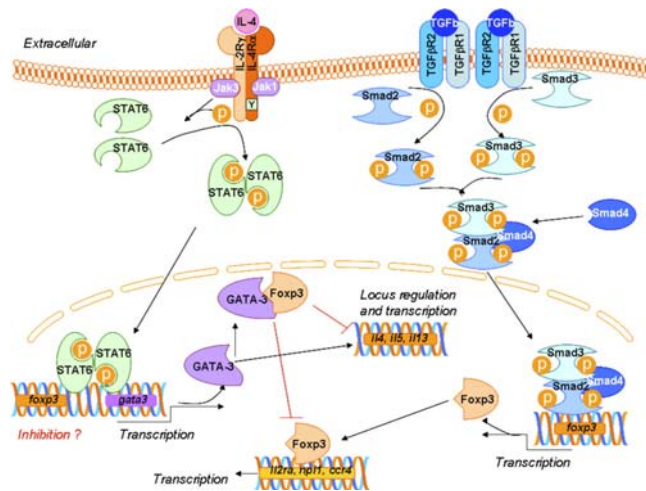


FIGURA 4: Via de senyalització en Limfòcits T mediada per IL4 i IL4R. CHAPOVAL, S; DASGUPTA, P; DORSEY, NJ; KEEGAN, AD. "Regulation of the T helper cell type 2 (Th2)/T regulatory cell (Treg) balance by IL-4 and STAT6". *Journal Leukocyte Biology*, 87(6) (2010): 1011-1018, doi: 10.1189/jlb.1209772.

En estudis posteriors es podria validar si aquests dos gens mantenen la seva funcionalitat degut a la seva interacció i que si un dels dos muta l'altre perdi afinitat o provoqui més susceptibilitat a una inflamació severa. També si hi ha possibilitat que *IL1RN* pugui unir-se a *IL4R* per activar la transducció de senyal en una via alternativa d'activació de la resposta immunitària. Tot i això, la interacció sembla que biològicament podria ser plausible i podria explicar un dels mecanismes de senyalització alterats durant la carcinogènesi gàstrica.

**Interacció PSCA-GPC5**

La interacció entre *PSCA* (*Prostate Stem Cell Antigen*) i el *miRNA* del Clúster 13\_2, que va presentar un *p-valor* molt al llindar del nivell de significació, podria suggerir que el *miRNA* de prop del gen *GPC5* (*Glypican - 5*) pot tenir un paper regulador en l'expressió de *PSCA*, que va ser inclòs a l'estudi com a gen candidat ja que en estudis previs s'ha associat amb el desenvolupament de la carcinogènesi gàstrica (29). En estudis futurs seria molt interessant validar si aquest *miRNA* pot tenir un paper important en la regulació de l'expressió de *PSCA* i s'hi s'han associat en altres estudis o a diferents Bases de Dades de *miRNA* com *miRBase*.

## 9. CONCLUSIONS

- Els resultats del projecte han mostrat que **5** interaccions diferents entre SNPs dels gens *MPO* i *CDH1*, *XRCC1* i *GAS6*, *ADH1B* i *NR5A2*, *IL4R* i *IL1RN* s'associen significativament amb alt risc de desenvolupar CG, amb *p-valor* inferior a  $10^{-4}$ . Aquestes interaccions són les millors candidates a estar associades amb la carcinogènesi gàstrica.
- Excepte *CDH1*, cap dels altres gens que intervenen en aquestes interaccions s'havia associat significativament (*Gene-based Permutation Test*) amb el CG en les anàlisis prèvies. Aquest estudi ha permès, doncs, seleccionar 7 altres gens que podrien ser importants en l'estudi del CG.
- Amb l'anàlisi de les 500 interaccions seleccionades un cop aplicat el filtratge segons la funció *Sinèrgia (Mètode 2)* només 4 interaccions entre els gens *IL4R* i *IL1RN*, *microRNA Clúster 19\_3* i *MBD1*, *HOXB3* i *desert-hCG\_2023776* (provinent d'un GWAs) i finalment *SOD2* i *CYP19A1*, van mostrar un *p-valor* inferior a 0.01, tot i que no van superar el llindar de significació.
- La interacció entre SNPs dels gens de *IL4R* i *IL1RN* va mostrar una associació significativa amb càncer gàstric en dues anàlisis independents, suggerint que té un paper important en els mecanismes de resposta inflamatòria en el procés de carcinogènesi gàstrica.
- En l'anàlisi d'interaccions el gen *PSCA* ha mostrat una interacció amb un SNP d'un miRNA del *Clúster 13\_2* (regió del gen *GPC5*) amb un *p-valor* de 0.0002, superior al llindar de significació establert. Tot i això, es tracta d'un gen que a més de presentar molta significació en el *Single SNP analysis* també va mantenir l'associació amb CG en el *Gene-based Permutation Test*, suggerint que pot tenir un paper molt important en el procés de desenvolupament de CG i que a més pot interaccionar amb altres factors genètics conferint més susceptibilitat a la malaltia.
- Globalment podem concloure que l'anàlisi d'interaccions ha permès identificar polimorfismes i gens que podrien estar implicats en el risc de càncer gàstric i que no s'havien identificat anteriorment amb les anàlisis univariants.
- També remarcar que cal ser prudent amb els resultats obtinguts i els possibles falsos positius que es poden obtenir amb una anàlisi com la d'interaccions on es realitzen una gran quantitat de proves d'hipòtesis. Tot i que es realitzin ajustos estadístics dels *p-valors* per proves múltiples, només una validació posterior en una mostra independent permetrà confirmar si les interaccions identificades tenen un efecte real sobre el risc de càncer gàstric.

## 10. EL FUTUR

Tots els resultats exposats en aquest *Treball de Final de Grau* són un punt de partida de cara a realitzar validacions posteriors de les dades, mitjançant la seva anàlisi en estudis independents de casos i controls, i ampliar la informació per avaluar la susceptibilitat real dels gens i variants identificats en la cohort europea **EPIC** sobre el desenvolupament del càncer gàstric.

Un cop analitzats i replicats els resultats obtinguts, futurs estudis podrien enfocar-se en l'anàlisi de desequilibri de lligament, seqüenciació, estudis funcionals *in silico* i assajos en el laboratori per a la identificació de les variants causals de les associacions i la validació experimental de l'expressió diferencial d'aquestes proteïnes en línies tumorals vers teixit sa i així poder avaluar-ne la rellevància funcional.

Paral·lelament, es podrien realitzar altres anàlisis independents per a identificar el perfil genètic de risc de CG a partir dels polimorfismes estudiats per seleccionar el conjunt que millor prediu la malaltia utilitzant el mètode *Area Under the ROC Curve Random Forest (AUC-RF)* desenvolupat per *Calle et al. any 2011* (4).

## 11. REFERÈNCIES

1. AGUDO, A; BONET, C; SALA, N; *et al.* "Hemochromatosis (HFE) gene mutations and risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) study". *Carcinogenesis*, (2013): 1-7, doi: 10.1093/carcin/bgt045.
2. BINGHAM, S. and RIBOLI, E. "Diet and Cancer – The European Prospective Investigation into Cancer and Nutrition". *Nature Reviews | Cancer*, 4 (2004): 206-215, doi:10.1038/nrc1298.
3. BUSLER, V. J; TORRES, V. J; MCCLAIN, M. S; TIRADO, O; FRIEDMAN, D. B; COVER, T. L. "Protein-Protein Interactions among *Helicobacter pylori* Cag Proteins". *Journal of Bacteriology*, 188(13) (2006): 4787-4800, doi:10.1128/JB.00066-06.
4. CALLE, ML; URREA, V, BOULESTEIX AL, MALATS N. "AUC-RF: A new strategy for genomic profiling with Random Forest". *Human Heredity*, 72 (2011): 121-132, doi: 10.1159/000330778.
5. CALLE, ML; URREA, V; VELLALTA, G; MALATS, N; STEEN, K. V. "Improving strategies for detecting genetic patterns of disease susceptibility in association studies". *Statistics in Medicine*, 27 (2008); 6532–6546, doi: 10.1002/sim.3431.
6. CALLE, ML; URREA, V; MALATS, N; STEEN, K. V. "MBMDR: An R package for exploring gene-gene interactions associated with binary or quantitative traits". *Bioinformatics*, 26(17) (2010): 2198-2199, doi:10.1093/bioinformatics/btq352.
7. CAMPALANS, A; MARSIN, S; NAKABEPPU, Y; O'CONNOR, T. R; BOITEUX, S; RADICELLA, P. J. "XRCC1 interactions with multiple DNA glycosylases: a model for its recruitment to base excision repair". *DNA Repair*, 4(7) (2005): 826-835, doi: 10.1016/j.dnarep.2005.04.014.
8. CATTART, T; CALLE, ML; DUDEK, S; MAHACHIE, JM; VAN LISHOUT, F; URREA, V; RITCHIE, MD; VAN STEEN, K. "Model-Based Multifactor Dimensionality Reduction for detecting gene-gene interactions in case-control data in the absence and presence of noise". *Annals of Human Genetics*, 75 (2011): 78-89, doi: 10.1111/j.1469-1809.2010.00604.x.
9. CATTART, T; URREA, V; NAJ, AC; DE LOBEL, L; DE WIT, V; FU, M; MAHACHIE, JM; SHEN, H; CALLE, ML; RITCHIE, MD; EDWARDS, TL; VAN STEEN, K. "FAM-MDR: a flexible family-base

- multifactor dimensionality reduction technique to detect epistasis using related individuals*". Public Library of Science ONE, 5(4), (2010): 1-15, doi: 10.1371/journal.pone.0010304.
10. DA COSTA, DM; NEVES-FILHO, EH; ALVES, MK; RABENHORST, SH. *"Interleukin polymorphisms and differential methylation status in gastric cancer: an association with Helicobacter pylori infection"*. Epigenomics, 5(2) (2013):167-75, doi: 10.2217/epi.13.7.
  11. DUDBRIDGE, F AND KOELEMAN, BPC. *"Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies"*. American Journal of Human Genetics, 75 (2004); 424-435.
  12. EL-OMAR, EM; CARRINGTON, M; CHOW, WH; MCCOLL, KEL; BREM, JH; YOUNG, HA *et al.* *"Interleukin-1 polymorphisms associated with increased risk of gastric cancer"*. Nature, 404 (6776) (2000):398-402.
  13. FERNÁNDEZ, L; BELLIDO, L; GARCÍA, P. *"Growth arrest-specific gene 6 (GAS6). "An outline of its role in haemostasis and inflammation"*. Thrombosis and Haemostasis, 100 (4) (2008): 604-610, doi: 10.1160/TH08-04-0253.
  14. FRAZER, K. A; MURRAY, S. S; SCHORK, N. J; TOPOL, E. J. *"Human genetic variation and its contribution to complex traits"*. Nature Reviews | Genetics, 10 (2009): 241-251, doi:10.1038/nrg2554.
  15. GABAY, C. and PALMER, G. *"Mutations in the IL1RN locus lead to autoinflammation"*. Nature Reviews Rheumatology, 5(9) (2009): 480-482, doi:10.1038/nrrheum.2009.177.
  16. GONZÁLEZ, C. A. and AGUDO, A. *"Carcinogenesis, prevention and early detection of gastric cancer: were we are and where we should go"*. International Journal of Cancer, 130 (2012): 745-753, doi: 10.1002/ijc.26430.
  17. GONZÁLEZ, JR; ARMENGOL, LL; SOLÉ, X; GUINÓ, E; MERCADER, JM; ESTIVILL, X; MORENO V. *"SNPassoc: an R package to perform whole genome association studies"*. Bioinformatics, 23 (2007): 644-645.
  18. GRAZIANO, F; HUMAR, B; GUILFORD, P. *"The role of the E-cadherin gene (CDH1) in diffuse gastric cancer susceptibility: from the laboratory to clinical practice"*. Annals of Oncology Official Journal of the

- European Society for Medical Oncology ESMO, 14(12) (2003): 1705-1713, doi: 10.1093/annonc/mdg486.
19. HUANG, H; CHANDA, P; ALONSO, A; BADER, JS; ARKING DE. "Gene-based tests of association". *PLoS Genet*, 7(7) (2011): 1-15 (e1002177), doi: 10.1371/journal.pgen.1002177.
  20. LAO-SIRIEIX, P; CALDAS, C; FITZGERALD, RC. "Genetic predisposition to gastro-oesophageal cancer". *Current opinion in genetics & development*, 20(3) (2010):210-217.
  21. LAZARUS, K. A; WIJAYAKUMARA, D; CHAND, A. L; SIMPSON, E. R; CLYNE, C. D. "Therapeutic potential of Liver Receptor Homolog-1 modulators". *The Journal of Steroid Biochemistry and Molecular Biology*, 130(3-5) (2012): 138-146, doi: 10.1016/j.jsbmb.2011.12.017.
  22. LI, W; ZHANG, L; MA, J; ZHANG, Y. et al. "Association between genetic variants of DNA base excision repair genes and evolution of precancerous gastric lesions in a Chinese population". *Carcinogenesis*, 30(3) (2009): 500-505, doi:10.1093/carcin/bgp018.
  23. LUCH, A. "Nature and nurture – lessons from chemical carcinogenesis". *Nature Reviews | Cancer*, 5 (2005):113-25, doi:10.1038/nrc1546.
  24. MATTAR R, MARQUES SB, DOS SANTOS AF, DO SOCORRO MONTEIRO M, IRIYA K, CARRILHO FJ. "A possible role of IL-1RN gene polymorphism in the outcome of gastrointestinal diseases associated with H. pylori infection". *Clinical and Experimental Gastroenterology*, 6 (2013):35-41. doi: 10.2147/CEG.S42260.
  25. MILNE, A. N; CARNEIRO, F; O'MORAIN, C; OVERHAUS, G. A. "Nature meets nurture: molecular genetics of gastric Cancer". *Human Genetics*, 126 (2009): 615–628, doi: 10.1007/s00439-009-0722-x.
  26. MOORE JH, WILLIAMS SM. "Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis". *Bioessays* (2005); 27:637–646.
  27. RIBOLI, E and HUNT, KJ. et al. "European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection". *Public Health Nutrition*, 5 (2002): 1113–1124, doi: 10.1079/PHN2002394.
  28. RITCHIE, MD; HAHN, LW; ROODI, N; BAILEY, LR; DUPONT, WD; PARL, FF and MOORE, JH. "Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer". *American Journal of Human Genetics*, 69 (2001): 138-147.

29. SALA, N; MUÑOZ, X; TRAVIER, N; AGUDO, A; DUELL, E.J. et al. "Prostate Stem Cell Antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: results from the EPIC-EurGast study". *International Journal of Cancer*, 130(10) (2012):2417-2427, doi:10.1002/ijc.26243.
30. SCHOTTENFELD, D; BEEBE-DIMMER, J. L. "Advances in cancer epidemiology: understanding causal mechanisms and the evidence for implementing interventions". *Annual Review of Public Health*, 26 (2005):37-60, doi: 10.1146/annurev.publhealth.26.021304.144402.
31. SMITH, M.G; HOLD G.L; TAHARA, E; EL-OMAR E.M. "Cellular and molecular aspects of gastric cancer". *World Journal Gastroenterology*, 12 (2006):2979-2990.
32. TANG, CS; FERREIRA, MA. "A gene-based test of association using canonical correlation analysis". *Bioinformatics*, 28(6) (2012):845-50, doi: 10.1093/bioinformatics/bts051.
33. VARADAN, V; MILLER, DM; ANASTASSIOU, D. "Computational inference of the molecular logic for synaptic connectivity in *C. elegans*". *Bioinformatics*, 22 (2006): 497–506.
34. WANG, SH; ZHENG, DZ; LAN, FH; DENG, XJ; ZENG, J; LI, CJ; WANG, R; ZHU, ZY. "Increased expression of *hLRH-1* in human gastric cancer and its implication in tumorigenesis". *Molecular and Cellular Biochemistry* (2008). Volume: 308, Issue: 1-2, Pages: 93-100.
35. YAMASHITA, K; SAKURAMOTO, S and WATANABE, M. "Genomic and Epigenetic Profiles of Gastric Cancer: Potential Diagnostic and Therapeutic Applications". *Surgery Today*, 41 (2011): 24–38, doi:10.1007/s00595-010-4370-5.
36. YANAGITA, M. "The role of the vitamin K-dependent growth factor *Gas6* in glomerular pathophysiology". *Current Opinion in Nephrology and Hypertension*, 13(4) (2004): 465-470.
37. YUSASA, Y. "Control of gut differentiation and intestinal-type gastric carcinogenesis". *Nature Reviews | Cancer*, 3 (2004): 592-600, doi: 10.1038/nrc1141.
38. ZHANG Y, LIU C, PENG H, ZHANG J, FENG Q. IL1 receptor antagonist gene IL1-RN variable number of tandem repeats polymorphism and cancer risk: a literature review and meta-analysis. *PLoS One*. 2012;7(9):e46017. doi:10.1371/journal.pone.0046017.



# **ANNEXES**

## 12. ANNEXES

### 12.1. MODEL-BASED MULTIFACTOR DIMENSIONALITY REDUCTION METHOD

El paquet *MB-MDR* (*Model-Based Multifactor Dimensionality Reduction Method*) proposat per *Calle et al. (2008)* (8) explora les possibles interaccions gen-gen en tres passos per identificar genotips multi-locus que poden estar associats amb el fenotip d'interès (Càncer Gàstric), i permet ajustar efectes marginals i de confusió.

La funció principal és *mbmdr* i la seva crida és la següent:

```
mbmdr(y,data,order,covar=NULL,exclude=NA, risk.threshold=0.1, output=NULL,  
adjust=c("none","covariates","main.effects","both"),  
first.model=NULL,list.models=NULL,use.logistf=TRUE,printStep1=FALSE,...)
```

S'indica el vector que conté la variable dependent o resposta, el *data frame* amb la informació dels SNPs codificats amb 0 (genotip homozigot *wildtype*), 1 (genotip heterozigot per l'al·lel variant) i 2 (genotip homozigot per l'al·lel variant), l'ordre de les interaccions entre els SNPs.

En el primer pas del mètode *MB-MDR*, cada combinació entre genotips dels dos SNPs és avaluat segons la seva associació amb la variable resposta, i es classifica segons alt risc, baix risc o no significatiu (categoria que reflecteix la no evidència d'associació), i tots els genotips de la mateixa classe es mesclen. El llindar de significació correspon a  $\alpha=0.1$  per defecte.

En el següent pas s'agrupen les combinacions entre SNPs significatives per categories, alt risc o baix risc, i es torna a dur a terme un altre test d'associació (*Wald Statistic Test*). El resultat aporta el paràmetre *Wald Statistic* per ambdues categories.

Finalment, s'explora la significació de les interaccions entre genotips de dos SNPs a través del test de permutació per models específics d'interacció (*Permutation Test*) en el màxim *Wald statistic* obtingut en l'anàlisi *MB-MDR*.

## 12.2. SCRIPTS APLICATS PER A L'ANÀLISI DEL MÈTODE 1 AMB R

**ANNEXES Taula 2:** *Scripts i funcions utilitzades en l'anàlisi MB-MDR del Mètode 1.*

<b>CARREGAR PAQUETS</b>	<code>library(SNPassoc)</code> <code>library(mbmdr)</code>
<b>LECTURA DE LA BASE</b>	<code>db&lt;-read.csv("eurgast_uncond_snp2011(2).csv",sep=";",na.string=c("</code>
<b>DE DADES</b>	<code>",","NA"),as.is=TRUE,header=TRUE)</code>
<b>CODIFICACIÓ</b>	<code>db\$COUNTRY&lt;-factor(db\$COUNTRY, levels=c(1, 2, 3, 4, 5, 6, 7, 8, 9,</code>
<b>VARIABLES</b>	<code>"B"),labels=c("France","Italy","Spain","UK","Holland","Greece","Germany","Sweedeen","Denmar k","Norway"))</code> <code>db\$SEX&lt;-factor(db\$SEX,levels=c(1,2),labels=c("Male","Female"))</code> <code>db\$TFF2_C_68_69INSCCTT&lt;-sapply(db\$TFF2_C_68_69INSCCTT,function(x) ifelse(x=="-/ ","11",x))</code> <code>db\$TFF2_C_68_69INSCCTT&lt;-sapply(db\$TFF2_C_68_69INSCCTT,function(x) ifelse(x=="-/ /CTT","12",x))</code> <code>db\$TFF2_C_68_69INSCCTT&lt;-as.integer(db\$TFF2_C_68_69INSCCTT)</code> <code>db\$IL1RN_VNTR &lt;-sapply(db\$IL1RN_VNTR,function(x) ifelse(x=="13","11",x))</code> <code>db\$IL1RN_VNTR &lt;-sapply(db\$IL1RN_VNTR,function(x) ifelse(x=="14","11",x))</code> <code>db\$IL1RN_VNTR &lt;-sapply(db\$IL1RN_VNTR,function(x) ifelse(x=="15","11",x))</code> <code>db\$IL1RN_VNTR &lt;-sapply(db\$IL1RN_VNTR,function(x) ifelse(x=="23","12",x))</code> <code>db\$IL1RN_VNTR &lt;-sapply(db\$IL1RN_VNTR,function(x) ifelse(x=="24","12",x))</code> <code>db\$IL1RN_VNTR &lt;-as.integer(db\$IL1RN_VNTR)</code> <code>names(db)[(which(names(db)=='ERBB2_rs1136201'))]&lt;-'rs1136201'</code> <code>names(db)[(which(names(db)=='FUT4_rs2230274'))]&lt;-'rs2230274'</code> <code>names(db)[(which(names(db)=='IL1RN_VNTR'))]&lt;-'rs2234663'</code> <code>names(db)[(which(names(db)=='TFF1c_41'))]&lt;-'rs13276'</code> <code>DB_p&lt;- db[,c(32:689,691:1128,1130:1321,1327:1329,1331:1334)]</code>
<b>ÚS SNPASSOC</b>	<code>db.ok&lt;- setupSNP(data=DB_p,colSNPs=c(1:1287,1289:length(DB_p)),sep="")</code>
<b>CREACIÓ BASE DE</b>	<code>db.ok2&lt;-as.data.frame(lapply(db.ok[,2:length(db.ok)],additive))</code>
<b>DADES DE TREBALL</b>	<code>db.ok3&lt;-as.data.frame(lapply(db.ok2[,1:length(db.ok2)],as.factor))</code> <code>SIG_db&lt;-</code> <code>db.ok3[,c(14,15,17,39,52,54,59,60,81,82,85,89,96,100,101,105,125,130,135,137,144,158,165,1</code> <code>66,168,200,213,216,221,225,227,233,241,247,261,269,293,294,296,298,306,312,329,331,335,3</code> <code>37,344,346,367,377,388,406,415,425,431,440,454,463,467,507,514,525,536,537,545,553,560,5</code> <code>65,568,574,590,601,624,630,638,641,644,670,671,677,679,702,707,714,720,736,743,765,776,8</code> <code>05,826,847,856,874,888,903,917,925,930,935,940,945,947,969,982,992,1045,1050,1055,1068,</code> <code>1075,1076,1086,1089,1090,1108,1110,1125,1144,1177,1184,1185,1186,1187,1202,1213,1226,</code> <code>1231,1237,1246,1251,1253,1261,1276,1279)]</code> <code>SNPsSIG&lt;- 1:ncol(db.ok3)</code>

	<pre>names(SNPsSIG) &lt;- colnames(db.ok3) SIGNIF &lt;- colnames(SIG_db) COLS &lt;- SNPsSIG[SIGNIF] MATRIU &lt;- NULL col_db.ok3 &lt;- 1:ncol(db.ok3) names(col_db.ok3) &lt;- colnames(db.ok3)</pre>
<b>BUCLE ANÀLISI STEP</b>	<pre>for(k in COLS){</pre>
<b>1 I 2 MB-MDR</b>	<pre>  matriuModels &lt;- cbind(k, (1:ncol(db.ok3))[-k])   MATRIU &lt;- rbind(MATRIU, matriuModels)   }   resMbmdr &lt;- mbmdr(y=db.ok\$GA_frst, data= db.ok3, order=2,   family=binomial(link=logit), adjust="main effects", list.models=MATRIU)</pre>
<b>SELECCIÓ</b>	<pre>pvalor &lt;- 0.0001</pre>
<b>INTERACCIONS AMB</b>	<pre>AUX &lt;- resMbmdr\$result[resMbmdr\$result\$MIN.P &lt; pvalor,]</pre>
<b><math>\alpha &lt; 10^{-4}</math></b>	<pre>interaccions1 &lt;- AUX[order(AUX\$MIN.P), 1] interaccions2 &lt;- AUX[order(AUX\$MIN.P), 2] POS1_int1 &lt;- col_db.ok3[as.character(interaccions1)] POS2_int2 &lt;- col_db.ok3[as.character(interaccions2)] POS_int &lt;- cbind(POS1_int1, POS2_int2) rownames(POS_int) = NULL</pre>
<b>STEP 3 MB-MDR</b>	<pre>PERM &lt;- mbmdr.PermTest(resMbmdr, 10000, POS_int)</pre>

**12.3. MÈTODE 1 - RESULTATS STEP 1 I 2 MB-MDR**

**12.3.1. MÈTODE 1 - RESULTATS STEP 1 MB-MDR**

A les taules mostrades a continuació (**ANNEXES Taula 3.1-3.5**) es presenten els resultats de l’STEP 1 de l’anàlisi per MB-MDR pel **Mètode 1**.

Els genotips es mostren en les categories **0** (homozigots *wildtype*), **1** (heterozigots) i **2** (homozigots variants). Per a cadascuna de les combinacions de genotips, es presenten el número de **casos** i **controls** i el seu corresponent paràmetre **beta** i **p-valor**. L’exponent del paràmetre *beta* resultant del test estadístic d’associació entre les diferents combinacions de genotips dels SNPs correspon a l’**Odds-Ratio**: **OR=e<sup>beta</sup>**.

Finalment, a **categoria** es presenta la classificació en les diferents categories de risc **H** (*High risk*), **L** (*Low risk*), **0** (*categoria no informativa – no hi ha evidència d’associació*).

**ANNEXES Taula 3.1:** Step 1 de l’Anàlisi MB-MDR de la interacció 1 (**Taula 4**) analitzades pel **Mètode 1**.

<b>rs2243828</b>	<b>rs2274240</b>	<b>casos</b>	<b>controls</b>	<b>beta</b>	<b>p-valor</b>	<b>categoria</b>
<b>0</b>	<b>0</b>	<b>96</b>	<b>263</b>	<b>0.2967</b>	<b>0.0316490</b>	<b>H</b>
<b>1</b>	<b>0</b>	<b>18</b>	<b>170</b>	<b>-0.7291</b>	<b>0.0043272</b>	<b>L</b>
<b>2</b>	<b>0</b>	<b>4</b>	<b>22</b>	<b>-0.1950</b>	<b>0.7214927</b>	<b>0</b>
<b>0</b>	<b>1</b>	<b>98</b>	<b>404</b>	<b>-0.3162</b>	<b>0.0175938</b>	<b>L</b>
<b>1</b>	<b>1</b>	<b>65</b>	<b>193</b>	<b>0.5298</b>	<b>0.0007932</b>	<b>H</b>
<b>2</b>	<b>1</b>	<b>5</b>	<b>26</b>	<b>-0.1760</b>	<b>0.7206064</b>	<b>0</b>
<b>0</b>	<b>2</b>	<b>50</b>	<b>100</b>	<b>0.1221</b>	<b>0.5086367</b>	<b>0</b>
<b>1</b>	<b>2</b>	<b>19</b>	<b>80</b>	<b>-0.3166</b>	<b>0.2279841</b>	<b>0</b>
<b>2</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>0.6798</b>	<b>0.2943893</b>	<b>0</b>

**ANNEXES Taula 3.2:** Step 1 de l'Anàlisi MB-MDR de la interacció 2 (Taula 4) analitzades pel Mètode 1.

<i>rs2107545</i>	<i>rs2274240</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	93	241	0.31422	0.025071	H
1	0	23	180	-0.64020	0.005403	L
2	0	4	26	-0.16627	0.758359	0
0	1	91	376	-0.31784	0.019632	L
1	1	70	206	0.47290	0.002067	H
2	1	6	37	-0.11761	0.791217	0
0	2	49	100	0.08759	0.637240	0
1	2	21	76	-0.22835	0.368965	0
2	2	4	9	0.49736	0.410210	0

**ANNEXES Taula 3.3:** Step 1 de l'Anàlisi MB-MDR de la interacció 3 (Taula 4) analitzades pel Mètode 1.

<i>rs2030404</i>	<i>rs6602902</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	51	132	0.4109	0.022309	H
1	0	44	209	-0.2618	0.147479	0
2	0	24	76	-0.1934	0.427924	0
0	1	38	160	-0.2402	0.215235	0
1	1	82	243	0.2459	0.098825	H
2	1	32	93	-0.1815	0.401641	0
0	2	4	54	-0.9472	0.069809	L
1	2	9	65	-0.2704	0.454969	0
2	2	15	22	10.018	0.003355	H

**ANNEXES Taula 3.4:** Step 1 de l'Anàlisi MB-MDR de la interacció 4 (Taula 4) analitzades pel Mètode 1.

<i>rs1229982</i>	<i>rs3790843</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	142	393	0.23684	0.059723	H
1	0	36	206	-0.37953	0.048187	L
2	0	4	28	-0.76817	0.153396	0
0	1	73	338	-0.31283	0.033011	L
1	1	50	150	0.50982	0.003805	H
2	1	7	24	0.09153	0.832966	0
0	2	19	53	0.03854	0.888434	0
1	2	5	35	-0.64805	0.178849	0
2	2	5	3	161.048	0.027990	H

**ANNEXES Taula 3.5:** Step 1 de l'Anàlisi MB-MDR de la interacció 5 (Taula 4) analitzades pel Mètode 1.

<i>rs8056488</i>	<i>rs315947</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	71	345	-0.2392	0.103072	0
1	0	101	275	0.2474	0.068262	H
2	0	7	66	-0.2161	0.590952	0
0	1	87	218	0.3255	0.023669	H
1	1	60	224	-0.4299	0.006998	L
2	1	11	45	0.4164	0.223105	0
0	2	8	51	-0.4435	0.249418	0
1	2	18	38	0.4850	0.097405	H
2	2	0	15	-135.127	0.971315	0

**12.3.2. MÈTODE 1 - RESULTATS STEP 2 MB-MDR**
**ANNEXES Taula 4: Resultat de l'Step 2 de l'anàlisi MB-MDR de les interaccions amb  $p$ -valor inferior a  $10^{-3}$ .**

<i>SNP1</i>	<i>SNP2</i>	<i>N<sub>H</sub></i>	<i>beta<sub>H</sub></i>	<i>W<sub>H</sub></i>	<i>p-valor<sub>H</sub></i>	<i>N<sub>L</sub></i>	<i>beta<sub>L</sub></i>	<i>W<sub>L</sub></i>	<i>p-valor<sub>L</sub></i>	<i>min. p-valor</i>
rs2243828	rs2274240	2	0.5183	18.172	2.019e-05	2	-0.5032	15.795	7.058e-05	2.019e-05
rs2107545	rs2274240	2	0.5099	17.615	2.704e-05	2	-0.5030	15.664	7.565e-05	2.704e-05
rs2030404	rs6602902	3	0.5296	16.033	6.225e-05	1	-0.9472	3.288	6.981e-02	6.225e-05
rs1229982	rs3790843	3	0.4936	15.858	6.827e-05	2	-0.4299	10.991	9.154e-04	6.827e-05
rs8056488	rs315947	3	0.4788	15.751	7.224e-05	1	-0.4299	7.274	6.998e-03	7.224e-05
rs9297976	rs7322734	3	0.4852	15.379	8.794e-05	1	-0.7784	5.772	1.628e-02	8.794e-05
rs8056488	rs315952	3	0.4704	15.065	1.039e-04	1	-0.4294	7.077	7.810e-03	1.039e-04
rs603884	rs315947	2	0.4850	14.765	1.218e-04	2	-0.4736	12.674	3.707e-04	1.218e-04
rs2294008	rs676800	2	0.4648	10.684	1.081e-03	2	-0.6260	14.676	1.277e-04	1.277e-04
rs1229982	rs6455468	2	1.6729	13.884	1.944e-04	1	-1.0900	4.156	4.148e-02	1.944e-04
rs603884	rs315952	2	0.4724	13.868	1.961e-04	2	-0.4613	11.843	5.789e-04	1.961e-04
rs12670097	rs10794281	3	0.4492	13.749	2.090e-04	1	-0.5217	5.302	2.130e-02	2.090e-04
rs8050048	rs2297518	2	0.4709	13.674	2.174e-04	2	-0.4832	13.600	2.262e-04	2.174e-04
rs13253389	rs11568818	3	0.4835	13.449	2.452e-04	2	-0.5121	11.149	8.410e-04	2.452e-04
rs2030404	rs6602901	3	0.4575	13.414	2.498e-04	0	NA	NA	NA	2.498e-04
rs1800896	rs12219606	1	0.7761	6.877	8.734e-03	2	-0.9248	13.061	3.015e-04	3.015e-04
rs12219606	rs1800896	1	0.7761	6.877	8.734e-03	2	-0.9248	13.061	3.015e-04	3.015e-04
rs10929303	rs1041981	2	0.4467	12.860	3.356e-04	2	-0.4580	12.987	3.137e-04	3.137e-04
rs12219606	rs4803455	3	0.4369	12.620	3.816e-04	0	NA	NA	NA	3.816e-04
rs283411	rs1543680	1	1.0566	12.457	4.164e-04	1	-0.8686	5.467	1.937e-02	4.164e-04
rs13409371	rs4525426	2	0.4603	12.431	4.222e-04	2	-0.4908	11.948	5.471e-04	4.222e-04
rs2057768	rs2297518	2	0.4373	12.301	4.528e-04	2	-0.4316	11.479	7.040e-04	4.528e-04
rs17545656	rs3783512	1	0.7381	8.347	3.864e-03	2	-0.5465	12.290	4.554e-04	4.554e-04
rs2030404	rs8065878	2	0.4263	12.263	4.621e-04	2	-0.4510	12.252	4.648e-04	4.621e-04
rs10188292	rs4525426	2	0.4573	12.215	4.742e-04	2	-0.5003	12.187	4.811e-04	4.742e-04
rs13227585	rs3135500	2	0.4192	12.118	4.994e-04	2	-0.4258	11.502	6.951e-04	4.994e-04
rs2530797	rs8178592	1	0.8426	12.080	5.096e-04	1	-0.5117	4.411	3.571e-02	5.096e-04
rs4251961	rs9436740	2	0.4586	12.043	5.197e-04	2	-0.4355	9.871	1.679e-03	5.197e-04
rs1125557	rs2107545	2	0.4276	11.998	5.324e-04	2	-0.4266	11.109	8.592e-04	5.324e-04
rs2107545	rs1125557	2	0.4276	11.998	5.324e-04	2	-0.4266	11.109	8.592e-04	5.324e-04
rs10902073	rs7916932	2	0.4663	11.829	5.831e-04	2	-0.4313	8.446	3.659e-03	5.831e-04
rs2854450	rs2174270	2	0.3747	9.846	1.702e-03	2	-0.4421	11.825	5.844e-04	5.844e-04
rs7012891	rs12219246	1	1.7987	11.750	6.083e-04	1	-0.7533	4.463	3.464e-02	6.083e-04
rs283411	rs707889	1	1.0339	11.729	6.153e-04	1	-0.7737	4.823	2.808e-02	6.153e-04
rs9929218	rs11121832	2	0.3982	10.885	9.693e-04	2	-0.4255	11.596	6.610e-04	6.610e-04
rs936469	rs3775260	1	2.2053	11.571	6.700e-04	0	NA	NA	NA	6.700e-04
rs502396	rs12612413	2	0.4000	9.831	1.716e-03	2	-0.5046	11.395	7.364e-04	7.364e-04
rs10853773	rs7226	1	3.0875	11.383	7.412e-04	0	NA	NA	NA	7.412e-04
rs502396	rs2860840	2	0.4971	11.352	7.535e-04	0	NA	NA	NA	7.535e-04
rs12818198	rs9543325	2	0.3857	9.644	1.900e-03	2	-0.4412	11.320	7.668e-04	7.668e-04



<b>rs9543325</b>	rs12818198	2	0.3857	9.644	1.900e-03	2	-0.4412	11.320	7.668e-04	7.668e-04
<b>rs17042407</b>	rs3088277	2	0.4051	11.258	7.930e-04	2	-0.4039	10.463	1.218e-03	7.930e-04
<b>rs13409371</b>	rs8113762	2	0.5817	11.243	7.992e-04	1	-0.6433	4.305	3.801e-02	7.992e-04
<b>rs17186418</b>	rs1800975	3	0.4377	11.222	8.085e-04	0	NA	NA	NA	8.085e-04
<b>rs1024610</b>	rs1078979	2	0.4027	11.177	8.283e-04	2	-0.4010	10.644	1.104e-03	8.283e-04
<b>rs2294008</b>	rs11877833	2	0.4060	8.834	2.957e-03	2	-0.5023	11.030	8.964e-04	8.964e-04
<b>rs11016908</b>	rs7334764	1	2.4231	10.935	9.434e-04	1	-0.8422	3.173	7.486e-02	9.434e-04

#### 12.4. PAQUET ESTADÍSTIC SINÈRGIA

Per a l'anàlisi de la *Sinèrgia* s'ha utilitzat la funció **Synergy** escrita per Víctor Urrea. En aquesta funció s'especifiquen els següents paràmetres:

***Synergy (ordre,dades,Ycol=1,keep=100,batch=900)***

L'ordre o dimensió de les interaccions (2 o 3), el *data frame* amb les dades sobre les que s'ha d'aplicar la funció (amb els SNPs transformats mitjançant *I'SNPassoc*).

*Ycol* equival al número de la columna on està la variable resposta; per defecte considera que la variable resposta està a la primera columna.

El paràmetre *keep* correspon al nombre d'interaccions a guardar; per defecte guarda les 100 millors interaccions amb un valor més alt de *Sinèrgia*.

Per últim, el paràmetre *batch* és un paràmetre computacional que avalua per blocs de la mida especificada les interaccions possibles. Requereix els paquets estadístics *utils* i *base*.

## 12.5. SCRIPTS APLICATS PER A L'ANÀLISI DEL MÈTODE 2 AMB R

ANNEXES Taula 5: *Scripts i funcions utilitzades en l'anàlisi MB-MDR del Mètode 2.*

<b>CARREGAR PAQUETS</b>	<code>library(utils)</code> <code>library(base)</code>
<b>CARREGAR FUNCIO</b>	<code>load("Sinergia_rc5.R")</code>
<b>SINÈRGIA</b>	
<b>EXECUCIÓ DE LA</b>	<code>SIN&lt;-Synergy(2,db.ok,Ycol=1,keep=500,batch=900)</code>
<b>FUNCIO SOBRE LA</b>	<code>MATRIU&lt;-SIN[,c(1,2)]</code>
<b>BASE DE DADES</b>	<code>MAT_MODEL&lt;-MATRIU-1</code>
<b>BUCLE ANÀLISI STEP</b>	<code>resMbmdr&lt;-mbmdr(y=db,ok\$GA_frst, data= db,ok3, order=2, exclude=NA, risk,threshold=0,1,</code>
<b>1 I 2 MB-MDR</b>	<code>family=binomial(link=logit), adjust=c ("main effects"), list,models=MAT_MODEL)</code>
<b>SELECCIÓ</b>	<code>pvalor &lt;- 0.0001</code>
<b>INTERACCIONS AMB</b>	<code>AUX &lt;- resMbmdr\$result[resMbmdr\$result\$MIN.P&lt;pvalor,]</code>
<b><math>\alpha &lt; 10^{-4}</math></b>	<code>interaccions1 &lt;- AUX[order(AUX\$MIN.P),1]</code> <code>interaccions2 &lt;- AUX[order(AUX\$MIN.P),2]</code> <code>POS1_int1 &lt;- col_db.ok3[as.character(interaccions1)]</code> <code>POS2_int2 &lt;- col_db.ok3[as.character(interaccions2)]</code> <code>POS_int &lt;- cbind(POS1_int1,POS2_int2)</code> <code>rownames(POS_int)=NULL</code>
<b>STEP 3 MB-MDR</b>	<code>PERM&lt;-mbmdr.PermTest(resMbmdr,10000,POS_int)</code>

**12.6. MÈTODE 2 - RESULTATS STEP 1 I 2 MB-MDR****12.6.1. MÈTODE 2 - RESULTATS STEP 1 MB-MDR**

A les taules mostrades a continuació (**ANNEXES Taula 6.1-6.4**) es presenten els resultats de l'Step 1 de l'anàlisi per MB-MDR pel **Mètode 2** (veure Annexes, apartat 12.4.1 per la codificació de les variables exposades a les taules).

**ANNEXES Taula 6.1:** Step 1 de l'Anàlisi MB-MDR de la interacció 1 (**Taula 7**) analitzades pel **Mètode 2**.

<i>rs8050048</i>	<i>rs315947</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	81	380	-0.2184	0.12341	0
1	0	81	222	0.2682	0.06905	H
2	0	5	38	-0.1396	0.77100	0
0	1	93	241	0.2643	0.06183	H
1	1	49	186	-0.4158	0.01676	L
2	1	8	29	0.3775	0.35027	0
0	2	11	52	-0.1693	0.61617	0
1	2	13	31	0.3662	0.27624	0
2	2	0	11	-136.761	0.97514	0

**ANNEXES Taula 6.2:** Step 1 de l'Anàlisi MB-MDR de la interacció 2 (**Taula 7**) analitzades pel **Mètode 2**.

<i>rs4239495</i>	<i>rs1443623</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	77	318	-0.22991	0.11049	0
1	0	85	247	0.24732	0.08379	H
2	0	19	60	-0.01272	0.96246	0
0	1	87	235	0.26647	0.06215	H
1	1	46	220	-0.40603	0.01998	L
2	1	17	43	0.17746	0.54468	0
0	2	11	51	-0.12151	0.71920	0
1	2	16	52	0.26718	0.36080	0
2	2	2	16	-0.77390	0.30367	0

ANNEXES Taula 6.3: Step 1 de l'Anàlisi MB-MDR de la interacció 3 (Taula 7) analitzades pel Mètode 2.

<i>rs2555111</i>	<i>rs10277279</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	50	191	0.08460	0.62148	0
1	0	113	312	0.07257	0.57877	0
2	0	39	194	-0.23167	0.21437	0
0	1	22	135	-0.41723	0.07968	L
1	1	82	231	0.05998	0.67923	0
2	1	34	112	0.24628	0.23083	0
0	2	12	24	0.84111	0.01903	H
1	2	6	49	-0.95766	0.02830	L
2	2	7	23	0.31814	0.46541	0

ANNEXES Taula 6.4: Step 1 de l'Anàlisi MB-MDR de la interacció 4 (Taula 7) analitzades pel Mètode 2.

<i>rs2758320</i>	<i>rs2470144</i>	<i>casos</i>	<i>controls</i>	<i>beta</i>	<i>p-valor</i>	<i>categoria</i>
0	0	74	240	0.09868	0.507315	0
1	0	31	102	-0.12403	0.562886	0
2	0	3	17	-0.53567	0.394497	0
0	1	114	414	0.11972	0.350115	0
1	1	47	194	-0.26123	0.135070	0
2	1	9	25	0.30617	0.436409	0
0	2	38	194	-0.37498	0.046592	L
1	2	40	75	0.54226	0.008373	H
2	2	4	15	-0.07092	0.900270	0

**12.6.2. MÈTODE 2 - RESULTATS STEP 2 MB-MDR**
**ANNEXES Taula 7: Resultat de l'Step 2 de l'anàlisi per MB-MDR del Mètode 2 – Interaccions amb  $p$ -valor<0.05.**

<b>SNP1</b>	<b>SNP2</b>	<b><math>N_H</math></b>	<b><math>\beta_{H_i}</math></b>	<b><math>W_H</math></b>	<b><math>p</math>-valor<math>_H</math></b>	<b><math>N_L</math></b>	<b><math>\beta_{L_i}</math></b>	<b><math>W_L</math></b>	<b><math>p</math>-valor<math>_L</math></b>	<b>min. <math>p</math>-valor</b>
rs8050048	rs315947	2	0,3789	9.363	0,002214	1	-0,4158	5.722	0,01676	0,002214
rs4239495	rs1443623	2	0,3554	8.691	0,003198	1	-0,406	5.414	0,019977	0,003198
rs2555111	rs10277279	1	0,8411	5.499	0,019027	2	-0,5873	7.645	0,005693	0,005693
rs2758320	rs2470144	1	0,5423	6.952	0,008373	1	-0,375	3.960	0,04659	0,008373
rs9436737	rs10829618	1	20.544	6.293	0,012125	0	NA	NA	NA	0,012125
rs1614972	rs1516629	1	0,7611	6.165	0,013033	0	NA	NA	NA	0,013033
rs762551	rs480643	1	0,9404	5,329	0,02098	1	-1,2124	3,972	0,04627	0,02098
rs1800935	rs1614972	1	12.853	5.276	0,021617	0	NA	NA	NA	0,021617
rs938890	rs1650697	1	13.436	5.210	0,022457	0	NA	NA	NA	0,022457
rs6608827	rs17412368	1	0,5589	5.096	0,023979	0	NA	NA	NA	0,023979
rs9462852	rs4939827	1	0,7396	4.957	0,025988	1	-0,5785	2.799	0,0943	0,025988
rs4699741	rs3775260	1	0,6919	4.399	0,035964	0	NA	NA	NA	0,035964
rs6509400	rs13062355	1	17.166	4.397	0,036012	0	NA	NA	NA	0,036012
rs13031342	rs1059840	1	0,7813	4.187	0,040742	0	NA	NA	NA	0,040742
rs2666433	rs2306371	0	NA	NA	NA	1	-1	4.137	0,04196	0,04196
rs2733534	rs2281610	1	0,8564	4.136	0,041989	0	NA	NA	NA	0,041989
rs778986	rs2424932	1	0,8228	4.065	0,04379	1	-12.859	3.012	0,08264	0,04379
rs2227869	rs1059519	1	1	4.003	0,04542	0	NA	NA	NA	0,04542
rs6602902	rs10493380	1	0,603	4.002	0,045444	0	NA	NA	NA	0,045444

**ANNEXES Taula 7: Resultat de l'anàlisi per MB-MDR – Interaccions amb  $p$ -valor<0.05. En blau es destaquen les interaccions que han estat classificades a la categoria de baix risc.**