

# VALIDACIÓ D'UN ESTUDI D'EXÒMICA EN UNA FAMÍLIA AFECTADA D'ESCLEROSI MÚLTIPLE

---

Estudiant: Alex Fernández Baena

Grau en Biotecnologia

Correu electrònic: alex\_feba@hotmail.com

Tutora: Rosa M. Araguas Sola

Cotutora: Ester Quintana Camps

Institució: Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta (IDIBGI)

Vistiplau tutor i cotutor:

Rosa M. Araguas Sola  
(Tutora)

Ester Quintana Camps  
(Cotutora)

Institució: IDIBGI  
Correus electrònics:

rosa.araguas@udg.edu  
equintana@idibgi.org

Data de dipòsit de la memòria a secretaria de coordinació: 29/05/2015



# ÍNDIX GENERAL

---

<i>Resum</i> .....	2
<i>Resumen</i> .....	3
<i>Abstract</i> .....	4
1. Introducció .....	5
1.1 Esclerosi múltiple .....	5
1.1.1 Etiologia i factors de risc .....	5
1.1.2 Tipus d'EM .....	6
1.2 Estudis en genètica d'esclerosi múltiple .....	6
1.2.1 <i>Human Leukocyte Antigen</i> (HLA) .....	7
1.2.2 <i>Genome Wide Association Studies</i> (GWAS).....	7
1.2.3 Exòmica.....	8
1.3 Introducció al projecte general .....	9
2. Objectives.....	10
3. Materials i mètodes .....	11
3.1 Descripció de la família.....	11
3.2 Obtenció del DNA dels subjectes .....	11
3.3 Selecció de gens de l'exòmica.....	12
3.4 Identificació del polimorfisme en les bases de dades .....	13
3.5 Amplificació dels fragments per PCR .....	13
3.5.1 Disseny de <i>primers</i> .....	13
3.5.2 Posada a punt de la PCR.....	13
3.5.3 PCR i comprovació en gel d'agarosa.....	14
3.6 Genotipatge .....	14
3.6.1 Genotipatge de HLA-DQB1 i HLA-DRB1.....	14
3.6.2 Seqüenciació pel mètode Sanger .....	16
3.6.3 Genotipatge per <i>assay</i> .....	16
3.7 Estadística .....	17
4. Resultats i discussió .....	18
4.1 Resultats exòmica i selecció de gens candidats .....	18
4.2 Validació de l'exòmica.....	21
4.2.1 Disseny de <i>primers</i> .....	21
4.2.2 Posada a punt de la PCR.....	22
4.2.3 Comprovació de l'amplificació dels fragments .....	23
4.2.4 Seqüenciació pel mètode Sanger .....	24
4.2.5 <i>Assay</i> .....	25
4.2.6 Resultats dels genotips dels gens analitzats .....	27
4.2.7 Concordança entre els genotips de l'exòmica i la seqüenciació.....	28
4.3 HLA .....	29
4.4 Estadística dels gens analitzats .....	30
5. Conclusions .....	32
6. Bibliografia .....	33

# RESUM

---

L'esclerosi múltiple (EM) és la primera causa de discapacitat neurològica no traumàtica en adults joves. Sabem que la genètica juga un paper fonamental en el seu desenvolupament. S'han descrit molts gens de susceptibilitat d'EM, però tots junts permeten explicar poc sobre el desenvolupament de la malaltia.

El present projecte es basa en l'estudi genètic d'una família d'ètnia gitana amb un nombre elevat d'individus afectats d'EM i amb un alt nivell d'endogàmia. Per tant, l'estudi genètic d'aquesta família presenta unes condicions excel·lents per la detecció de gens de susceptibilitat a l'EM. El treball ha consistit en la validació de diferents canvis genètics identificats en un estudi d'exòmica, realitzat en un número reduït d'individus respecte el total d'individus de la família que participen en l'estudi (4 individus afectats d'EM i 1 individu sa com a control pels al·lels de referència). Partint dels resultats de l'estudi d'exòmica, s'han seleccionat, per a la seva validació en tots els individus de la família, aquells gens que, per la funció de la proteïna resultant i la conseqüència funcional del canvi genètic, poden estar involucrats en el desenvolupament de l'EM. Per fer el disseny del genotipatge s'ha buscat el gen, s'han localitzat els canvis, s'ha seleccionat la regió d'interès i s'han establert les condicions òptimes per a la seva amplificació. La validació dels canvis seleccionats s'ha realitzat pel mètode de seqüenciació per Sanger o per RT-PCR amb sondes TaqMan<sup>®</sup>. No s'han observat diferències significatives entre el genotip dels individus sans i afectats per l'EM en cap dels gens analitzats. Degut a la baixa cobertura en algunes regions analitzades en l'estudi d'exòmica, s'ha observat una baixa concordança entre el genotipatge proporcionat per l'exòmica i l'obtingut en la seva validació.

El HLA és el factor genètic que més s'ha associat a la susceptibilitat a EM, per aquest motiu s'han analitzat els gens DRB1 i DQB1 mitjançant el sistema Luminex<sup>®</sup>, no trobant en cap individu els genotips DRB1\*15:01 i DQB1\*06:02 considerats de risc a EM. Per altra banda, el genotip DRB1\*03 es troba molt present en aquesta família.

Els gens estudiats en aquest projecte no permeten explicar la presència d'EM en aquesta família.

# RESUMEN

---

La esclerosis múltiple (EM) es la primera causa de discapacidad neurológica no traumática en adultos jóvenes. Sabemos que la genética juega un papel fundamental en su desarrollo. Se han descrito muchos genes de susceptibilidad a la EM, pero todos juntos permiten explicar poco sobre el desarrollo de la enfermedad.

El presente proyecto se basa en el estudio genético de una familia de etnia gitana con un número elevado de individuos afectados de EM y con un alto nivel de endogamia. Por tanto, el estudio genético de esta familia presenta unas condiciones excelentes para la detección de genes de susceptibilidad a la EM. El trabajo ha consistido en la validación de diferentes cambios genéticos identificados en un estudio de exómica, realizado en un número reducido de individuos respecto del total de individuos de la familia que participan en el estudio (4 individuos afectados de EM y 1 individuo sano como control por los alelos de referencia). Partiendo de los resultados del estudio de exómica, se han seleccionado, para su validación en todos los individuos de la familia, aquellos genes que, por la función de la proteína resultante y la consecuencia funcional del cambio genético, pueden estar involucrados en el desarrollo de la EM. Para hacer el diseño del genotipado se ha buscado el gen, se han localizado los cambios, se ha seleccionado la región de interés y se han establecido las condiciones óptimas para su amplificación. La validación de los cambios seleccionados se ha realizado por el método de secuenciación por Sanger o por RT-PCR con sondas TaqMan<sup>®</sup>. No se han observado diferencias significativas de un genotipo entre los individuos sanos y afectados por la EM en ninguno de los genes analizados. Debido a la baja cobertura en algunas regiones analizadas en el estudio de exómica, se ha observado una baja concordancia entre el genotipado proporcionado por la exómica y el obtenido en su validación.

El HLA es el factor genético que más se ha asociado a la susceptibilidad a EM, por este motivo se han analizado los genes DRB1 y DQB1 mediante el sistema Luminex<sup>®</sup>, no encontrando en ningún individuo los genotipos DRB1\*15:01 y DQB1\*06:02 considerados de riesgo en EM. Por otra parte, el genotipo DRB1\*03 se encuentra muy presente en esta familia.

Los genes estudiados en este proyecto no permiten explicar la presencia de EM en esta familia.

# ABSTRACT

---

Multiple sclerosis (MS) is the leading reason of nontraumatic neurological disability in young adults. We know that genetics plays a role in its development. It has described many susceptibility genes for MS, but together they are not able to explain the development of the disease. This project is based on the genetic study of a gypsy family with a large number of individuals affected by MS and a high level of inbreeding. Therefore, the genetic study of this family has excellent conditions to detect MS susceptibility genes.

This project aims to validate genetic changes identified in an exome study realized in a small number of individuals of the total family (4 individuals affected by MS and 1 healthy individual used as control for the reference alleles). On the basis of the results of the exome study, it has been selected for validation in all individuals of the family, those genes in which the function of the resulting protein and the functional consequence of genetic change may be involved in the development of MS. Previously to the genetic study, genes sequences has been searched, the changes localized and identified the regions of interest. Then, the optimal conditions for amplification have been established. Validation of selected changes has been made by Sanger sequencing or RT-PCR with TaqMan<sup>®</sup> probes. No significant differences were observed for the genotype between healthy and affected individuals in any analyzed gene. Due to the low coverage in some regions analyzed in the exome study, it has observed low concordance between genotyping provided by exome study and the one obtained in the validation.

Because of the importance of HLA in MS susceptibility, *DRB1* and *DQB1* genes have been analyzed by Luminex<sup>®</sup> system. None of the studied individuals presents the genotypes DRB1\*15:01 and DQB1\*06:02 considered at risk of MS. However, the DRB1\*03 genotype is common in this family.

The genes studied in this project are not able to explain the presence of MS in the family.

# 1. INTRODUCCIÓ

---

## 1.1 ESCLEROSI MÚLTIPLE

L'esclerosi múltiple (EM) és una malaltia neurològica crònica del sistema nerviós central (SNC), d'etiologia no coneguda i, aparentment, de base autoimmunità. Els models actuals de patogenicitat parlen de dos mecanismes que es superposen, la inflamació i la neurodegeneració, ja que l'EM es caracteritza per la presència d'infiltrats inflamatoris, desmielització, lesió axonal i pèrdua neuronal en el SNC. És una malaltia de curs progressiu i causa una discapacitat física i cognitiva (Ebers et al., 2008; Oksenberg et al., 2008; Oksenberg & Baranzini, 2010).

La malaltia és una de les causes més comunes de discapacitat neurològica en adults joves i s'estima que més de 2 milions de persones tenen EM a tot el món (Kingwell et al., 2013). L'EM és més freqüent en dones que en homes (2-3:1), i aquesta incidència està augmentant a mesura que passa el temps (Greer & McCombe, 2011; Harbo et al., 2013).

L'eina més important pel diagnòstic de l'EM és la ressonància magnètica (RM) i es basa en demostrar la disseminació espacial (aparició d'almenys dues lesions de localització diferent en el SNC) i temporal (aparició d'almenys dues lesions en el SNC separades en el temps). En el diagnòstic, també són importants els símptomes clínics i la detecció de bandes oligoclonals d'immunoglobulina G (IgG) en líquid cefaloraquídi (LCR). Malauradament, en l'actualitat no existeixen mesures terapèutiques curatives, però el diagnòstic precoç i els tractaments disponibles modifiquen parcialment el curs de la malaltia i milloren el pronòstic.

### 1.1.1 ETIOLOGIA I FACTORS DE RISC

Actualment l'etiologia específica de l'EM no és coneguda, però se sap que no hi ha una única causa de la malaltia, sinó que l'EM és el resultat d'una complexa interacció entre els diferents factors genètics, ambientals i epigenètics (American Neurological Association, 2011; Giovannoni & Ebers, 2007; Zuvich et al., 2009).

L'EM és una malaltia de genètica complexa, ja que el component genètic no s'hereta segons un patró mendelià simple (Oksenberg et al., 2008) i possiblement aquest fet està associat amb fenòmens epistàtics (Oksenberg et al., 2008; Zuvich et al., 2009). Els estudis epidemiològics, la diferent susceptibilitat entre els grups racials i ètnics, i els casos d'agregació familiar, indiquen la clara implicació del component genètic en la susceptibilitat a l'EM. A més, les variants genètiques també poden determinar el moment de l'aparició dels símptomes, el curs de la malaltia, i la resposta al tractament (Ebers & Sadovnick, 1994; Oksenberg & Baranzini, 2010; Sadovnick et al., 1996). La susceptibilitat genètica global de la malaltia ve determinada per interaccions complexes de variants al·lèliques de molts gens, cadascun dels quals contribueix al risc global de la malaltia (Zuvich et al.,

2009). Malgrat la importància del factor genètic, és estrany trobar famílies amb 5 o més individus amb EM en successives generacions (Ebers et al., 2008; Dyment et al., 2012).

D'altra banda, hi ha una forta evidència que els factors ambientals influeixen en la heterogeneïtat i la distribució geogràfica de l'EM. Entre els factors ambientals més implicats en la patogènia de l'EM destaquen els baixos nivells de vitamina D, el tabac i la infecció del virus Epstein-Barr (VEB) (Giovannoni & Ebers, 2007). Aquests i d'altres possibles factors ambientals interactuen sobre els factors de predisposició genètica mitjançant canvis o modificacions epigenètiques que finalment seran els responsables del desenvolupament de l'EM (G. Ebers, 2008; Giovannoni & Ebers, 2007).

### 1.1.2 TIPUS D'EM

El curs clínic de la malaltia i el seu pronòstic evolutiu és heterogeni i difícil de predir. Típicament, l'EM comença com un trastorn episòdic i es desenvolupa amb el temps en una malaltia progressiva. La primera fase de la malaltia està descrita per l'aparició de dos tipus d'episodis primaris: RIS i CIS. Un RIS (*Radiologically Isolated Syndrome*) o Síndrome Radiològicament Aïllat (SRA) es defineix com una presentació sense símptomes clínics manifestos però amb resultats de la RM altament suggestius d'EM. D'altra banda un CIS (*Clinically Isolated Syndrome*) o SCA (Síndrome Clínicament Aïllat) s'utilitza per descriure un primer episodi de símptomes neurològics durant almenys 24 hores i és causada per la inflamació i desmielinització en un o més llocs en el SNC. Un percentatge important d'aquests dos tipus de pacients acabaran desenvolupant EM (Barrie, 2009). Per tant, com que molts RIS acaben desenvolupant EM, en aquest projecte era essencial la realització d'una RM a tots els participants de l'estudi, fossin o no afectats per EM.

Així doncs, si un individu pateix un segon brot o augmenta progressivament la discapacitat, ja es considera una EM clínicament definida (EMCD). L'EMCD s'ha classificat tradicionalment des d'un punt de vista fenotípic clínic en:

- EM remitent-recurrent (EMRR): els pacients presenten nous brots de la malaltia sense progressió entre ells, que poden remetre totalment o deixar seqüeles.
- EM secundària progressiva (EMSP): després d'una fase de brots els pacients empitjoren lentament amb un increment progressiu de la discapacitat sense brots.
- EM primària progressiva (EMPP): els pacients presenten des de l'inici una forma progressiva de la malaltia sense brots.

## 1.2 ESTUDIS EN GENÈTICA D'ESCLEROSI MÚLTIPLE

En els últims anys s'han realitzat nombroses aproximacions per identificar les regions cromosòmiques involucrades en la susceptibilitat a desenvolupar EM. A continuació s'esmenten els principals estudis que s'han fet en genètica d'EM i que han sigut de gran utilitat pel desenvolupament del projecte d'estudi.



### 1.2.1 HUMAN LEUKOCYTE ANTIGEN (HLA)

En l'actualitat, l'única regió del genoma que s'ha associat clara i consistentment amb EM és el Complex Major d'Histocompatibilitat (*Major Histocompatibility Complex*, MHC) o també anomenat en humans com Antigen Leucocitari Humà (*Human Leukocyte Antigen*, HLA) (Oksenberg et al., 2008). Aquest locus, localitzat en el braç curt del cromosoma 6 (6p21), explica el 30% del component genètic de la malaltia (Haines et al., 1996). El HLA està classificat en tres classes, però són els gens que formen part del HLA de classe II els que s'han associat a EM. Aquests, s'expressen en la superfície de les cèl·lules presentadors d'antígens (cèl·lules dendrítiques, macròfags i limfòcits B) i codifiquen per proteïnes presentadores de pèptids exògens als limfòcits T Helper (Traherne, 2008). El conjunt de gens del complex presenten una baixa taxa de recombinació, ja que es troben lligats, i en conseqüència, certs al·lels de gens del HLA formen haplotips. Els gens d'aquest complex són altament polimòrfics (Haines et al., 1996; Traherne, 2008). Determinats haplotips del HLA s'han associat, no únicament amb una major susceptibilitat a presentar la malaltia, sinó que també semblen determinar el grau de severitat. Pel que fa a la codificació dels al·lels, el primer número del genotip del gen determina la combinació de molts SNPs (*Single Nucleotide Polymorphism*) i fa referència al grup al·lèlic. El segon número defineix els canvis d'aminoàcid en la proteïna en un grup al·lèlic determinat.

Concretament l'haplotip DR2 del HLA de classe II, el qual compren els al·lels DRB1\*15:01 i DQB1\*06:02, s'ha associat en EM en tots els grups ètnics, però particularment en els caucàsics (Fernández et al., 2004), mostrant la major associació el HLA-DRB1\*15:01 (Ramagopalan et al., 2007). L'any 2008, Fernández i col·laboradors van realitzar un estudi sobre l'associació de HLA de classe II i l'EM en població gitana. En aquest estudi es van observar diferències significatives pels al·lels DRB1\*15:01 i DQB1\*06:02 entre els controls i afectats d'EM d'ètnia gitana. En principi amb la presència dels al·lels ja es considera susceptibilitat d'EM, per tant, no es parla d'homozigosi en aquests al·lels.

### 1.2.2 GENOME WIDE ASSOCIATION STUDIES (GWAS)

Encara que no es conegui la totalitat dels gens involucrats en l'EM, s'han aconseguit grans avenços en els últims anys gràcies al desenvolupament dels estudis d'associació de genoma complet (*Genome Wide Association Studies*, GWAS). Els GWAS es basen en la tecnologia dels *microarrays* i permeten analitzar més de 100.000 SNPs de tot el genoma de manera simultània en molts individus controls i afectats per una malaltia. Així doncs, aquesta tecnologia pot determinar si un polimorfisme està més present en els individus afectats per una malaltia que en els controls (Hoffjan & Akkad, 2010). El seu principal inconvenient és l'elevada probabilitat d'obtenir falsos positius, per això, és imprescindible replicar l'estudi en un altre col·lecció de mostres. A més, generalment, només identifiquen variacions comunes (amb freqüències del al·lel menor superiors al 5%). Tots els GWAS

realitzats en EM han destacat, amb una forta significació estadística, el clàssic locus de risc HLA-DRB1\*15:01. L'any 2010, Oksenberg i Baranzini, mitjançant estudis de GWAS, van descriure els 12 gens que, en cert grau, s'associaven al desenvolupament d'EM: CD58, EVI5, RGS1, IL7R, IL2RA, CD6, TNFRSF1A, CLEC16a, IRF8, CD226, TYK2 i CD40. De cada gen, es va identificar un SNP que es trobava en més freqüència entre els pacients afectats d'EM que en els controls. Tots els nous al·lels de risc per l'EM que s'han identificat amb estudis de GWAS són bastant freqüents, però exerceixen un efecte individual modest en el risc i actuen de forma independent (Oksenberg & Baranzini, 2010). Actualment es coneixen almenys 150 gens associats a la susceptibilitat a l'EM.

### 1.2.3 EXÒMICA

L'exòmica és la seqüenciació de l'exoma complet (*Whole Exome Sequencing*, WES) que pertany a les tecnologies de Seqüenciació de Nova Generació (*Next Generation Sequencing*, NGS) o de Seqüenciació Massiva en Paral·lel (*Massively Parallel Sequencing*). Aquesta tècnica permet l'anàlisi de totes les regions genòmiques codificants (exons) de diferents mostres en un únic procés de seqüenciació i en un temps reduït. La WES és una alternativa de menor cost i més senzilla d'interpretar que la Seqüenciació del Genoma Complet (*Whole Genome Sequencing*, WGS) ja que la WES analitza l'exoma que compren aproximadament l'1,5% del genoma, el component que codifica per la producció de proteïnes (International Human Genome Sequencing Consortium, 2001). Aquesta plataforma ha provocat un gran impacte en la genètica mèdica, ja que permet determinar les variacions de totes les regions codificants o exons (generalment SNPs) entre pacients afectats d'una malaltia i controls sans (Rabbani et al., 2014). Malgrat l'elevada capacitat de donar resultats i la possibilitat d'analitzar tot el genoma, aquestes tècniques de seqüenciació massiva són relativament cares i poden cometre errors de lectura. Per tant, després d'un estudi d'exòmica cal la corresponent validació amb un altre mètode de seqüenciació més sensible, per poder confirmar els resultats.

Existeix un estudi d'exòmica en una família amb molts membres afectats d'EM (Dyment et al., 2012) en el qual, prèviament s'havia realitzat un estudi de lligament sense observar cap gen que s'associés amb la malaltia. D'altra banda l'estudi d'exòmica va detectar un SNP en el gen TYK2 que es trobava en un 72% dels afectats d'EM i en un 47% dels no afectats. Així doncs, van identificar una nova variant en la susceptibilitat de l'EM. També van observar que en aquest gen hi ha una prioritat de transferència de l'al·lel mutat respecte el *wild type*. Per altra banda, van extrapolar aquest gen de susceptibilitat a altres famílies sense trobar que la variant del TYK2 estigués més representada en els afectats que els sans. Per tant, quan s'estudia una família concreta, és possible que els resultats genotípics no es puguin extrapolar a tota la població. Cal destacar que aquest mateix gen, ja havia estat descrit l'any 2010 per Oksenberg i Baranzini associat a la susceptibilitat a EM.

### 1.3 INTRODUCCIÓ AL PROJECTE GENERAL

Per la realització del present treball d'investigació he col·laborat amb el grup de recerca d'EM de la Unitat de Neuroimmunologia i Esclerosi Múltiple (UNIEM) que pertany al Institut d'Investigació Biomèdica de Girona (IDIBGI). El projecte en el qual he participat es basa en l'estudi dels factors genètics d'una família gitana amb un número elevat de membres amb EM. Concretament he participat en la validació d'un estudi d'exòmica realitzat en un nombre reduït d'individus de la família per la identificació de gens de susceptibilitat a EM. L'exòmica es va desenvolupar de manera externa (Sistemas Genómicos S.L) i a part d'analitzar tot l'exoma, també s'han analitzat algunes zones intròniques flanquejants definides per la pròpia empresa. Així doncs, l'exòmica ha identificat un gran número de canvis en diferents gens presents en els membres afectats de la família i que no porta el membre sa estudiat. Entre aquests gens, alguns tenen funcions que podrien estar relacionats amb la patogènia de la malaltia. La validació d'aquests canvis en tots els membres de la família ens indicarà la implicació d'aquests en la malaltia.

Per tal d'arribar a la situació actual de l'estudi, és a dir, a la validació de l'exòmica mitjançant el genotipatge dels gens candidats seleccionats en tota la família, prèviament s'han realitzat una sèrie de passos que s'esmenten a continuació. En primer lloc es va plantejar la possibilitat de realitzar un estudi genètic a la família, pel fet de tenir molts membres afectats d'EM. A més, l'alta endogàmia present en la família fa pensar en un augment de la càrrega genètica, de manera que l'EM es comporti com una malaltia monogènica. Per tant, tot i ser l'EM una malaltia de genètica complexa, en la qual intervenen nombrosos loci, les circumstàncies de la família eren úniques per la identificació de gens de susceptibilitat en EM.

Un cop obtingudes les mostres de DNA dels subjectes d'estudi, es van analitzar els dos al·lels del HLA que més s'han relacionat amb la malaltia (HLA-DRB1\*15:01 i HLA-DQB1\*06:02). Tot i ser el primer pas de l'estudi, s'ha hagut de repetir en etapes més avançades del projecte, ja que s'han anat incorporant nous membres de la família. Posteriorment, es van analitzar els 12 gens que més s'han relacionat en EM després del HLA (Oksenberg & Baranzini, 2010). El que es va observar va ser que, en la família estudiada, no hi havien diferències significatives entre els afectats i no afectats respecte els genotips de risc del HLA i dels 12 gens de susceptibilitat. Aquest fet fa pensar que en la família d'estudi hi ha un factor genètic que podria explicar la causa de la malaltia i que possiblement encara no s'ha associat a l'EM. És per aquest motiu que es va procedir a la realització d'un estudi d'exòmica.

## 2. OBJECTIVES

---

The main objective of this project is to identify susceptibility genes for multiple sclerosis (MS) in a gypsy family, by validating a Whole Exome Sequencing (WES) study.

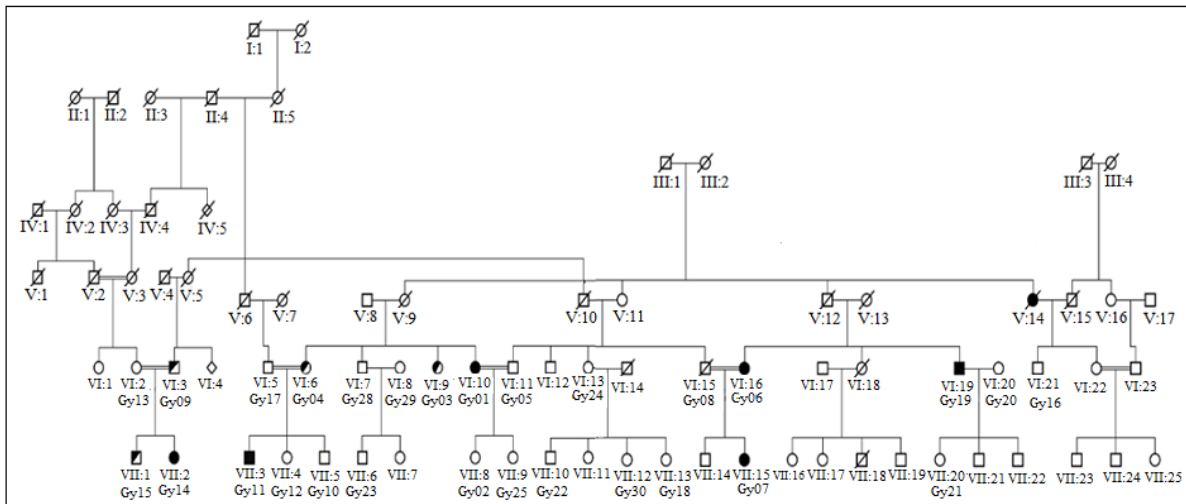
According to the main objective, the secondary objectives are:

1. Genotyping of the HLA complex by HLA Luminex<sup>®</sup> system, and assess the presence of risk genotype locus for MS in the studied family.
2. Localize the mutations in the candidate genes previously selected from an exome study.
3. Validation of genetic changes using Sanger sequencing or genotyping by RT-PCR assay with TaqMan<sup>®</sup> probes.
4. Genotype-phenotype correlation in patients studied.

# 3. MATERIALS I MÈTODES

## 3.1 DESCRIPCIÓ DE LA FAMÍLIA

El treball actual es basa en l'estudi d'una família d'ètnia gitana amb una alta endogàmia tal i com s'observa en l'arbre genealògic de la família (*Figura 1*) on es troba un número elevat d'individus amb EM. Es disposa de mostres de DNA de 28 membres de la família, dels quals aproximadament la meitat, estan afectats d'EM. La família d'estudi és molt extensa i alguns membres de la família viuen a diverses zones d'Espanya. En l'estudi s'inclou els membres de la família majors d'edat que han acceptat la participació de l'estudi, havent firmat el consentiment informat. També s'han inclòs els menors d'edat només si estan afectats, i sempre que el mateix afectat i els seus progenitors o tutors hagin firmat el consentiment informat. Els pacients han estat estudiats clínicament i amb proves complementàries com RM, així com també s'han recollit variables demogràfiques (edat, sexe, edat d'inici de la malaltia) i variables clíniques (taxa de brots anual, localització dels brots). Tenint en compte les característiques de la malaltia, que pot estar present sense que l'afectat hagi tingut símptomes, s'ha fet una RM a tots els membres no afectats majors d'edat que han participat en l'estudi, a fi de detectar els fenotips intermedis de la malaltia. Es van observar molts casos de RIS. Aquest fet suggereix la possibilitat d'un patró d'herència autosòmic recessiu o semidominant en aquesta família per l'EM.



*Figura 1:* Arbre genealògic de la família. Els afectats d'EM es representen de color negre i els sans de blanc. Els membres RIS estan marcats amb ambdós colors. La codificació Gy01, Gy02, etc, assenyalen els individus que han participat en l'estudi.

## 3.2 OBTENCIÓ DEL DNA DELS SUBJECTES

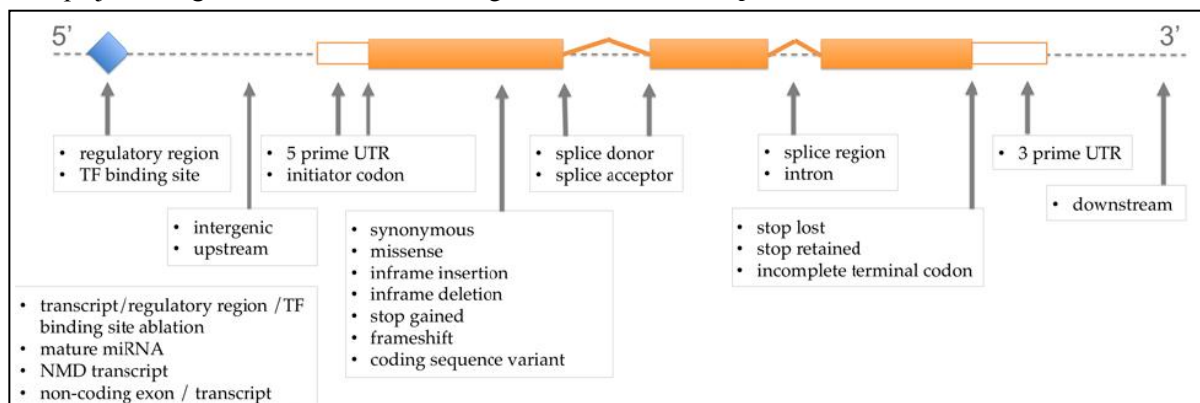
Per tal de poder realitzar l'estudi genètic, cal obtenir el DNA dels membres de la família que participen en l'estudi. En primer lloc, s'han realitzat extraccions de sang per punció venosa per part d'un infermer/a en un tub d'extracció de sang de 10ml amb anticoagulant EDTA (BD Vacutainer™). Seguidament s'ha obtingut el *buffy coat* o capa leucocitària un cop s'han centrifugat els tubs

d'extracció a 19°C, 2000g durant 10 minuts. Posteriorment s'ha extret i purificat el DNA mitjançant el kit comercial QIAamp® DNA Blood Mini Kit de la companyia QIAGEN®. S'ha seguit el *Spin Protocol* del *QIAamp DNA Mini and Blood Mini Handbook* que proporciona la mateixa companyia. Finalment, s'ha quantificat el DNA usant l'espectrofotòmetre NanoDrop™ per determinar si la puresa i la concentració del DNA extret de totes les mostres són les adequades. Les mostres s'han conservat a -20°C i s'han identificat mitjançant un codi numèric per mantenir l'anonimat d'aquestes.

### 3.3 SELECCIÓ DE GENS DE L'EXÒMICA

Un cop obtinguts els resultats de l'exòmica enviats per Sistemas Genómicos S.L en els quals hi ha una llista de 250 gens amb diferent genotip entre els afectats i l'individu sa, es va fer una selecció i prioritització dels gens candidats a ser validats. Per aquesta prioritització es va tenir en compte la funció de la proteïna resultant i la conseqüència funcional del canvi genètic. La funció de la proteïna d'aquests gens hauria d'estar relacionada amb la malaltia, i per tant, ha estat un equip de neuròlegs juntament amb els investigadors del projecte els encarregats d'escollir i fer una prioritització dels gens candidats. Durant l'estada, he après a examinar la funció i implicació dels gens. Per obtenir aquesta informació, s'han utilitzat dos bases de dades bioinformàtiques: *Pubmed* (<http://www.ncbi.nlm.nih.gov/pubmed>) i *Genecards* (<http://www.genecards.org/>).

La majoria dels canvis genètics identificats són SNPs, però pel que fa a la conseqüència d'aquests canvis, se n'han observat de molts tipus diferents (*Figura 2*). No obstant això, els canvis més interessants són: aquells que afecten a donadors o acceptors de *splicing*, pèrdua o guany del codó stop, variants en el marc de lectura (*frameshift variant*) degut a una inserció o deleció no múltiple de tres parells de bases, canvis en el codó d'inici de la traducció (ATG) i canvis d'un aminoàcid (*missense variant*). Aquests canvis, són de gran interès, ja que a priori fan pensar que són els més perjudicials perquè poden provocar que la proteïna resultant no sigui funcional, no sigui tan eficient, que s'expressi en baixa quantitat o fins i tot evitar la traducció d'aquesta. Per altra banda, tenen menys rellevància tots aquells canvis sinònims deguts a la degeneració del codi genètic, insercions o deleccions que mantenen el marc de lectura, o aquells polimorfismes que afectin a zones intròniques flanquejants, regions 5'UTR i 3'UTR, regions *downstream* i *upstream* i RNAs no codificants.



*Figura 2:* Classificació dels polimorfismes d'acord amb la conseqüència funcional o la posició en els transcrits.

### 3.4 IDENTIFICACIÓ DEL POLIMORFISME EN LES BASES DE DADES

Per seguir l'estudi genètic, el següent pas és cercar la posició del polimorfisme en la seqüència del gen per tal de poder dissenyar els *primers* i amplificar la regió d'interès. La gran majoria dels canvis detectats en l'exòmica són polimorfismes que ja han estat descrits i es consideren no patològics o comuns en la població. Aquests canvis es troben registrats amb la nomenclatura *rs* a la dbSNP (*Database of Single Nucleotide Polymorphisms*) (<http://www.ncbi.nlm.nih.gov/snp/>). D'altra banda, també s'han detectat una sèrie de canvis no descrits prèviament com a polimorfismes, codificats per la nomenclatura HGVS (*Human Genome Variation Society*). La nomenclatura HGVS<sub>c</sub> codifica pel canvi en la seqüència de DNA del gen i la HGVS<sub>p</sub> codifica el canvi en la seqüència d'aminoàcids de la proteïna. Per cercar la regió d'interès, s'ha utilitzat la base de dades del Ensembl (<http://www.ensembl.org/index.html>). És important partir de la seqüència genòmica del gen (amb introns i exons), per tal de poder amplificar el fragment d'interès.

### 3.5 AMPLIFICACIÓ DELS FRAGMENTS PER PCR

#### 3.5.1 DISSENY DE PRIMERS

Un cop s'han identificat els polimorfismes dels gens candidats, cal obtenir els *primers* de la regió d'interès per poder amplificar els fragments i continuar amb la validació de l'exòmica. Per construir els *primers* o encebadors (*forward* i *reverse*) d'una seqüència de DNA s'ha utilitzat l'aplicació bioinformàtica Primer3 (<http://primer3.ut.ee/>). Si es copia la seqüència d'interès del Ensembl amb els *links* dels polimorfismes descrits, el programa és capaç de reconèixer aquests polimorfismes i no es construeixen *primers* a sobre dels mateixos. Això permet fer *primers* d'una regió més conservada i assegurar que la regió s'amplifiqui en tots els pacients. Cada primer té uns 20pb i els fragments d'interès són d'uns 200-500pb aproximadament. El programa també determina la temperatura d'*annealing* dels dos *primers* ja que interessa que la temperatura sigui similar en els dos *primers* perquè funcionin de la mateixa manera en la reacció de PCR. Finalment, es demanen els *primers* a l'empresa Sigma-Aldrich<sup>TM</sup> per tal que els seqüenciïn i els enviïn.

#### 3.5.2 POSADA A PUNT DE LA PCR

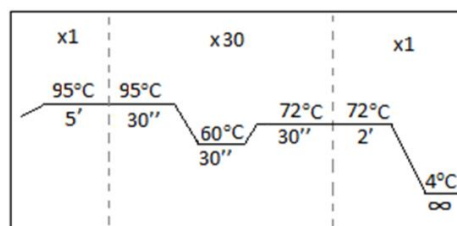
- **Objectiu de la tècnica:** Ajustar les condicions de l'amplificació dels fragments per PCR.
- **Fonament de la tècnica:** Consisteix en una PCR (veure metodologia de la PCR) però el termociclador utilitzat permet fer un gradient de temperatura (de 57°C a 62°C) per determinar la temperatura òptima per l'*annealing* dels *primers* al DNA motlle. A més temperatura, fins un límit, s'augmenta la rigorositat de l'*annealing* i s'eviten els mals aparellaments dels *primers*. Primer es realitza una amplificació d'una mostra per cada fragment, sense l'addició de MgCl<sub>2</sub>. Si en la comprovació de la PCR no s'observa una bona amplificació, es repeteix l'amplificació amb unes condicions que afavoreixin la reacció, com per exemple: l'addició de MgCl<sub>2</sub>, augmentar el nombre de cicles de la PCR, augmentar el *timing* de cada cicle, etc.

### 3.5.3 PCR I COMPROVACIÓ EN GEL D'AGAROSA

- **Objectiu de la tècnica:** Amplificar els fragments d'interès de les mostres de DNA.
- **Protocol:**
  1. Preparar la Mix en un tub *epENDORF* de 1,5ml segons les condicions de la *Taula 1*. Afegir tots els reactius excepte el DNA que es posa individualment en els tubs de 0,2ml per reaccions de PCR (MicroAmp®). Cal preparar un control negatiu i calcular el volum per alguna mostra de més per evitar que s'esgoti la Mix per l'error del pipeteig.
  2. Fer un spin a les mostres abans de posar-les en el termociclador. Executar el programa d'amplificació específic per cada fragment de DNA. El programa estàndard d'amplificació (*Figura 3*) es basa en una amplificació a 60°C de temperatura d'*annealing*, 30 segons cada etapa del cicle i 30 cicles.
  3. Finalment, es comprova la PCR mitjançant una electroforesi en gel d'agarosa al 2%.

*Taula 1:* Components i volums (µl/reacció) per l'amplificació d'una mostra de DNA.

Components Mix	Condicions x1
Master Mix	10µl
Primers (10µM)	0,5µl de cada primer (forward i reverse)
MgCl <sub>2</sub> (25mM)	Opcional
DNA (100ng/µl)	1µl
H <sub>2</sub> O d'injectable	Volum fins a 25µl
<b>Total</b>	<b>25µl</b>



*Figura 3:* Programa estàndard d'amplificació.

## 3.6 GENOTIPATGE

### 3.6.1 GENOTIPATGE DE HLA-DQB1 I HLA-DRB1

- **Objectiu de la tècnica:** Determinar si els pacients tenen almenys una còpia dels al·lels HLA-DRB1\*15:01 o HLA-DQB1\*06:02.
- **Fonament de la tècnica:** Per determinar quins al·lels del HLA estan presents en una mostra de DNA, s'ha utilitzat el *kit* de tipificació LIFECODES HLA-SSO que utilitza oligonucleòtids amb especificitat de seqüència (*Sequence-Specific Oligonucleotide*, SSO). La tècnica es basa en una amplificació del locus HLA amb diversos parells de primers (PCR multiplex) seguit d'una hibridació amb sondes específiques pels SNPs unides a microesferes (*beads*). Cada una d'aquestes es distingeix per un color de fluorescència característic. Aleshores aquestes sondes s'hibriden amb el DNA amplificat i es llegeix per citometria de flux amb el sistema Luminex®. L'aparell genera dos làsers, un de vermell i un de verd. El vermell identifica els diferents *beads* i el verd capta la fluorescència que genera l'estreptavidina quan s'uneix a una doble cadena de DNA.



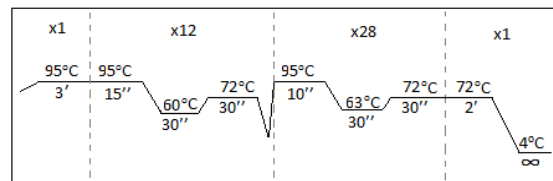
- **Protocol:** S'ha seguit el protocol de l'empresa Diagnostica Longwood.

### Amplificació del DNA

1. Preparar la Mix en un tub *epENDORF* de 1,5ml segons les condicions de la *Taula 2*. Afegir tots els reactius excepte el DNA que es posa individualment en els tubs de 0,2ml (MicroAmp®). A part de les mostres, cal preparar un control negatiu, control positiu i calcular el volum per alguna mostra de més per evitar que s'esgoti la Mix degut a l'error del pipeteig. No totes les mostres de DNA estan a la mateixa concentració, per tant s'han de fer els càlculs corresponents per afegir 80ng de DNA.
2. Fer un *spin* a les mostres i executar el programa d'amplificació en el termociclador (*Figura 4*).

*Taula 2:* Components i volums (µl/reacció) per l'amplificació d'una mostra de DNA.

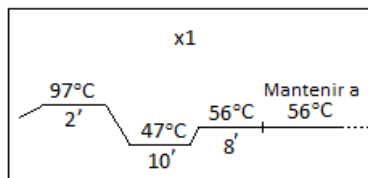
Components Mix	Condicions x1
Master Mix	6µl
Taq DNA polimerasa	0,2µl
Mostra de DNA	80ng
H <sub>2</sub> O d'injectable	Volum fins a 20µl
<b>Total</b>	<b>20µl</b>



*Figura 4:* Programa d'amplificació del locus HLA.

### Hibridació

1. Pre-escalfar els *beads* 10 minuts a 55-60°C, sonicar-los en el bany d'ultrasons durant 1 minut i vortejar-los per desfer els agregats de les microesferes.
2. Afegir a cada pou: 5µl de DNA amplificat + 15µl dels *beads*.
3. Executar el programa d'hibridació en el termociclador (*Figura 5*).



*Figura 5:* Programa d'hibridació.

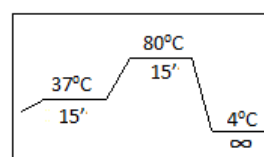
4. Un cop ha acabat el programa d'hibridació, mantenint els tubs en el termociclador a 56°C, afegir 170µl de la Solució d'estreptavidina a les mostres.
5. Abans d'analitzar les mostres en el Luminex® (LABScan™ 100 Luminex XMAP™ Technology) cal fer els corresponents rentats i calibracions a l'aparell. Finalment s'analitzen els resultats amb el programa *Quick-Type*.

### 3.6.2 SEQÜENCIACIÓ PEL MÈTODE SANGER

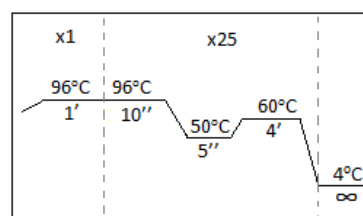
- **Objectiu de la tècnica:** Genotipatge dels polimorfismes mitjançant seqüenciació Sanger.
- **Protocol:**
  1. Purificar el producte de PCR barrejant 2,5µl d'aquest amb 1µl d'Exosap<sup>®</sup> en una placa de 96 pous MicroAmp<sup>®</sup> i executar el programa Exosap en el termociclador (*Figura 6*).
  2. Preparar la Mix en un tub *ependorf* de 1,5ml segons les condicions de la *Taula 3*. Afegir tots els reactius excepte el producte d'Exosap que es posa individualment en cada pouet d'una nova la placa MicroAmp<sup>®</sup>. A part de les mostres, cal preparar un control negatiu i calcular el volum per alguna mostra de més per evitar que s'esgoti la Mix degut a l'error del pipeteig. Fer un pols de centrifuga a la placa abans de posar-la al termociclador.
  3. Executar el programa de seqüenciació al termociclador (*Figura 7*).
  4. Precipitació del DNA:
    - a. Afegir 80µl de la solució: 78,1% d'etanol (95%), 3,8% d'acetat de sodi (3M i pH=4,6) i 18,1% d'H<sub>2</sub>O Milli-Q<sup>®</sup>.
    - b. Deixar a temperatura ambient 10-13 minuts i centrifugar a 2000g, 45 minuts.
    - c. Retirar el sobrenedant per inversió i fer un *spin* amb la placa capgirada.
    - d. Deixar assecar a temperatura ambient durant 15 minuts.
    - e. Afegir a cada mostra 150µl d'etanol al 70%.
    - f. Centrifugar a 2000g durant 10 minuts i retirar el sobrenedant.
    - g. Deixar assecar a temperatura ambient 15 minuts.
  5. Afegir 10µl de formamida a cada mostra.
  6. Llegir les mostres al seqüenciador i analitzar els resultats amb el programa *SeqScope*.

*Taula 3:* Components i volums (µl/reacció) per la reacció de seqüenciació d'una mostra de DNA.

Components Mix	Condicions x1
<b>BigDye</b>	0,5µl
<b>Buffer BigDye</b>	1µl
<b>Primer (5µM) forward o reverse</b>	0,25µl
<b>H<sub>2</sub>O d'injectable</b>	2,25µl
<b>Producte d'Exosap</b>	1µl
<b>Total</b>	<b>5µl</b>



*Figura 6:* Programa d'Exosap.



*Figura 7:* Programa de seqüenciació.

### 4.6.3 GENOTIPATGE PER ASSAY

- **Objectiu de la tècnica:** Genotipatge dels polimorfismes mitjançant *assays* amb sondes TaqMan<sup>®</sup> de la companyia Applied Biosystems.
- **Fonament de la tècnica:** Aquesta tècnica es basa en un genotipatge per RT-PCR (*Real Time PCR*) a partir de 2 primers (*forward* i *reverse*) que acoten molt específicament la zona del SNP

d'interès i un sonda TaqMan que hibrida sobre de la seqüència del polimorfisme. La sonda TaqMan està marcada amb dues molècules: una molècula fluorescent (R) i un captador o *quencher* (Q). El captador si està col·locat a prop de la molècula fluorescent absorbeix totes les emissions. En una situació de PCR, en el moment d'*annealing*, hibriden els tres oligonucleòtids a la regió complementaria del DNA motlle. El *primer forward* allarga la seqüència i en el moment que topi amb la sonda TaqMan, es separen les molècules R i Q, aleshores s'emetrà fluorescència. En funció de la fluorescència detectada, es pot conèixer la base de la posició d'interès, ja que ha hibridat una de les sondes, de les quals es coneix la base i el color associats.

- **Protocol:**

1. Preparar la Mix en un tub *epENDORF* de 1,5ml segons les condicions de la *Taula 4*. Afegir tots els reactius excepte el DNA que es posa individualment en la placa MicroAmp®. A part de les mostres, cal preparar un control negatiu, un control positiu i calcular el volum per alguna mostra de més per evitar que s'esgoti la Mix degut a l'error del pipeteig.
2. Fer un pols de centrífuga a la placa.
3. Analitzar la placa al termociclador RT (Sequence Detection System – Applied Biosystems).

*Taula 4:* Components i volums (µl/reacció) per la reacció RT-PCR d'una mostra de DNA.

Components Mix	Condicions x1
H <sub>2</sub> O d'injectable	10,875µl
TaqMan Universal PCR Genotyping Master Mix	12,5µl
Assay	0,675µl
DNA (5-20ng/µl)	1µl
<b>Total</b>	<b>25,05µl</b>

### 3.7 ESTADÍSTICA

L'objectiu de l'estudi és la detecció de gens de susceptibilitat a EM en la família. Per tant, s'espera veure uns canvis en algun genotip i que aquests estiguin més representats en els afectats que en els no afectats d'EM. Per comprovar si hi ha diferències significatives en aquesta distribució, s'ha realitzat un anàlisi estadístic mitjançant el programa SPSS. Per fer l'anàlisi, s'ha realitzat un contrast d'independència amb l'estadístic khi quadrat ( $\chi^2$ ) de *Pearson* o test exacte de *Fisher* segons conveniència. Una variable és el fet de ser afectat o no afectat per l'EM i l'altre variable és el genotip de l'individu el qual es troba dividit en tres categories: homozigot per l'al·lel *wild type*, heterozigot i homozigot per l'al·lel variant. Així doncs, un cop definides les freqüències observades en la taula de contingència, s'obté un *p-valor*. En l'anàlisi s'ha considerat que un *p-valor*<0,05 (interval de confiança del 95%) s'accepta la hipòtesi alternativa, és a dir, hi ha diferències significatives. I un *p-valor*>0,05 s'accepta la hipòtesi nul·la, no hi ha diferències significatives.

## 4. RESULTATS I DISCUSSIÓ

---

Es disposa d'una família amb molts individus afectats d'EM i amb un alt nivell de consanguinitat, el que fa pensar que el component genètic pot ser important per explicar la causa de la malaltia. Després d'analitzar els gens prèviament descrits i relacionats amb la susceptibilitat a EM, sense observar resultats concloents, es va procedir a la realització de l'estudi d'exòmica en un nombre reduït d'individus respecte el total de la família.

### 4.1 RESULTATS EXÒMICA I SELECCIÓ DE GENS CANDIDATS

L'estudi d'exòmica es va realitzar mitjançant la comparació de 4 mostres de membres afectats d'EM (Gy01, Gy06, Gy07, Gy11) amb 1 membre sa de la família, el qual conserva tots els polimorfismes propis de la família però que en principi no porta els patològics (Gy05). L'anàlisi d'exòmica es va realitzar en una empresa externa (Sistemas Genómicos S.L., València) i va donar com a resultat la identificació d'aproximadament 250 canvis genètics amb expressió diferenciada entre l'individu sa i els afectats per EM. La premissa de l'anàlisi era determinar que l'individu sa (Gy05) fos homozigot per l'al·lel de referència i com a mínim algun dels afectats tingui un genotip diferent.

Alguns dels paràmetres més importants dels resultats de l'exòmica en referència a cadascun dels canvis obtinguts són:

- **Nom del gen.**
- **Posició genòmica.**
- **Al·lel de referència (*wild type*) i al·lel variant.**
- **Genotip de cada mostra.**
- **Profunditat (*Depth*):** És la cobertura (*coverage*), és a dir, el nombre total de "lectures" d'un polimorfisme.
- **Probabilitat del genotip (*Var/Depth*):** Es determina la freqüència de l'al·lel variant respecte la cobertura. Aquest paràmetre és una *ratio* entre el número de cops que s'ha llegit l'al·lel variant i la profunditat. El seu valor oscil·la entre 1-0, de manera que un valor d'1 indica homozigot per l'al·lel variant, 0 homozigot per l'al·lel de referència i 0,5 heterozigot. Però en la majoria de casos, no s'observen aquests valors tan exactes, ja que sempre hi ha errors en la lectura. Així doncs, la probabilitat del genotip segons aquesta raó aritmètica està determinada en rangs:
  - 0-0,12: Probablement homozigot per l'al·lel de referència (*Probable homozygous for the reference allele*)<sup>1</sup>.
  - 0,12-0,35: Genotip dubtós, homozigot per l'al·lel de referència o heterozigot (*Uncertain heterozygous*)<sup>1</sup>.

---

<sup>1</sup> Entre parèntesis estan representats els genotips segons la nomenclatura dels resultats de l'exòmica.

- 0,35-0,65: Probablement heterozigot (*Probable heterozygous*)<sup>1</sup>.
  - 0,65-0,85: Genotip dubtós, heterozigot o homozigot per l'al·lel variant (*Uncertain homozygous for the variant allele*)<sup>1</sup>.
  - 0,85-1: Probablement homozigot per l'al·lel variant (*Probable homozygous for the variant allele*)<sup>1</sup>.
- **Codificació del canvi:** Segons la nomenclatura HGVS i *rs*.
  - **Conseqüència funcional del canvi genètic.**

Per tal de seleccionar els gens més interessants per a la validació en el total de la família, es va tenir en compte la conseqüència funcional del canvi genètic identificat i la funció de les proteïnes afectades. El meu treball s'ha basat en la validació dels canvis detectats en els següents 10 gens candidats: CALN1, CD207, EME1, EMP2, FAIM2, HLA-DQB1-AS1, IL17RC, IL27RA, KIR2DL1 i el TMEM175.

Per una banda, la *Taula 5* mostra els paràmetres més importants dels resultats de l'exòmica per aquests 10 gens estudiats. D'altra banda, la *Taula 6* mostra la conseqüència funcional del canvi genètic dels gens estudiats i la funció de la proteïna codificada la qual s'ha cercat utilitzant dos bases de dades bioinformàtiques: *Pubmed* (<http://www.ncbi.nlm.nih.gov/pubmed>) i *Genecards* (<http://www.genecards.org/>).

Tal i com s'observa en la *Taula 6*, la funció de la proteïna dels gens estudiats està relacionada ja sigui amb el sistema immunitari, les vies del SNC o en l'apoptosi. Per tant, és possible que un mal funcionament d'aquestes proteïnes estigui implicat en el desenvolupament de la malaltia. Pel que la conseqüència funcional del canvi genètic, en el cas dels gens TMEM175, IL27RA i CD207, presenten un canvi *missense*, per tant és més probable que s'alteri el correcte funcionament de la proteïna. El SNP (rs670771) del KIR2DL1, al localitzar-se en una *splice region variant*, pot influir en la síntesi del mRNA i en conseqüència en la síntesi de la proteïna. La inserció de 3 nucleòtids del gen EME1, manté la pauta de lectura del gen, és a dir, només s'incorpora un aminoàcid. Però podria ser que aquesta inserció afectés a l'estructura tridimensional de la proteïna, o bé, al centre actiu d'aquesta. Els SNPs del FAIM2 i CALN1 són variants intròniques, que malgrat no ésser regió codificant, pot contenir regions reguladores que afectin indirectament la transcripció d'algun gen. Els SNPs dels gens HLA-DQB1-AS1 i EMP2 són variants al extrem 3' del gen (*downstream*) i aquests canvis tenen menys possibilitat de perjudicar la síntesi de la proteïna. Finalment, els SNPs dels gens IL17RC i KIR2DL1 (rs75200129) provoquen un canvi sinònim degut a la degeneració del codi genètic, i al mantenir-se la seqüència d'aminoàcids de la proteïna, a priori, no és perjudicial per la síntesi de la proteïna. No obstant això, tot i que alguns gens han presentat conseqüències funcionals de menys rellevància, s'han analitzat per la funció d'aquestes proteïnes.

Taula 5: Resultats de l'exòmica pels gens analitzats.

Gen	Cromosoma	Posició Genòmica	Al·lel Ref.	Al·lel Var.	Genotip Gy05	Depth Gy05	Var/Depth Gy05	Genotip Gy01	Depth Gy01	Var/Depth Gy01	Genotip Gy06	Depth Gy06	Var/Depth Gy06	Genotip Gy07	Depth Gy07	Var/Depth Gy06	Genotip Gy07	Depth Gy07	Var/Depth Gy06	Genotip Gy07	Depth Gy06	Var/Depth Gy06	Genotip Gy01	Depth Gy01	Var/Depth Gy01	Genotip Gy01	Depth Gy01	Var/Depth Gy05	Genotip Gy05	Al·lel Ref.	Al·lel Var.
<b>CALN1</b>	7	71674018	G	A	Probable Homo Ref.	32	0,09	Uncertain Hetero	42	0,21	Uncertain Hetero	32	0,19	Probable Hetero	27	0,37	Probable Hetero	32	0,21	Uncertain Hetero	56	0,21	Uncertain Hetero	rs200910610							
		71674019	A	G	Probable Homo Ref.	32	0,09	Uncertain Hetero	40	0,23	Uncertain Hetero	31	0,26	Uncertain Hetero	26	0,31	Uncertain Hetero	26	0,2	Uncertain Hetero	54	0,2	Uncertain Hetero	rs201883338							
<b>CD207</b>	2	71061091	G	A	Homo Ref.	16	0	Probable Hetero	16	0,38	Probable Hetero	13	0,54	Uncertain Homo Var.	13	0,69	Uncertain Homo Var.	25	0,52	Probable Hetero	25	0,52	rs115727537								
<b>EME1</b>	17	48452978	A	AAGC	Homo Ref.	-	0	Probable Hetero	39	0,54	Probable Hetero	21	0,5	Probable Hetero	17	0,5	Probable Hetero	19	0,5	Probable Hetero	19	0,5	rs144820580								
<b>EMP2</b>	16	10626607	G	A	Homo Ref.	22	0	Probable Hetero	23	0,57	Uncertain Hetero	25	0,16	Probable Hetero	22	0,5	Probable Hetero	21	0,62	Probable Hetero	21	0,62	rs12930183								
		10626608	A	G	Probable Homo Ref.	8	0,12	Probable Hetero	8	0,5	Uncertain Hetero	7	0,14	Probable Hetero	2	0,5	Probable Homo Var.	7	1	Probable Homo Var.	7	1	rs12934615								
<b>FAIM2</b>	12	50273450	T	A	Homo Ref.	6	0	Uncertain Homo Var.	6	0,83	Uncertain Homo Var.	9	0,67	Probable Homo Var.	5	1	Probable Homo Var.	1	1	Probable Homo Var.	1	1	rs116224073								
<b>HLA-DQB1-AS1</b>	6	32627460	T	G	Homo Ref.	-	0	Probable Homo Var.	4	1	Probable Homo Var.	3	1	Probable Homo Var.	3	1	Probable Homo Var.	4	0,75	Uncertain Homo Var.	4	0,75	rs9273430								
<b>IL17RC</b>	3	9975247	C	G	Homo Ref.	-	0	Probable Homo Var.	1	1	Probable Homo Var.	3	1	Probable Homo Var.	3	1	Probable Homo Var.	1	1	Probable Homo Var.	1	1	rs183956								
<b>IL27RA</b>	19	14142763	T	G	Homo Ref.	6	0	Uncertain Hetero	6	0,1	Uncertain Hetero	8	0,25	Uncertain Hetero	5	0,2	Uncertain Hetero	6	0,33	Uncertain Hetero	6	0,33	ENST000002633 79.2:c.79T>G								
<b>KIR2DL1</b>	19	55294931	T	C	Homo Ref.	13	0	Uncertain Hetero	13	0,23	Uncertain Homo Var.	5	0,8	Uncertain Homo Var.	4	0,75	Uncertain Homo Var.	11	0,27	Uncertain Hetero	11	0,27	rs 670771								
		55294969	A	G	Homo Ref.	22	0	Probable Hetero	31	0,35	Probable Homo Var.	18	0,94	Probable Homo Var.	17	0,94	Probable Homo Var.	30	0,23	Uncertain Hetero	30	0,23	rs75200129								
<b>TMEIM175</b>	4	951812	C	T	Homo Ref.	11	0	Uncertain Hetero	9	0,22	Probable Hetero	10	0,6	Probable Homo Var.	10	0,9	Probable Hetero	11	0,55	Probable Hetero	11	0,55	rs147975675								

Taula 6: Conseqüència funcional del canvi genètic i funció de la proteïna dels gens estudiats.

Gen	Conseqüència funcional del canvi genètic	Funció de la proteïna
<b>CALN1</b>	rs200910610 - Variant intrònica.	Proteïna d'unió al calci en el SNC (calineurona). Involucrada en la fisiologia de les neurones, en la memòria i aprenentatge.
	rs201883338 - Variant intrònica.	
<b>CD207</b>	<i>Missense variant.</i> HGVS: ENSP00000386378.3:p.Arg84His. (canvi d'una Arginina per una Histidina en la posició 84 de la seqüència d'aminoàcids.)	La proteïna codificada és la Langerina i s'expressa en les cèl·lules de Langerhans, que són cèl·lules dendrítiques de l'epidermis i la mucosa (cèl·lules presentadores d'antígens).
<b>EME1</b>	<i>Inframe insertion</i>	Endonucleasa implicada en la reparació de danys en el DNA i en el manteniment de l'estabilitat genòmica.
<b>EMP2</b>	rs12930183 - Variant a la regió 3' UTR.	Proteïna epitelial de membrana. S'ha associat a: l'endocitosi, senyalització cel·lular, proliferació cel·lular, migració cel·lular, adhesió cel·lular i mort cel·lular.
	rs12934615 - Variant a la regió 3' UTR.	
<b>FAIM2</b>	Variant intrònica.	Proteïna antiapoptòtica.
<b>HLA-DQB1-AS1</b>	Variant a la regió 3' UTR	No codifica per cap proteïna. No hi ha descrita una funció però el fet de ser un RNA antisentit de la seqüència del HLA-DQB1, fa pensar que pot fer disminuir la presència d'aquest.
<b>IL17RC</b>	Canvi sinònim.	Receptor de la Interleuquina 17.
<b>IL27RA</b>	<i>Missense variant</i> HGVS: ENSP00000263379.1:p.Phe27Val (canvi d'una Fenilalanina per una Valina en la posició 27 de la seqüència d'aminoàcids).	Subunitat alfa del receptor de la Interleuquina 27.
<b>KIR2DL1</b>	rs670771 - Variant en una regió de <i>splicing</i> .	Receptor tipus immunoglobulina expressat en les cèl·lules assassines naturals (NK) i en els limfòcits T, implicat en la prevenció de la lisi cel·lular.
	rs75200129 - Canvi sinònim.	
<b>TMEM175</b>	<i>Missense variant.</i> HGVS: ENSP00000264771.4:p.Ser348Leu (canvi d'una Serina per una Leucina en la posició 348 de la seqüència d'aminoàcids).	Proteïna transmembrana. Associada amb la malaltia de Parkinson.

## 4.2 VALIDACIÓ DE L'EXÒMICA

Com ja s'ha comentat, tot estudi d'exòmica s'ha de validar mitjançant un altre mètode de seqüenciació per corroborar resultats, ja que l'exòmica pot cometre errors (Rabbani et al., 2014). Per tant, s'han validat els genotips d'interès per la resta d'individus de la família.

### 4.2.1 DISSENY DE PRIMERS

Previ a l'estudi de validació, cal identificar els canvis en la seqüència del gen per tal d'escollir la regió on es dissenyaran els primers i que finalment s'amplificarà i es seqüenciarà. En la Taula 7 s'observen els primers dissenyats per l'amplificació dels fragments mitjançant el programa Primer3 (<http://primer3.ut.ee/>).

Taula 7: Característiques dels primers dissenyats per l'amplificació dels fragments d'interès.

Gen	Primer forward	Primer reverse	Fragment (pb)	T <sup>a</sup> annealing (°C) (Forward - Reverse)
<b>CALN1</b>	5'_AGAAGTCAAGG GAGAAGGGG_3'	5'_GCTGGGGATTTG AAAGGCAT_3'	296	58,34 - 58,80
<b>CD207</b>	5'_TCAGCCTATTCC TTCCCTTC_3'	5'_CTGGGATTTGGGC ATTTAAGGT_3'	374	58,87 - 58,89
<b>EMP2</b>	5'_CGCAAATAGAG TTCCGGAGC_3'	5'_TCTCCTCTCCTCC CATCACAG_3'	357	59,07 - 58,99
<b>HLA-DQB1-AS1</b>	5'_TGAAACACCAG TGACCAAGG_3'	5'_ACCTTGAAAAGA CACTGAAAGCA_3'	225	57,95 - 58,98
<b>IL17RC</b>	5'_CCAACCTGCCAG ACTTCCT_3'	5'_CCCGCCACTGAA ACGGAG_3'	256	59,24 - 60,74
<b>IL27RA</b>	5'_GCAGAGAGTGT GTGTGTGTG_3'	5'_CCATCCCCAGTG TCGCC_3'	581	59,07 - 60,44
<b>KIR2DL1</b>	5'_GGGACCTCAGC CACCTATG_3'	5'_TGTCTTGGGCCTC TGAGAAG_3'	359	59,17 - 59,02
<b>TMEM175</b>	5'_GATGTGAAGGA GAGGTTTCAGC_3'	5'_AGATGCTGGCCA GGAAGAT_3'	298	58,88 - 59,78

En la taula hi ha representada la seqüència dels primers, la mida del fragment amplificat i la temperatura d'annealing dels primers per l'amplificació dels fragments que comprenen els polimorfismes dels gens validats pel mètode Sanger.

En el mateix moment que es dissenyen els primers, es determina el primer que s'utilitzarà per fer la reacció de seqüenciació. Pels fragments dels gens CALN1, CD207, HLA-DQB1-AS1, IL17RC i TMEM175 s'ha utilitzat el primer forward. D'altra banda, pels fragments dels gens EMP2, IL27RA i KIR2DL1 s'ha utilitzat el primer reverse. A la reacció de seqüenciació només es pot utilitzar un primer, en funció del que s'utilitzi s'obté la seqüència directe o inversa. Per escollir quin primer és el més adequat s'analitza si dins del fragment amplificat hi ha polimorfismes *indel*, o moltes repeticions d'un mateix nucleòtid, i en quina posició es troba respecte el polimorfisme d'interès. En el cas dels fragments del FAIM2 i EME1, s'han observat una gran quantitat de *indels* a prop del SNP d'interès. És per aquest motiu que aquests dos gens s'han genotipat per *assay*. Cal dir que ja s'havia intentat seqüenciar aquests dos gens pel mètode Sanger però no va ser possible llegir les seqüències, degut a l'elevada presència de dobles pics.

#### 4.2.2 POSADA A PUNT DE LA PCR

Per ajustar les condicions de l'amplificació dels fragments per PCR, en primer lloc s'han preparat 4 reaccions d'una mostra de DNA per cada fragment. Aquests tubs s'han col·locat en les files corresponents del termociclador per tal de que el rang de temperatura d'annealing quedi de 61°C a 57°C. Un cop finalitzada la PCR s'ha fet la comprovació en gel d'agarosa. Alguns fragments han mostrat una bona amplificació, i per tant s'ha determinat la temperatura d'annealing amb un programa de 30 cicles. En els casos en que la posada a punt no ha donat una bona senyal d'amplificació, s'ha afegit 1µl de MgCl<sub>2</sub> observant així una bona amplificació en el cas del TMEM175.



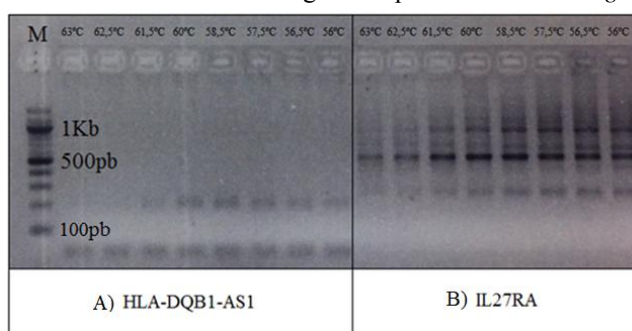
En el cas del HLA-DQB1-AS1 s'han observat dues bandes, una corresponent al fragment amplificat (aproximadament 200pb) però amb una banda poc marcada i l'altre banda té una mida inferior a 100pb. Aquesta banda correspon als *primers* hibridats entre ells, i això indica que els *primers* no han hibridat amb el fragment d'interès (*Figura 9A*). En aquest cas, per tal d'afavorir la reacció, s'ha incrementat la quantitat de MgCl<sub>2</sub> de la mostra, els cicles d'amplificació i el temps d'*annealing* a 1 minut, aconseguint així una bona amplificació del fragment. D'altra banda en el cas del IL27RA s'observen múltiples bandes. Això indica que el *primer* hibrida en diferents llocs, de manera inespecífica, per això s'amplifiquen fragments de diferents mides (*Figura 9B*). Per millorar aquesta reacció s'han establert condicions més restrictives, en aquest cas s'ha incrementat la temperatura d'*annealing*.

Les condicions finals per l'amplificació dels fragments analitzats estan representades en la *Taula 8*.

*Taula 8:* Condicions per l'amplificació dels gens estudiats.

Gen	Programa	MgCl <sub>2</sub>
CALN1	60°C i 30c	-
CD207	60°C i 30c	-
EMP2	60°C i 30c	-
HLA-DQB1-AS1	60°C, 1' d' <i>annealing</i> i 35c	1µl
IL17RC	60°C i 30c	-
IL27RA	61°C i 30c	-
KIR2DL1	60°C i 30c	-
TMEM175	60°C i 30c	1µl

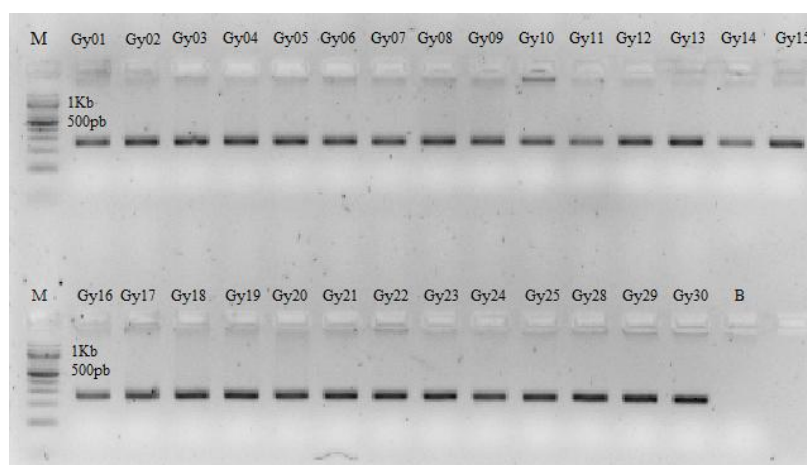
*Figura 9:* Comprovació de la posada a punt de dos fragments d'una mostra amb tot el rang de temperatura d'*annealing*.



Els °C fan referència a la temperatura d'*annealing*. La c són els cicles d'amplificació.

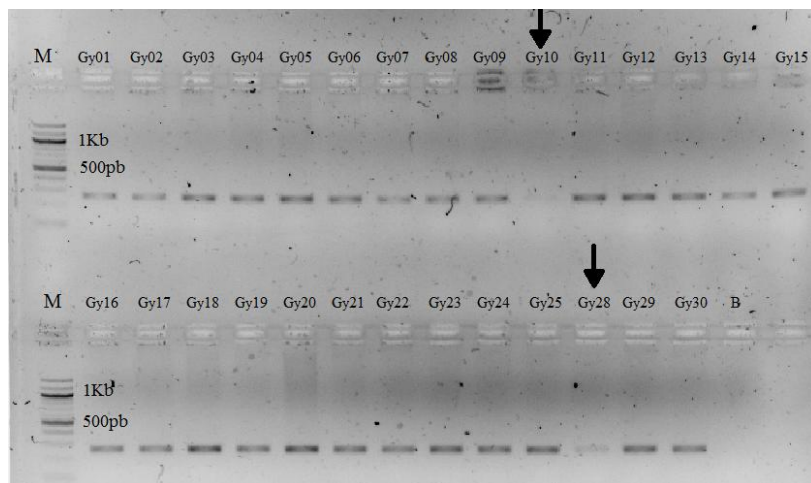
#### 4.2.3 COMPROVACIÓ DE L'AMPLIFICACIÓ DELS FRAGMENTS

Un cop determinades les condicions d'amplificació dels fragments d'interès, s'han amplificat per la resta de mostres de la família. La *Figura 10* és un exemple d'aquesta amplificació.

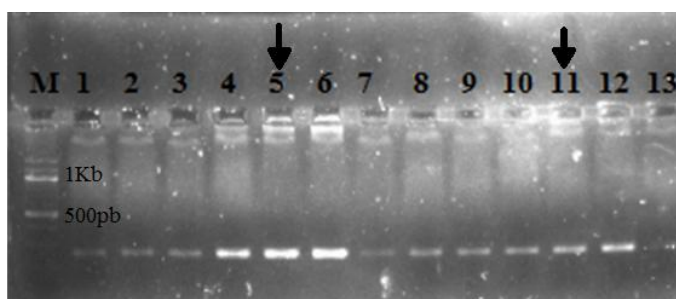


*Figura 10:* Comprovació de l'amplificació d'un fragment per totes les mostres.

Cal comentar que en alguns casos s'ha hagut de repetir alguna mostra, ja que el DNA d'alguns pacients no han amplificat en determinats fragments. Tot i que com s'ha observat en diverses vegades, el Gy10 i Gy28 han presentat complicacions a l'hora d'amplificar alguns fragments. Aquest fet és degut a que la concentració d'aquestes mostres de DNA mesurada amb el NanoDrop són bastant baixes, per tant una opció per corregir aquest error és afegir més quantitat de DNA. Aquest problema s'ha observat amb el fragment del gen IL17RC (*Figura 11*). En aquest cas, per tal d'amplificar el fragment d'interès per aquestes dos mostres de DNA, s'ha repetit el procés d'amplificació amb diferents quantitats de DNA i amb DNAs procedents de diferents extraccions fins a obtenir una correcta amplificació del fragment (*Figura 12*). S'ha observat que les mostres que millor han amplificat corresponen al carril 5 pel Gy10 (1µl de DNA rehidratat amb solució AE, un tampó d'elució del *kit* d'extracció de QIAGEN) i al carril 11 pel Gy28 (3µl de DNA).



*Figura 11:* Comprovació de l'amplificació del fragment del gen IL17RC.



*Figura 12:* Comprovació de l'amplificació del fragment del gen IL17RC en un gel d'agarosa 2% pel Gy10 i Gy28.

Llegenda de la Figura 12:

1. Gy10 0,5µl DNA
2. Gy10 1µl DNA
3. Gy10 1,5µl DNA
4. Gy10 0,5µl DNA AE
- 5. Gy10 1µl DNA AE**
6. Gy10 1,5µl DNA AE
7. Gy28 0,5µl DNA
8. Gy28 1,5µl DNA
9. Gy28 2µl DNA
10. Gy28 2,5µl DNA
- 11. Gy28 3µl DNA**
12. Gy1 1µl DNA (Control +)
13. Blanc

#### 4.2.4 SEQÜENCIACIÓ PEL MÈTODE SANGER

La majoria dels gens han sigut seqüenciats utilitzant el mètode Sanger. Els resultats d'aquesta seqüenciació estan representats en electroferogrames. La *Figura 13* és un exemple d'un electroferograma d'un dels fragments analitzats.

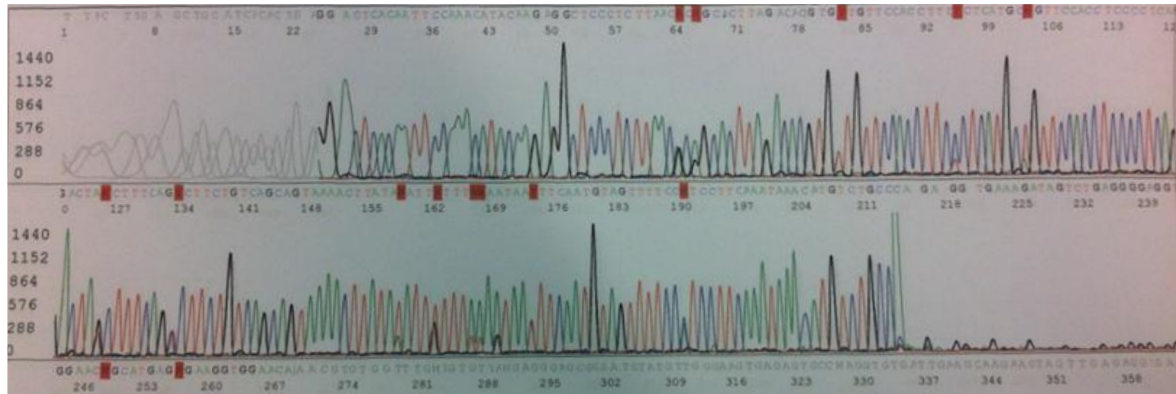


Figura 13: Exemple d'un electroferograma per la seqüenciació d'un fragment per una mostra.

El principal problema d'aquesta tècnica és que molts dels fragments analitzats presenten variats polimorfismes, tant *indels* com SNPs, a prop del canvi d'interès. Aquest fet ha dificultat la lectura de les seqüències i, en alguns casos, s'ha hagut de repetir l'amplificació i seqüenciació d'alguns fragments. Una possible solució a aquest problema seria genotipar per *assay*. Però, a part de ser una tècnica cara, no es permeten dissenyar *assays* per genotipar fragments els quals hi ha polimorfismes molt propers al SNP d'interès. A més, en la lectura dels electroferogrames d'alguns gens estudiats, s'han detectat una sèrie de problemes o fets que s'han de tenir en compte per interpretar els resultats i que s'esmenten a continuació:

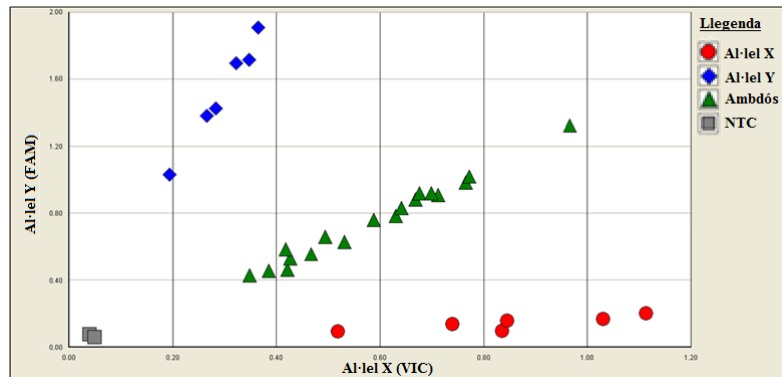
- **Soroll de fons:** Una possible causa de la presència de pics de baixa intensitat és la contaminació d'un altre DNA o per l'altre *primer* utilitzat en la PCR que no s'ha acabat d'eliminar. També, pot ser que la PCR no sigui específica i hi hagin amplicons contaminants en la mostra.
- **No s'ha detectat senyal en l'electroferograma:** Això pot ser degut a que el *primer* no ha reconegut correctament el motlle, baixa concentració o mal estat de DNA o de *primer*.
- **Polymerase slip:** Quan hi ha la presència de moltes repeticions d'un nucleòtid, la polimerasa "rellisca" i introdueix nucleòtids de més, dificultant la lectura de la seqüència.
- **Heterozigots:** Presenten un doble pic en una posició determinada. En principi s'espera trobar una intensitat del 50% per cada al·lel, però molts cops s'observa un pic més destacat.
- **Indels:** La seqüència es llegeix bé però a partir d'un *indel*, l'electroferograma mostra dobles pics i es complica la lectura de la seqüència. Això succeeix quan la inserció o deleció és en heterozigosi.

#### 4.2.5 ASSAY

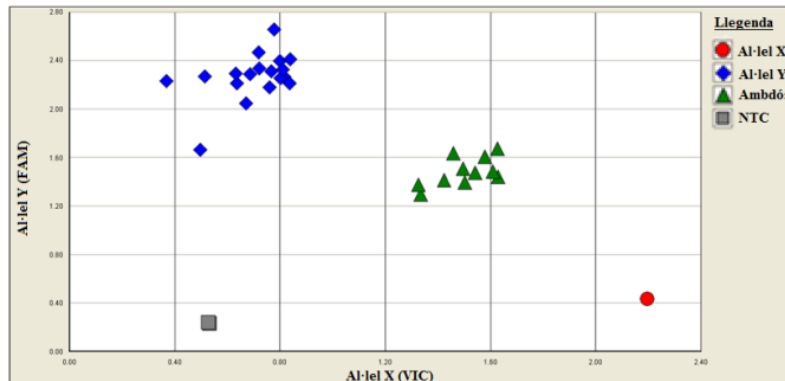
Pel genotipatge d'aquells fragments els quals comprenen molts SNPs i *indels* en el fragment amplificat, i que per tant, presenten dificultats en la lectura de la seqüenciació per Sanger, s'ha genotipat per RT-PCR utilitzant la tecnologia dels *assays* mitjançant sondes TaqMan<sup>®</sup>. El resultat de l'*assay* s'analitza amb el programa del ordinador 7000 system SDS Software, que proporciona el

genotip d'interès dels fragments dels gens EME1 i FAIM2. A més, proporciona una gràfica (*Figura 14 i 15*) que mostra la distribució dels genotips de les mostres pel gen estudiat. Aquesta distribució s'obté mitjançant la mesura de l'absorbància dels dos marcadors (VIC i FAM) associats cadascun a un dels dos al·lells. En aquesta gràfica, s'observen 4 núvols de punts que representen: els individus homozigots per l'al·lel Y, homozigots per l'al·lel X, heterozigots i els NTC (*No Template Control*). El NTC és una mostra que no conté DNA (control negatiu) i determina el soroll de fons.

En el cas del EME1, el marcador VIC està associat a l'al·lel A que és el *wild type* i el marcador FAM a l'al·lel variant AAGC. Per tant quan es detecten ambdós marcadors, l'individu és heterozigot (A/AAGC). La *Figura 14* il·lustra que la majoria dels individus són heterozigots, però també hi ha amb menor proporció, individus homozigots per l'al·lel de referència i variant. Pel que fa el FAIM2, el marcador VIC està associat a l'al·lel T que és el *wild type* i el marcador FAM a l'al·lel variant. En aquest cas la *Figura 15* mostra que tots els individus són homozigots per variant o bé heterozigots, excepte un individu que és homozigot per l'al·lel de referència.



*Figura 14:* Discriminació al·lèlica pel fragment del EME1.



*Figura 15:* Discriminació al·lèlica pel fragment del FAIM2.

## 4.2.6 RESULTATS DELS GENOTIPS DELS GENS ANALITZATS

Els resultats del genotip dels gens analitzats per seqüenciació Sanger i per *assay* es recullen a la *Taula 9*.

Pacient	CALN1 rs200910610	CALN1 rs201883338	CD207	EME1	EMP2 rs12930183	EMP2 rs12934615	FAIM2	HLA-DQB1- AS1	IL17RC	IL27RA	KIR2DL1 rs670771	KIR2DL1 rs75200129	TMEM175
Gy01	A/G	A/G	G/A	A/AAGC	G/A	A/G	A/A	G/G	G/G	T/T	-	-	C/T
Gy02	A/G	A/G	G/A	A/A	G/G	A/G	T/A	G/G	G/G	T/T	T/C	A/G	C/C
Gy03	A/G	A/G	G/G	A/AAGC	G/A	A/G	A/A	G/G	G/G	T/T	T/C	A/G	C/T
Gy04	A/G	A/G	A/A	A/AAGC	G/A	A/G	A/A	G/G	G/G	T/T	T/C	A/G	C/T
Gy05	A/G	A/G	G/G	A/A	G/G	A/G	T/A	G/G	G/G	T/T	T/T	A/A	C/C
Gy06	A/G	A/G	G/A	A/AAGC	G/G	A/G	T/A	G/G	G/G	T/T	C/C	G/G	C/T
Gy07	A/G	A/G	G/A	A/AAGC	G/A	A/G	A/A	T/T	G/G	T/T	-	-	C/T
Gy08	A/G	A/G	G/G	A/AAGC	G/A	A/G	A/A	G/G	G/G	T/T	T/C	A/G	C/T
Gy09	A/G	A/G	G/G	A/A	-	-	T/A	G/G	G/G	T/T	T/C	A/G	C/C
Gy10	A/G	A/G	G/A	AAGC/AAGC	G/A	A/A	A/A	G/G	G/G	T/T	T/C	A/G	C/C
Gy11	A/G	A/G	G/A	A/AAGC	-	-	A/A	T/T	G/G	T/T	T/C	A/A	C/T
Gy12	A/G	A/G	G/A	AAGC/AAGC	-	-	A/A	G/G	G/G	T/T	T/T	A/G	C/C
Gy13	A/G	A/G	G/G	A/AAGC	G/A	A/G	A/A	G/G	G/G	T/T	T/C	A/G	C/T
Gy14	A/G	A/G	G/G	A/A	G/A	A/G	A/A	G/G	G/G	T/T	T/C	A/G	C/C
Gy15	A/G	A/G	G/G	A/AAGC	-	-	A/A	G/G	G/G	T/T	T/C	A/G	C/T
Gy16	A/G	A/G	G/G	A/AAGC	G/G	A/G	A/A	G/G	G/G	T/T	T/C	A/G	-
Gy17	A/G	A/G	G/G	AAGC/AAGC	G/A	A/G	A/A	T/T	G/G	T/T	T/C	A/G	C/C
Gy18	A/G	A/G	G/G	A/A	-	-	A/A	G/G	G/G	T/T	T/T	A/A	C/C
Gy19	A/G	A/G	G/A	A/AAGC	G/A	A/G	T/A	G/G	G/G	T/T	T/C	A/G	T/T
Gy20	A/G	A/G	G/A	A/AAGC	G/A	A/G	T/A	G/G	G/G	T/T	T/T	A/A	T/T
Gy21	A/G	A/G	G/A	AAGC/AAGC	G/A	A/G	T/T	G/G	G/G	T/T	T/C	A/G	T/T
Gy22	A/G	A/G	G/G	AAGC/AAGC	G/G	A/G	A/A	T/T	G/G	T/T	T/C	A/G	C/T
Gy23	A/G	A/G	G/A	A/A	G/A	A/G	A/A	G/G	G/G	T/T	T/T	A/A	C/T
Gy24	A/G	A/G	G/G	A/AAGC	-	-	A/A	G/G	G/G	T/T	T/C	A/G	C/C
Gy25	A/G	A/G	G/A	A/AAGC	G/G	A/G	T/A	G/G	G/G	T/T	T/T	A/A	C/C
Gy28	A/G	A/G	G/G	AAGC/AAGC	G/G	A/G	T/A	G/G	G/G	T/T	T/C	A/G	C/C
Gy29	A/G	A/G	G/G	A/AAGC	G/G	A/G	T/A	T/T	G/G	T/T	T/C	A/G	C/T
Gy30	A/G	A/G	G/G	A/AAGC	G/A	A/G	T/A	T/T	G/G	T/T	T/C	A/G	C/C

*Taula 9:* Taula dels genotips dels gens validats de l'estudi d'exòmica.

De color verd s'ha marcat l'individu que es va determinar com a membre sa en l'estudi d'exòmica i de color vermell els afectats d'EM analitzats en l'exòmica. Els - són fragments que no s'han pogut genotipar.

#### 4.2.7 CONCORDANÇA ENTRE ELS GENOTIPS DE L'EXÒMICA I LA SEQÜENCIACIÓ

Un cop s'han analitzat els resultats de la validació de l'exòmica, s'ha observat que, en alguns casos, els genotips no es corresponen amb el genotip proporcionat per l'exòmica. El principal problema és quan el genotip del control sa (Gy05) proporcionat per l'exòmica (*Taula 5*) no correspon amb el genotip validat (*Taula 9*). Aquest fet s'ha observat en el gen CALN1, EMP2 (rs12934615), FAIM2, HLA-DQB1-AS1 i IL17RC.

En el cas dels dos SNPs del CALN1 i del EMP2 (rs12934615) el que ha succeït és que en el Gy05, en l'exòmica s'ha llegit algun al·lel variant però no els suficients com per determinar que aquest individu és heterozigot. És per això que en el resultat de l'exòmica el genotip del Gy05 l'anomena "probable homozigot per l'al·lel de referència". En el FAIM2 només s'han fet 6 lectures del Gy05 per la posició d'interès. Aquesta baixa cobertura ha donat un biaix en el genotipatge. Per tant, en la validació dels canvis d'aquests gens, s'ha vist que realment aquest individu és heterozigot i no homozigot per l'al·lel de referència.

En el cas del HLA-DQB1-AS1 i IL17RC no s'ha llegit cap còpia de la posició analitzada en el Gy05, de manera que s'ha atorgat automàticament un valor de 0 en la freqüència de l'al·lel variant. Tot i considerar-se homozigot per l'al·lel de referència, en la validació s'ha observat que és homozigot per l'al·lel variant en el cas del HLA-DQB1-AS1 i IL17RC. Aquest fet també s'ha observat en el EME1 però en aquest cas, per coincidència, els genotips obtinguts en l'exòmica dels individus estudiats són els mateixos que en la validació.

Un altre error de l'exòmica és quan la *ratio* entre el número de cops que s'ha llegit l'al·lel variant i la profunditat es troba en un rang dubtós entre dos genotips. Aquest genotip dubtós es determina com a *Uncertain heterozygous*, si el rang es troba entre l'homozigot per l'al·lel de referència o heterozigot, o *Uncertain homozygous for the variant allele*, si el rang es troba entre l'heterozigot o homozigot per l'al·lel variant. En el cas del IL27RA, s'han detectat diferències entre tots els individus afectats estudiats en l'exòmica determinats com *Uncertain heterozygous* i el membre sa. Malgrat que pel Gy05 els genotips de l'exòmica i la validació han coincidit, en la validació s'ha observat que els individus afectats són realment homozigots per l'al·lel de referència.

Un altre fet a comentar és quan només es llegeix una còpia d'una seqüència determinada. En aquests casos l'exòmica determina aquest individu com homozigot, però realment no és possible saber si és heterozigot o homozigot. A més, també hi ha casos amb una cobertura molt baixa i per tant hi ha menys probabilitat de que aquell genotip sigui correcte.

No obstant això, en el cas del EMP2 (rs12930183), CD207, TMEM175 i KIR2DL1 els genotips de l'exòmica i la validació han coincidit per tots els individus estudiats en l'estudi d'exòmica.

### 4.3 HLA

Per determinar quins al·lels del HLA estan presents en els membres de la família, s'ha utilitzat el sistema Luminex. Tal i com s'observa en la *Taula 10*, els al·lels DRB1\*15:01 i DQB1\*06:02, que més s'han associat a l'EM, no s'han trobat en els membres de la família, exceptuant el Gy22 que presenta un al·lel DQB1\*06. Per tant es descarta que aquests al·lels tinguin implicació genètica en el desenvolupament de l'EM en aquesta família. D'altra banda, s'observa que l'al·lel HLA-DRB1\*03 és bastant freqüent en aquesta família. Prèviament s'havia associat el genotip HLA-DRB1\*03 amb susceptibilitat a EM en població de Sardenya (Marrosu et al., 2001).

*Taula 10:* Genotips del HLA-DRB1 i HLA-DQB1.

<b>Pacient</b>	<b>HLA-DRB1</b>	<b>HLA-DQB1</b>
Gy01	03/04	02/03
Gy02	03/03	02/02
Gy03	03/04	02/03
Gy04	03/04	02/03
Gy05	03/07	02/03
Gy06	03/14	02/05
Gy07	07/14	03/05
Gy08	03/07	02/03
Gy09	07/08	02/04
Gy10	04/08	03/04
Gy11	04/14	03/05
Gy12	03/08	02/04
Gy13	03/03	02/02
Gy14	03/07	02/02
Gy15	03/07	02/02
Gy16	03/11	02/03
Gy17	08/14	04/05
Gy18	03/07	02/03
Gy19	03/14	02/05
Gy20	03/04	02/03
Gy21	03/04	02/03
Gy22	03/13	02/06
Gy23	04/04	03/03
Gy24	03/07	02/03
Gy25	03/07	02/03
Gy28	11/11	03/03
Gy29	08/14	04/05
Gy30	14/16	05/05

#### 4.4 ESTADÍSTICA DELS GENS ANALITZATS

Per comprovar si hi ha diferències significatives d'un genotip entre els individus afectats d'EM i els sans, s'ha realitzat una prova de  $\chi^2$  mitjançant el programa SPSS. Només s'ha fet l'anàlisi estadístic per aquells gens els quals s'ha observat una distribució heterogènia en la família per a un genotip dels fragments estudiats. Així doncs, en el cas del CALN1, IL17RC i IL27RA tots els individus de la família tenen el mateix genotip (Taula 9), i per tant no es pot aplicar cap anàlisi estadístic.

Es pren com a exemple l'anàlisi estadístic de FAIM2 i CD207. Els individus RIS són difícils de classificar, ja que, malgrat no considerar-se EMCD, podrien tractar-se d'individus afectats que encara es troben en estadis inicials de la malaltia. Per això, en l'anàlisi estadístic s'ha realitzat considerant els RIS com afectats (Taula 11, 12 pel FAIM2 i 15 pel CD207) i com a no afectats (Taula 13, 14 pel FAIM2 i 16 pel CD207). A més, l'anàlisi s'ha dut a terme considerant els 3 genotips possibles (homozigot per l'al·lel de referència, heterozigot i homozigot per l'al·lel variant) (Taula 11 i 13 pel FAIM2) i considerant la presència o absència de l'al·lel variant (Taula 12 i 14 pel FAIM2). En el cas del CD207 només s'han analitzats els dos genotips presents en la família. Cal comentar que en cap dels casos analitzats s'han trobat diferències significatives ( $p$ -valor<0,05) en la proporció d'un genotip entre els afectats i no afectats. Tot i que en algun cas s'ha observat que el  $p$ -valor està molt pròxim al llindar, pel fet que el genotip homozigot per l'al·lel de referència s'ha trobat només en els individus sans. A més, en aquests casos, s'ha observat que pel genotip heterozigot i homozigot per l'al·lel variant hi ha una distribució homogènia entre els afectats i no afectats d'EM.

Taula 11: Taula de contingència del FAIM2 considerant RIS afectat i els tres genotips possibles.

Genotip	Sa	Afectat + RIS
Homozigot al·lel referència	1 (5,9%)	0 (0%)
Heterozigot	6 (35,3%)	5 (45,5%)
Homozigot al·lel variant	10 (58,8%)	6 (54,5%)
<i>p</i> -valor* *test exacte de Fisher	0,824	

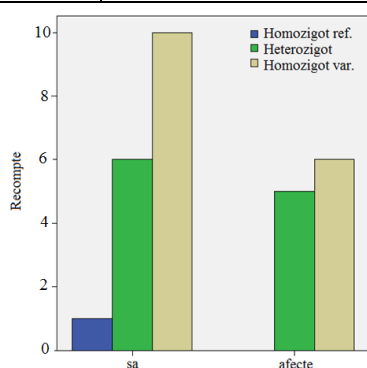


Figura 16: Representació gràfica de la Taula 11.

Taula 12: Taula de contingència del FAIM2 considerant RIS afectat i absència/presència de l'al·lel variant.

Genotip	sa	Afectat + RIS
Absència al·lel variant	1 (5,9%)	0 (0%)
Presència al·lel variant	16 (94,1%)	11 (100%)
<i>p</i> -valor* *test exacte de Fisher	1,000	

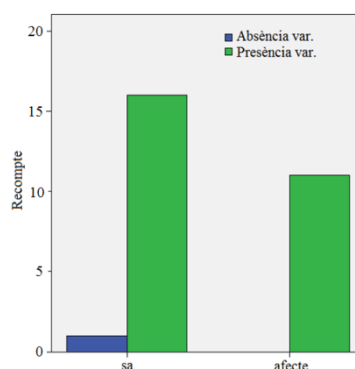


Figura 17: Representació gràfica de la Taula 12.



Taula 13: Taula de contingència del FAIM2 considerant RIS sa i els tres genotips possibles.

Genotip	Sa + RIS	Afectat
<b>Homozigot al·lel referència</b>	1 (4,8%)	0 (0%)
<b>Heterozigot</b>	8 (38,1%)	3 (42,9%)
<b>Homozigot al·lel variant</b>	12 (57,1%)	4 (57,1%)
<b><i>p</i>-valor*</b> <b>*test exacte de Fisher</b>	1,000	

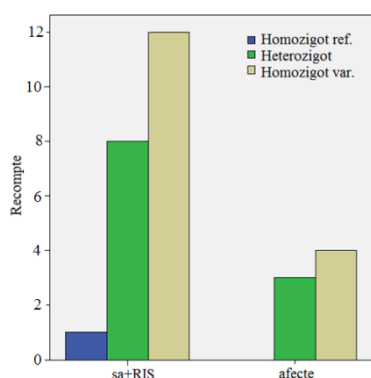


Figura 18: Representació gràfica de la Taula 13.

Taula 14: Taula de contingència del FAIM2 considerant RIS sa i absència/presència de l'al·lel variant.

Genotip	Sa + RIS	Afectat
<b>Absència al·lel variant</b>	1 (4,8%)	0 (0%)
<b>Presència al·lel variant</b>	20 (95,2%)	7 (100%)
<b><i>p</i>-valor*</b> <b>*test exacte de Fisher</b>	1,000	

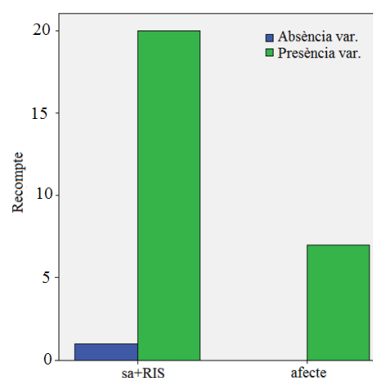


Figura 19: Representació gràfica de la Taula 14.

Taula 15: Taula de contingència del CD207 considerant RIS afectat i dos genotips possibles.

Genotip	sa	Afectat + RIS
<b>Homozigot al·lel referència</b>	10 (58,8%)	5 (50%)
<b>Heterozigot</b>	7 (41,2%)	5 (50%)
<b><i>p</i>-valor*</b> <b>*test <math>\chi^2</math> de Pearson</b>	0,656	

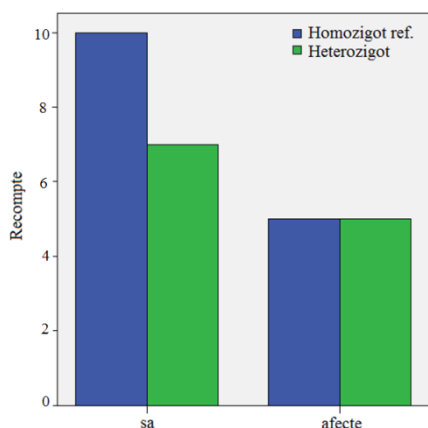


Figura 20: Representació gràfica de la Taula 15.

Taula 16: Taula de contingència del CD207 considerant RIS sa i dos genotips possibles.

Genotip	Sa + RIS	Afectat
<b>Homozigot al·lel referència</b>	13 (65%)	2 (28,6%)
<b>Heterozigot</b>	7 (35%)	5 (71,4%)
<b><i>p</i>-valor*</b> <b>*test exacte de Fisher</b>	0,185	

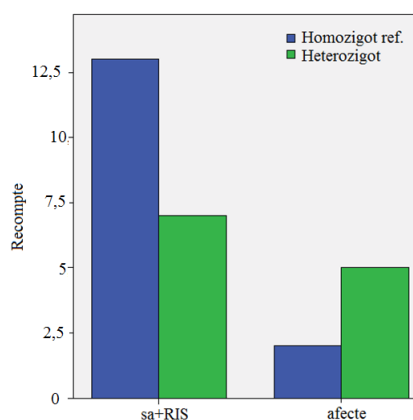


Figura 21: Representació gràfica de la Taula 16.

## 5. CONCLUSIONS

---

- The riskiest genotypes of MS (DRB1\*15:01 and DQB1\*06:02) are not present in the studied family. On the other hand, the HLA-DRB1\*03 allele is common in this family.
- Many of the validated changes are located in zones with a large number of polymorphisms, which has hampered the primer design and validation.
- It has observed a low concordance between genotyping provided by exome study and the genotyping obtained in the validation. Most errors are justified by the inadequate coverage in the exome study. It's important to consider the coverage in the selection of candidate genes.
- In the genotype-phenotype correlation of the studied family, no significant differences were detected between the affected and unaffected by MS for the studied genes.
- The genes: CALN1, CD207, EME1, EMP2, FAIM2, HLA-DQB1-AS1, IL17RC, IL27RA, KIR2DL1 and TMEM175, could not explain the presence of MS in the family.

## 6. BIBLIOGRAFIA

---

- American Neurological Association. (2011). Decoding Multiple Sclerosis. *Annals of Neurology*, 70(6), A5–A7.
- Barrie, J. H. (2009). The diagnosis of multiple sclerosis and the clinical subtypes. *Ann Indian Acad Neurol*, 12(4), 226–230.
- Dyment, D., Ebers, G., Cader, M., et al. (2012). Exome sequencing identifies a novel multiple sclerosis susceptibility variant in the TYK2 gene. *Neurology*, 79(5), 406–11.
- Ebers, G. (2008). Environmental factors and multiple sclerosis. *Lancet Neurol*, 7(3), 268–77.
- Ebers, G., Dyment, D., Cader, M., et al. (2008). A genome scan in a single pedigree with high prevalence of multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 79(2), 158–62.
- Ebers, G & Sadovnick, A. (1994). The role of genetic factors in multiple sclerosis susceptibility. *J Neuroimmunol.*, 54(1-2), 1–17.
- Fernández, O., Fernández, V., Alonso, A., et al. (2004). DQB1\*06:02 allele shows a strong association with multiple sclerosis in patients in Malaga, Spain. *Journal of Neurology*, 251(4), 440–4.
- Fernández, O., Fernández, V., Martínez, V., et al. (2008). Multiple sclerosis in Gypsies from southern Spain: prevalence, mitochondrial DNA haplogroups and HLA class II association. *Tissue Antigens*, 71(5), 426–33.
- Giovannoni, G & Ebers, G. (2007). Multiple sclerosis: the environment and causation. *Current Opinion in Neurology*, 20(3), 261–8.
- Greer, J & McCombe, P. (2011). Role of gender in multiple sclerosis: clinical effects and potential molecular mechanisms. *Journal of Neurology*, 234, 7–18.
- Haines, J., Ter-Minassian, M., Bazyk, A., Gusella, J., et al. (1996). A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. *Nature Genetics*, 13(4), 469–471.
- Harbo, H., Gold, R & Tintoré, M. (2013). Sex and gender issues in multiple sclerosis. *Therapeutic Advances in Neurological Disorders*, 6(4), 237–48.
- Hoffjan, S & Akkad, D. (2010). The genetics of multiple sclerosis: an update 2010. *Mol Cell Probes*, 24(5), 237–43.
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(February), 860–921.
- Kingwell, E., Marriott, J., Jetté, N., et al. (2013). Incidence and prevalence of multiple sclerosis in Europe: a systematic review. *BMC Neurology*, 13, 128.
- Marrosu, M., Murru, R., Murru, M., et al. (2001). Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Human Molecular Genetics*, 10(25), 2907–2916.
- Oksenberg, J & Baranzini, S. (2010). Multiple sclerosis genetics - Is the glass half full, or half empty? *Nature Reviews. Neurology*, 6(8), 429–37.
- Oksenberg, J., Baranzini, S., Sawcer, S & Hauser, S. (2008). The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nature Reviews. Genetics*, 9(7), 516–26.
- Rabbani, B., Tekin, M & Mahdih, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, 59(1), 5–15.
- Ramagopalan, S., Morris, A., Dyment, D., et al. (2007). The inheritance of resistance alleles in multiple sclerosis. *PLoS Genetics*, 3(9), 1607–13.
- Sadovnick, A., Dyment, D., Ebers, G & Risch, N. (1996). Evidence for genetic basis of multiple sclerosis. *The Lancet*, 347(9017), 1728–1730.
- Traherne, J. (2008). Human MHC architecture and evolution: implications for disease association studies. *Int J Immunogenet*, 35(3), 179–192.
- Zuvich, R., McCauley, J., Pericak-Vance, M & Haines, J. (2009). Genetics and Pathogenesis of Multiple Sclerosis. *Seminars in Immunology*, 21(6), 328–333.