

**ANÀLISI DE LA VARIACIÓ GENÈTICA
DE LES REGIONS CFTR I GBA
EN POBLACIONS HUMANES DE TOT EL MÓN**

EVA MATEU i MORANTE

2001

Dipòsit legal: 24691-2002
ISBN: 84-699-8735-6

Departament de Ciències Experimentals i de la Salut

UNIVERSITAT POMPEU FABRA



**ANÀLISI DE LA VARIACIÓ GENÈTICA
DE LES REGIONS CFTR I GBA
EN POBLACIONS HUMANES DE TOT EL MÓN**

Memòria presentada per Eva Mateu i Morante per a optar al grau de Doctor en
Biologia per la Universitat Pompeu Fabra.

Aquest treball s'ha realitzat sota la direcció de Jaume Bertranpetit i Busquets, a la Unitat d'Antropologia del Departament de Biologia Animal de la Facultat de Biologia, Universitat de Barcelona, i a la Unitat de Biologia Evolutiva del Departament de Ciències Experimentals i de la Salut, de la Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra.

Programa de doctorat: Biologia Animal II: Antropologia Biològica,
Universitat de Barcelona (bienni 1995-1997).

Jaume Bertranpetit i Busquets
Director de la tesi

Eva Mateu i Morante

Als meus pares
A en Carles

Is a dream a lie if it don't come true?

(Bruce Springsteen)

AGRAÏMENTS

Són moltes les persones que directa o indirectament han col·laborat en la realització d'aquesta tesi. Des d'aquí els hi vull agrair la seva ajuda a totes elles:

A en Jaume, director d'aquesta tesi, per haver confiat en mi des del primer dia i haver-me introduït, amb el seu entusiasme constant, en el món de la recerca.

A en Francesc per la seva AJUDA durant tot aquest temps, per la seva paciència inesgotable a contestar una pregunta darrera l'altra, per totes les anàlisis estadístiques fetes i totes les lectures i correccions dels escrits.

A en David per haver sigut el meu mestre de laboratori i haver-me ensenyat la teoria i la pràctica de les tècniques que s'hi utilitzen (encara que per sort a mig camí van sorgir els seqüenciadors automàtics!). I perquè sempre he pogut comptar amb ell!

A l'Anna, la meva altra mestra de laboratori, per totes les estones que hem rigut juntes, totes les anades de compres, els viatges i mil coses més que hem compartit. I també per haver ideat una manera de tipar SNPs amb el seqüenciador automàtic!.

A tots els antics companys d'Antropologia però molt especialment a la Natalie, la Neus (malgrat que sempre faci més piscines que jo), l'Ana-de-Madrit, la Blanca (companya d'aventures i desventures becàries), l'Araceli (la meva tennista preferida), la Bàrbara i també la Lourdes. I als de Genètica: al Marc (amb h!), a la Rebe i al Lorens per totes les hores que hem passat junts des d'aquella festa Toblerone.

A tots els nous companys de Biologia Evolutiva però molt especialment a la Mònica (sempre disposada a donar un cop de mà), l'Stéphanie (la millor veïna que es pot desitjar en un bon barri) i l'Oscar (amb els seus meravellosos mapes). I també als nous companys del CEXS (immunos, genètics...) amb qui hem compartit més que uns cafès.

Als antics, però també nous, companys: la Cristina (i la seva germana Susana!), l'Elena (que va obrir el camí a les tesis de ciències a la Pompeu), la Rosa (i els megahaplotips que hem pogut fer conjuntament) i, molt especialment, al Jordi (company insubstituïble de pràctiques) i a l'Aida (que més d'un dia es va barallar amb el GBA).

I a tots ells per totes les bones estones passades al llarg d'aquests anys tant al laboratori com a les múltiples sortides i festetes que hem fet.

A tot l'equip de la Unitat de Seqüenciació dels SCT de la UB.

Als companys de la biblioteca de l'Àrea del Mar de la UPF per facilitar-me l'accés a tants articles.

A en David Moreso per saber solucionar sempre els problemes informàtics.

A tot el personal administratiu i de suport de la UPF i molt especialment a la Lydia Garcia i la Carme Abelló per la seva inestimable ajuda amb la burocràcia.

Als laboratoris que m'han acollit al llarg d'aquest temps: a tot l'equip de l'Ernest Beutler, de l'Scripps Research Institute a La Jolla, i molt especialment a la Carol West; a en Xavier Estivill i a tot el grup de Fibrosi Quística de l'IRO: la Teresa Casals, la Loli Ramos, el Javi Gimenez i la Sara Larriba.

A la Laura i la Neus (que ja ho han passat) i a la Mònica.

Al George, la Banana, el Pat i la Beni i els seus plans *destroyer*.

A la Xus i la Marian que també ho han patit.

A la Vicky que sempre s'hi ha interessat.

A la família Mateu Morante i a la família Ryhr Miralles.

A en Carles, la columna del meu temple, per *tolerar-me* tant.

Als meus pares per ser-hi sempre.

Finalment, un agraïment molt especial a totes aquelles persones anònimes que es van deixar treure una mica de sang fent possible la realització d'estudis d'aquesta mena.

Durant la realització d'aquest treball he comptat amb l'ajuda d'una beca de l'Institut Universitari de Salut Pública de Catalunya (projecte ISP-10/A), una beca de formació d'investigadors de la Universitat de Barcelona i d'un contracte d'ajudant d'escola universitària de la Universitat Pompeu Fabra. A més a més el treball ha comptat amb l'ajuda dels següents organismes i projectes: Fundació La Marató de TV3, projecte "La història natural de la fibrosi quística: interpretació geogràfica de la variació genètica"; Dirección General de Enseñanza Superior e Investigación Científica, Ministerio de Educación y Cultura, projectes PB95-0267-C02-01 i PB98-1064; Direcció General de Recerca, Generalitat de Catalunya, Grup de Recerca Consolidat 1996SGR00041 i 1998SGR00009.

ÍNDEX

PRESENTACIÓ	13
INTRODUCCIÓ	17
1. La diversitat humana	21
1.1. Polimorfismes utilitzats en l'estudi de la diversitat humana	22
1.1.1. Haplotips	24
1.2. Equilibri i desequilibri de lligament	25
1.2.1. Factors que afecten el desequilibri de lligament	26
1.2.2. Mesura del desequilibri de lligament	27
1.2.3. Desequilibri de lligament, distància genètica i població	28
1.2.4. Desequilibri de lligament i edat de la població	30
1.2.5. Límit del desequilibri de lligament	31
2. La regió genòmica 7q31: el gen CFTR	32
2.1. Clínica de la fibrosi quística	33
2.1.1. Relació genotip-fenotip	35
2.2. Família de gens transportadors ABC	37
2.2.1. Estructura de CFTR	37
2.3. Funció de CFTR	41
2.4. Mutacions al gen CFTR	42
2.4.1. Tipus moleculars de mutació	44
2.4.2. La mutació $\Delta F508$ i la seva distribució geogràfica	46
2.4.3. Les mutacions G542X, G551D, N1303K i W1282X i la seva distribució geogràfica	47
2.4.4. Altres mutacions menys freqüents	59
2.5. Per què la CF és una malaltia tan freqüent? Avantatge dels heterozigots	60
2.6. Origen de les mutacions més freqüents	61
2.6.1. Edat de les principals mutacions CF	62
2.6.2. Haplotips per cromosomes mutants i per cromosomes normals	62
2.6.3. Població d'origen de les mutacions $\Delta F508$, G542X i N1303K	63

3. La regió genòmica 1q21: els gens GBA i PKLR	65
3.1. Altres gens a la regió genòmica 1q21	66
3.2. La regió GBA	66
3.2.1. Clínica de la malaltia de Gaucher	67
3.2.1.1. Relació genotip-fenotip	70
3.2.2. Estructura del gen GBA	70
3.2.3. Estructura del pseudogèn GBA	71
3.2.4. Funció de GBA	71
3.2.5. La saposina C	72
3.2.6. Mutacions al gen GBA	72
3.2.6.1. La mutació N370S o 1226G	73
3.2.6.2. La mutació L444P o 1448C	73
3.2.6.3. La mutació 84GG	74
3.2.6.4. La mutació IVS2+1	74
3.2.6.5. Altres mutacions puntuals	74
3.2.6.6. Al·lels complexos o recombinants	74
3.2.6.7. Distribució poblacional de les mutacions més freqüents	75
3.2.7. Malaltia de Gaucher i avantatge dels heterozigots en la població jueva	76
3.2.8. Polimorfismes i haplotips al gen GBA	77
3.2.9. Origen de les mutacions més freqüents	78
3.3. El gen PKLR	79
3.3.1. Clínica de la deficiència de piruvat quinasa	79
3.3.2. Estructura de PKLR	80
3.3.3. Funció de PKLR	81
3.3.4. Mutacions al gen PKLR i distribució geogràfica	81
3.3.5. Polimorfismes, haplotips i mutacions associades al gen PKLR	82
 OBJECTIUS	 85

MATERIALS I MÈTODES	89
1. Poblacions estudiades	91
1.1. Origen i obtenció del DNA de les diferents poblacions	92
2. Anàlisi de microsatèl·lits i substitucions nucleotídiques	93
3. Anàlisi estadística de les dades	94
RESULTATS	95
Capítol I: <i>Allele frequencies in a worldwide survey of a CA repeat in the first intron of the CFTR gene</i>	97
Capítol II: <i>Worldwide genetic analysis of the CFTR region</i>	105
Capítol II. Apèndix	123
Capítol III: <i>Can a place of origin of the main CF mutations be recognized?</i>	133
Capítol IV: <i>The PKLR-GBA region shows almost complete linkage disequilibrium over 70 kb in a set of worldwide populations</i>	155
DISCUSSIÓ	197
BIBLIOGRAFIA	217
ADRECES ELECTRÒNIQUES D'INTERÈS	241
APÈNDIX I	245
<i>Selection shaping variability on a human pseudogene</i>	249
APÈNDIX II	289
<i>A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea</i>	293

NOTA: Les Taules i Figures dels Capítols III i IV dels Resultats i les de l'Apèndix I segueixen una numeració independent. Els Capítols I i II de Resultats i l'Apèndix II també segueixen una numeració independent a la de la resta del treball.

PRESENTACIÓ

El treball que presentem a continuació és una contribució als estudis de diversitat del genoma humà i pretén estudiar la variació genètica existent a nivell mundial en dues regions del genoma on es troben ubicats dos gens, causants de malaltia, ben coneguts, el gen CFTR i el gen GBA, en cromosomes d'individus no afectats. Mutacions en aquest dos gens produeixen la fibrosi quística en el primer cas i la malaltia de Gaucher en el segon. La fibrosi quística és la malaltia autosòmica recessiva més comuna en poblacions d'origen europeu. La malaltia de Gaucher és la malaltia lisosòmica d'acumulació lipídica més freqüent. L'estudi analitza la variació genètica en diferents polimorfismes d'ambdós gens (tant microsatèl·lits com substitucions nucleotídiques) per tal de calcular les freqüències al·lèliques i la seva diferent distribució geogràfica; reconstrueix els haplotips i analitza també la seva distribució geogràfica; i analitza l'extensió i distribució geogràfica del desequilibri de lligament entre *loci*. Pel cas de l'estudi del gen GBA, hem ampliat la regió cap a 5' del gen, abastant fins al gen PKLR (gen que codifica per a la piruvat quinasa). A més a més, pel cas de CFTR, l'estudi pot ajudar a entendre l'origen de les mutacions més freqüents causants de fibrosi quística.

INTRODUCCIÓ

En la introducció que ve a continuació descriurem els factors implicats en la generació de la diversitat humana i les eines que tenim per a estudiar aquesta diversitat i seguidament passarem a una descripció detallada de cadascuna de les dues regions genòmiques que hem analitzat en aquest treball: la regió 7q31 i la regió 1q21.

1. LA DIVERSITAT HUMANA

La diversitat del genoma humà està estretament lligada a la seva recent evolució. La distribució de la variació en la seqüència del DNA (*DeoxyriboNucleic Acid*), a dins i entre les diferents poblacions humanes, és el resultat de diversos factors evolutius moleculars o genòmics com són la mutació, la recombinació i la conversió gènica; i de factors poblacionals com la deriva, la selecció i la migració que han operat al llarg de la història de la nostra espècie.

La **mutació** és un error en la còpia d'una seqüència de DNA en cèl·lules de la línia germinal i que és transmesa a la descendència. Es poden donar substitucions nucleotídiques, insercions i delecions d'un o més nucleòtids, i inversions d'un segment de DNA. La probabilitat que un sol nucleòtid muti varia molt depenent de la zona del genoma i del seu efecte fenotípic, però és en general un procés força infreqüent, de l'ordre de $\sim 10^{-9}$ substitucions nucleotídiques per nucleòtid i any.

La **recombinació** és l'intercanvi recíproc de material genètic que es produeix en la meiosi per l'entrecreuament entre els diferents cromosomes homòlegs.

La **conversió gènica**, en canvi, suposa l'existència d'un al·lel donador d'un fragment de seqüència, que no es modifica, i un al·lel acceptor, que sí resulta modificat. En aquest cas l'intercanvi de material genètic no és equitatiu i una de les variants es perd en la transmissió a la descendència.

La **deriva** és un mecanisme basat en l'atzar que té efectes importants si la població és petita i està aïllada. En aquest tipus de població la transmissió dels al·lells, d'una generació a l'altra, pot variar molt simplement per pur atzar, fent variar les seves freqüències al·lèliques. En poblacions grans el seu efecte no és tan evident, malgrat que també hi pot actuar.

La **selecció natural**, condicionada per la pressió ambiental, és el mecanisme que afavoreix unes variants al·lèliques o un genotip en front d'altres variants que no seran afavorides i per tant la seva probabilitat de passar a la descendència serà menor. La variant afavorida és la que augmentarà de freqüència en la població fins arribar a fixar-se. Aquesta selecció positiva s'ha de diferenciar del que s'anomena selecció negativa o purificadora, que elimina els al·lells deleteris de la població (al·lells letals o causants de malaltia). Hi ha un elevat nombre d'al·lells, però, que no tenen cap avantatge ni desavantatge selectiu, constituint el que s'anomena variació neutra. Cal ressaltar que la selecció actua sobre el fenotip i, indirectament, sobre els gens.

La **migració** és l'intercanvi d'individus entre diferents poblacions. El seu efecte serà la disminució de la diferència genètica entre les poblacions que es barregin. Aquest efecte dependrà, però, de les diferències genètiques inicials entre les poblacions i del nombre d'individus que migrin.

Els patrons de variació observats tant en els cromosomes autosòmics i sexuals (DNA nuclear), com en el DNA mitocondrial són resultat i alhora reflex d'aquests factors. Hem de tenir en compte, però, que la variabilitat observada a les poblacions humanes actuals és força petita, segurament degut al fet de ser una espècie molt recent.

1.1. Polimorfismes utilitzats en l'estudi de la diversitat humana

Hom diu que un locus és polimòrfic si dos o més al·lells coexisteixen en la població i les freqüències d'almenys dos d'ells són superiors a l'1%. Trobem diferents tipus de polimorfismes al DNA, com són els microsatèl·lits o STRPs (*Short Tandem Repeat Polymorphisms*), minisatèl·lits o VNTRs (*Variable Number of Tandem Repeats*), substitucions nucleotídiques o SNPs (*Single Nucleotide*

Polymorphisms) i polimorfismes d'inserció/deleció d'un o més nucleòtids (com per exemple insercions polimòrfiques de seqüències *Alu*). Aquest treball es centra especialment en dos tipus: els STRPs i els SNPs.

Els **microsatèl·lits o STRPs** són repeticions en tàndem d'una unitat bàsica de seqüència de DNA, d'entre dos i sis nucleòtids. En aquest sentit s'anomenen respectivament di-, tri-, tetra-, penta- i hexanucleòtids (a partir de 7 nucleòtids passen a anomenar-se VNTRs). Són polimorfismes molt abundants i àmpliament estesos per tot el genoma humà. Hom estima que hi ha un STRP cada 30-60 kilobases (kb) de DNA. El polimorfisme consisteix en la variació del nombre d'unitats repetides. En mutar, el microsatèl·lit guanya o perd una o més unitats bàsiques, essent els guanys o pèrdues d'una sola repetició els més freqüents (Brinkmann et al. 1998). Aquest fenomen genera l'alt grau de polimorfisme observat en molts microsatèl·lits. No obstant, és possible que aquesta apreciació sigui deguda al procés de descobriment de STRPs en el genoma, que consisteix sovint en l'ús de sondes amb $(CA)_{10}$, les quals detectaran els microsatèl·lits amb un nombre més gran de repeticions. I aquests són precisament els que presenten una taxa de mutació més alta (Ellegren, 2000). Per tant, és possible que l'espectre de variabilitat dels microsatèl·lits hagi estat subestimat. Degut a la seva elevada taxa de mutació (entre 10^{-5} - 10^{-2} / generació, depenent de l'autor; Weber i Wong 1993; Chakraborty et al. 1997; Brinkmann et al. 1998; sabem que la taxa de mutació depèn de molts factors genètics com nombre inicial de repeticions, tipus i llargada de la repetició, puresa de les repeticions...) els microsatèl·lits generen elevada quantitat d'al·lels, fet que facilita els estudis evolutius (Bowcock et al. 1994a; Pérez-Lezaun et al. 1997, entre d'altres) o els estudis de determinades regions genòmiques (Kidd et al. 1998; Tishkoff et al. 2000b, entre d'altres). Un factor, però, que pot emascarar les diferències evolutives en els microsatèl·lits és l'homoplàsia, que es dona quan els diferents al·lels d'un locus son idèntics per estat però no per descendència, degut al propi procés de mutació i producció de nous al·lels.

Les **substitucions nucleotídiques o SNPs** són substitucions d'un sol nucleòtid en la seqüència del DNA. La seva taxa de mutació, equivalent a la taxa de mutació del genoma en general ($\sim 10^{-9}$) (Li et al. 1996) és menor que la dels microsatèl·lits, essent més estables evolutivament. Molt sovint es poden considerar fruit d'esdeveniments únics en l'evolució, i per tant els al·lels idèntics ho són per descendència. Hi ha heterogeneïtat, donada tant pel tipus de substitució com pel context, essent el cas més conegut el dels duets CpG.

STRPs i SNPs, a més a més de la seva utilitat pels estudis de la diversitat humana en el camp evolutiu, són marcadors molt utilitzats en genètica clínica per establir un diagnòstic a partir d'estudis de segregació en famílies i per fer consell genètic (Casals et al. 1996).

1.1.1. Haplotips

Un haplotip és la combinació d'al·lels de diversos polimorfismes presents en un determinat segment de cromosoma. Si aquesta regió és prou petita, la probabilitat que la recombinació hi actuï és baixa i, per tant, l'haplotip es transmetrà íntegre a la descendència. Els haplotips es poden definir a partir de qualsevol combinació de tipus de polimorfismes. Els haplotips, per si mateixos, són alhora un altre tipus de marcador genètic, molt útils, per exemple, per a la reconstrucció de processos migratoris humans o per establir efectes fundadors en determinades mutacions que produeixen malalties genètiques. La combinació de STRPs i SNPs ens permet definir un nucli evolutiu estable d'SNPs (amb una de mutació baixa) on s'ha anat acumulant la variació dels STRPs (amb alta taxa de mutació). Els haplotips més conservats es donen en regions amb baixes taxes de recombinació.

En el genoma diploide, la determinació independent dels genotips per als diferents polimorfismes no permet d'establir l'haplotip, ja que no proporciona informació sobre la fase, és a dir, quin al·lel es troba a cadascun dels dos

cromosomes homòlegs. Cal recórrer al genotipat de nuclis familiars (preferentment els pares de l'individu) per tal d'escatir quina combinació d'al·lels ha heretat de cada progenitor. Altrament, si no es disposa de nuclis familiars (com sol ser el cas en la majoria d'estudis de genètica de poblacions) cal recórrer a l'estimació estadística de les freqüències haplotípiques a la mostra. Els models que es plantegen per a aquesta inferència no són resolubles algebraicament, pel que cal recórrer a l'algoritme EM (*Expectation-Maximization*) (Dempster et al. 1977; Excoffier i Slatkin 1995; Slatkin i Excoffier 1996) per tal de trobar-ne una aproximació numèrica. Aquest algoritme, que es troba implementat en diferents programes, assumint equilibri Hardy-Weinberg a tot els *loci* implicats, dóna les estimes de màxima versemblança de les freqüències haplotípiques. Hom ha vist que no hi ha diferències estadístiques significatives en l'estimació de freqüències haplotípiques tant si la fase és coneguda com si no ho és (Tishkoff et al. 2000a), en casos en que el nombre de *loci* considerats sigui petit.

L'estimació de les freqüències haplotípiques ha demostrat ser una valuosa eina en els estudis de genètica de poblacions en els últims anys (Tishkoff et al. 1996, 1998; Kidd et al. 1998, entre d'altres).

1.2. Equilibri i desequilibri de lligament

Hom anomena **equilibri de lligament** la situació en què tots els possibles haplotips o possibles combinacions de diferents al·lels tenen lloc en freqüències que depenen només de l'associació a l'atzar de les freqüències al·lèliques individuals. Per exemple, per a un sistema de dos *loci* o dos gens (A i B) dial·lèlics ($A_1, A_2; B_1, B_2$) amb unes freqüències: $p_1 = 0.3, p_2 = 0.7$ (tal que $p_1 + p_2 = 1$); $q_1 = 0.8, q_2 = 0.2$ (on també $q_1 + q_2 = 1$), tenim quatre possibles haplotips (A_1B_1, A_1B_2, A_2B_1 i A_2B_2); en general, el nombre màxim d'haplotips és el producte del nombre d'al·lels de cada locus), que sota equilibri de lligament es trobaran respectivament a unes freqüències esperades: $p_1 \times q_1, p_1 \times q_2, p_2 \times q_1, p_2 \times q_2$ (és a dir: 0.24, 0.06, 0.56 i 0.14).

El **desequilibri de lligament** (*Linkage Disequilibrium*, LD) en canvi, és l'associació, no produïda per l'atzar, d'al·lels de diferents *loci* en un cromosoma concret. En aquest sentit, també és LD l'associació preferent d'una mutació o al·lel causant de malaltia amb al·lels específics de loci propers; quan una mutació apareix de nou en una població (per nova mutació o per introducció a través de la immigració en la població d'un portador d'aquell al·lel mutat) necessàriament va acompanyada d'un *background* genètic (un haplotip determinat de marcadors propers). Per recombinacions produïdes en el pas d'una generació a la següent, el desequilibri de lligament anirà disminuint gradualment. En el desequilibri de lligament, les freqüències haplotípiques observades s'allunyen de les esperades sota l'associació a l'atzar de les freqüències al·lèliques individuals. En l'exemple anterior dels *loci* A i B, sota LD, les freqüències haplotípiques observades podrien ser: 0.1, 0.2, 0.7 i 0 que clarament difereixen de les anteriors i si la mostra fos suficientment gran les diferències serien significatives.

1.2.1. Factors que afecten el desequilibri de lligament

Diversos factors poblacionals i genòmics poden alterar el desequilibri de lligament; alguns d'aquests factors tendiran a fer-lo minvar, mentre altres el faran augmentar.

El desequilibri de lligament entre *loci*, com ja hem comentat, va desapareixent gradualment al llarg del temps (en generacions) depenent de la freqüència de **recombinació** entre els *loci*. Així, el desequilibri disminuirà a mida que passin generacions tant més de pressa com més alta sigui la taxa de recombinació (Jorde 1995). Un altre factor que pot erosionar desequilibri és la **mutació**, entesa com a modificadora de les freqüències al·lèliques de diferents marcadors lligats (Calafell et al. 2001).

Factors que podem fer augmentar el desequilibri de lligament són, per exemple, la deriva, la migració o la selecció. En poblacions petites, la **deriva** pot

canviar aleatòriament la freqüència d'alguns haplotips, més enllà de les seves freqüències d'equilibri. Aquest fenomen és especialment rellevant quan un nombre reduït d'individus funden una població. En aquest cas, la població es pot fundar a partir d'un nombre petit de cromosomes que continguin una representació dels haplotips allunyada de l'equilibri simplement per un efecte aleatori sobre una mostra petita. La **migració** implica una barreja de poblacions i en aquest cas també es genera LD ja que les noves freqüències al·lèliques donen unes freqüències haplotípiques d'equilibri que no es corresponen amb les freqüències haplotípiques en la població mesclada (per exemple, si tenim dues poblacions inicials amb unes freqüències haplotípiques 0.24, 0.06, 0.56 i 0.14 en una població i 0.06, 0.24, 0.14 i 0.56 en l'altre, les freqüències haplotípiques de la població mesclada en igual proporció seran 0.15, 0.15, 0.35 i 0.35 que no es corresponen a les freqüències haplotípiques en l'equilibri de la població mesclada: 0.25, 0.25, 0.25 i 0.25). Finalment, la **selecció**, en afavorir un determinat al·lel també pot actuar augmentant el desequilibri de lligament al fer-nos variar les freqüències al·lèliques (Terwilliger et al. 1998; Terwilliger i Weiss 1998).

1.2.2. Mesura del desequilibri de lligament

Hi ha diverses mesures per a quantificar el desequilibri de lligament (D' , Δ , δ , d , Q , entre d'altres) (Devlin i Risch 1995; Guo 1997) però el component bàsic de totes elles és el coeficient de desequilibri (D) que és la diferència entre les freqüències haplotípiques observades i les esperades. Aquestes mesures però, assumeixen l'existència de només dos al·lells per locus (cas dels SNPs) en casos de parelles de *loci*. Així doncs:

$$D = \text{freqüències observades} - \text{freqüències esperades}$$

I per tant, $D = 0$ correspondrà a una situació d'equilibri de lligament, essent D dependent de les freqüències al·lèliques en la població. La mesura més utilitzada de totes és una variant de D , el coeficient D' , introduïda per Lewontin (1964) que es defineix com:

$$D' = D / D_{\max}$$

INTRODUCCIÓ

On D_{\max} ve determinat pel valor més petit de p_1q_2 i p_2q_1 , és a dir, $\min(p_1q_2, p_2q_1)$. D' pot variar entre -1 i 1 , havent-hi LD si és significativament diferent de zero. Aquest coeficient resulta molt útil per comparacions de desequilibri de lligament entre *loci*, ja que el seu rang és més independent de les freqüències al·lèliques (Zapata i Visedo 1995). La seva significació es pot analitzar amb l'estadístic χ^2 però també amb mètodes de permutació.

Pel cas de diversos *loci* multial·lèlics (cas dels STRPs), una mesura recentment definida (Zhao et al. 1999) és el coeficient ξ . Aquesta mesura, que permet també fer comparacions entre poblacions, ens quantifica la desviació que trobem sobre la barreja a l'atzar dels diferents al·lèls dels diferents *loci* i en aquest cas la seva significació (p) pot ser calculada també amb un test de permutació. L'estima de ξ ve definida per:

$$\hat{\xi} = \sqrt{2v} \frac{1}{n} \left(\frac{t - \mu}{\sigma} \right)$$

On n és el nombre d'individus, v és la complexitat (graus de llibertat) del sistema haplotípic per població, μ i σ són, respectivament, la mitja i la desviació *standard* de la distribució de les permutacions i t és l'estadístic observat de la raó de versemblança.

1.2.3. Desequilibri de lligament, distància genètica i població

Si D o D' difereixen significativament de zero, es correspon a una situació de LD. El desequilibri de lligament tendirà a ser menor com més allunyats estiguin dos *loci* i també disminuirà al llarg del temps com a resultat de la recombinació. En aquest sentit, el desequilibri de lligament dóna una estima de la freqüència de recombinació i per tant, de la distància física que hi ha entre dos *loci* (Jorde 2000).

Així, el desequilibri de lligament és una eina útil per a localitzar **gens mendelians** causants de malaltia, ja que com més proper estigui un marcador a

un gen candidat, més probable serà que estiguin en LD el gen i el marcador o marcadors. A diferència dels tradicionals estudis de lligament, on estudiant les recombinacions produïdes entre dos o més *loci* en famílies ens permet fer una aproximació a la possible zona on es troba un gen, el desequilibri de lligament incorpora l'efecte de moltes generacions de recombinacions permetent una millor resolució del mapatge gènic. El gen CFTR, mutacions en el qual causen la fibrosi quística, és un bon exemple de gen localitzat a partir del fort LD que es va observar entre diferents marcadors i la malaltia (Estivill 1987; Kerem et al. 1989). Altres gens causants de malaltia localitzats amb anàlisi de desequilibri de lligament són el de la malaltia de Huntington (MacDonald et al. 1992), el de la malaltia de Wilson (Bowcock et al. 1994b) o el gen de l'hemocromatosis (Feder et al. 1996). De fet, el clonatge posicional es basa en l'existència de LD entre alguna variant d'un polimorfisme proper al gen i la variant genètica causant de la malaltia.

Per a la cerca de **gens per a malalties complexes**, on podem tenir l'acció de diferents gens o de l'ambient, l'anàlisi genètica es fa més difícil i en aquest cas és important escollir poblacions humanes on hom pugui preveure que el desequilibri de lligament sigui major. Com que el desequilibri de lligament és un reflex de les recombinacions passades, poblacions amb diferents històries demogràfiques tindran diferents patrons de desequilibri.

En aquest sentit podem diferenciar dos tipus de poblacions: aquelles que al llarg de la seva història han mantingut una grandària efectiva constant i les que, en un moment donat, van patir una expansió. Les primeres acostumen a ser poblacions petites i aïllades mentre que les segones són poblacions actualment amb un gran nombre d'individus. La deriva, que afecta a les poblacions petites, farà, per tant, que augmenti el desequilibri de lligament en aquestes poblacions i no en les que s'han expandit (Slatkin 1994; Laan i Pääbo 1997).

Una altra distinció que hom ha de fer, però, és si aquesta expansió de la població ha estat recent o si és una expansió antiga. En el cas de l'expansió recent sí que trobarem un fort desequilibri de lligament entre *loci* separats per

INTRODUCCIÓ

diversos centiMorgans (cM) de distància. Aquest és el cas de la població finlandesa, que va ser fundada per un petit nombre d'individus ara fa uns 2.000 anys i actualment té més de 5 milions de persones. Un nombre petit de fundadors augmenta la probabilitat que una mutació causant de malaltia, associada a un haplotip concret, fos introduïda en la població per un sol individu, i si la fundació és prou recent, la mutació actualment encara estarà associada al mateix haplotip, ja que les recombinacions encara no hauran trencat el desequilibri de lligament existent. En la població finlandesa, on més de 30 malalties genètiques (la majoria d'elles autosòmiques recessives) hi són especialment freqüents (de la Chapelle i Wright 1998), el mapatge per LD ha permès de descobrir la localització dels gens causants d'aquestes malalties.

A partir de l'èxit en el mapatge de gens *mendelians*, en poblacions recentment expandides a partir de pocs individus, els estudis de cerca de *loci* relacionats amb malalties complexes s'han dirigit cap a poblacions petites i aïllades que hagin mantingut una grandària constant de la població i on la deriva hi ha actuat augmentant el desequilibri de lligament. Aquestes poblacions, també amb pocs individus fundadors i amb poca barreja d'altres poblacions, raonablement haurien de tenir menys variabilitat genètica i, per tant, també menys variabilitat en gens que afectin malalties complexes tals com la diabetis, malalties psiquiàtriques, malalties cardiovasculars, etc. Exemples d'aquests aïllats genètics i malalties estudiades són, entre d'altres, l'asma i la població de l'illa de Tristan da Cunha (Zamel et al. 1996) o l'esquizofrènia i la població de Palau a la Micronèsia (Myles-Worsley et al. 1999). Estudis recents (Lonjou et al. 1999; Eaves et al. 2000; Jorde et al. 2000) però, posen en dubte el diferent comportament de les poblacions aïllades genèticament i per tant que realment siguin més útils que la població general per al mapatge de gens relacionats amb malalties complexes.

1.2.4. Desequilibri de lligament i edat de la població

D'altra banda, és a les poblacions *antigues* (com les africanes) on trobem valors més baixos de desequilibri de lligament per a molts *loci* estudiats (Kidd et

al. 1998, 2000; Tishkoff et al. 1996, 1998, entre molts d'altres). Tots aquests estudis donen suport al model de l' *Out of Africa* (sortida d'Àfrica) en l'evolució dels humans moderns, amb un marcat efecte fundador associat a la sortida i posterior expansió fora de l'Àfrica dels humans moderns ara fa uns 100.000 anys. Tots aquests estudis troben menys LD a les poblacions sud saharianes que no pas a les poblacions no africanes, i arriben a la conclusió que un efecte fundador (després d'un coll d'ampolla) va establir un patró de LD, per als *loci* estudiats, que s'ha preservat en totes les poblacions no africanes estudiades fins els nostres dies.

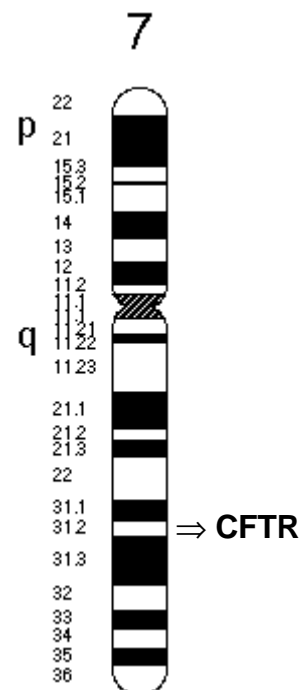
1.2.5. Límit del desequilibri de lligament

Si parlem de distància genètica (en kb) són molts els estudis que busquen el límit del desequilibri de lligament, normalment entre SNPs però també utilitzant STRPs, al llarg del genoma, per tal de facilitar la localització de gens per desequilibri. Kruglyak (1999), basant-se en simulacions per ordinador, prediu que, a la població general, el desequilibri de lligament no és probable que s'estengui més enllà de les 3 kb. Ara bé, com que sabem que diferents poblacions amb diferents històries demogràfiques tindran diferents límits de desequilibri, aquest valor pot ser molt variable. Per exemple, Gordon et al. (2000), en la població *afrikaner* de Sudàfrica, troba un significatiu LD entre *loci* distants a ~5.000 kb. Entre aquests dos valors (3 kb i 5.000 kb) se'n troben d'intermitjos, com per exemple l'estudi de Peterson et al. (1995) que, en població finlandesa, troba desequilibri de lligament fins a ~1.000 kb.

2. LA REGIÓ GENÒMICA 7q31: EL GEN CFTR

L'any 1989 va ésser identificat i clonat el gen **CFTR** (*Cystic Fibrosis Transmembrane Conductance Regulator*) (Kerem et al. 1989; Riordan et al. 1989; Rommens et al. 1989), les mutacions en el qual produeixen la fibrosi quística. El gen, d'unes 250 kb, compren 27 exons i es troba al braç llarg del cromosoma 7 (7q31.2) (Figura 1). La fibrosi quística és la malaltia autosòmica recessiva més comuna en poblacions d'origen europeu. La seva incidència, com a terme mitjà, és d'un cada 2.500 naixements, essent-ne una de cada 25 persones portadora.

Figura 1. Localització cromosòmica del gen CFTR



Han estat descrites més de 970 mutacions al gen CFTR que produeixen la malaltia (Cystic Fibrosis Mutation Data Base). Gairebé el 70% dels cromosomes mutats (cromosomes CF) porten una deleció de 3 parelles de bases (bp) al codó 508 (mutació $\Delta F508$). Només quatre altres mutacions (anomenades G542X, N1303K, G551D i W1282X) tenen freqüències globals en els cromosomes CF superiors a l'1% (TCFGAC 1994, Estivill et al. 1997). Tal com serà comentat més

endavant, les freqüències d'aquestes mutacions varien a les diferents poblacions europees.

2.1. Clínica de la fibrosi quística

L'elevada quantitat de mutacions diferents, variacions a nivell de genotip o heterogeneïtat al·lèlica, fa que el fenotip CF (*Cystic Fibrosis*, MIM 219700), malgrat ser una malaltia monogènica, sigui molt variable entre individus no emparentats i també dins d'una mateixa família. El terme *fibrosi quística* va ser introduït per Anderson (1938) en anomenar "fibrosi quística del pàncreas" la destrucció, produïda per la malaltia, de la funció exocrina del pàncreas.

La forma clàssica i més comuna de la malaltia afecta el tracte respiratori, on es presenta amb una obstrucció crònica pulmonar; i el pàncrees, produint una insuficiència pancreàtica exocrina. També es dona una elevada concentració de clor a la suor i infertilitat masculina. Quan es va descriure la fibrosi quística, el pronòstic de vida dels malalts era de mesos o pocs anys, no sobrevivint a la infància. Actualment, l'esperança de vida d'un pacient als països desenvolupats és d'uns 40 anys.

Al **tracte respiratori**, l'acumulació d'un moc espès per manca d'hidratació, obstrueix les vies i fa que es produeixin infeccions recurrents, causades majoritàriament per *Pseudomonas aeruginosa* i per *Staphylococcus aureus* (Welsh et al. 1995). Els primers símptomes són la tos persistent seguida de bronquiolitis i bronquitis. La colonització bacteriana es produeix per una combinació de l'augment de la viscositat del moc juntament amb una disminució en la mobilitat dels cilis de la mucosa, que dificulta la neteja de partícules de les vies aèries, creant un ambient bacterià favorable. *P. aeruginosa* s'adhereix a les cèl·lules aèries epitelials i és la pròpia proteïna CFTR, en el seu estat normal, la que fa de receptor i internalitza via endocitosi els bacteris dins la cèl·lula, netejant així les vies aèries i contribuint al mecanisme de defensa de l'hoste (Pier et al.

INTRODUCCIÓ

1996, Pier 2000) (Figura 2). En absència de CFTR, *P. aeruginosa* colonitza les vies aèries per una internalització defectuosa.



Figura 2. *Pseudomonas aeruginosa* penetrant a cèl·lules traqueals d'un ratolí infectat. Micrografia electrònica feta per GB Pier. MicrobeLibrary.org

Les infeccions respiratòries es tornen cròniques i de difícil eradicació un cop establertes. La patologia respiratòria és la causa de mort més freqüent dels pacients, que acaben patint hipoxèmia, hipertensió pulmonar i *cor pulmonale*.

La **insuficiència pancreàtica exocrina** provoca una deficiència d'enzims pancreàtics que causa una mala digestió de greixos i proteïnes amb un quadre de distensió abdominal i esteatorrea. La mala digestió provoca una malnutrició que afecta el creixement de l'individu.

Les **glàndules sudorípares** dels pacients tenen una elevada concentració de clor, sodi i potassi. El nombre de glàndules és normal, essent el principal

problema la dificultat per a la reabsorció de clor. La pell es torna salada, podent presentar cristalls de sal després d'una forta suada.

Més del 95% dels pacients masculins presenten alteracions als conductes de Wolff. Els conductes deferents, l'epidídim, i les vesícules seminals estan atrofiats o són absents. Només un 2-3% dels homes afectats són fèrtils. L'agenèsia congènita de conductes deferents (CBAVD, *Congenital Bilateral Absence of Vas Deferens*; MIM 277180), que comporta infertilitat masculina, es pot donar de forma aïllada (deguda a mutacions, normalment en heterozigosi, en el gen CFTR) o com una manifestació més de la fibrosi quística.

A la dona la fertilitat també es pot veure afectada, essent la falta d'ovulació o la presència d'un moc espès a l'úter, que impediria la migració dels espermatozous, les causes més freqüents d'infertilitat.

Síntomes menys comuns són el *meconium ileus* (MI) del nou-nat o obstrucció intestinal que es dona al naixement, diabetis o disfunció pancreàtica endocrina que cursa amb pancreatitis i alteracions al fetge que poden acabar en cirrosi. Sembla que el *meconium ileus* es dona amb l'actuació sinèrgica d'un altre locus (*cystic fibrosis modifier* a 19q13) (Zielinski et al. 1999).

El tractament de la fibrosi quística es centra principalment en la prevenció i el control de les infeccions respiratòries, en millorar la fluïdesa de les secrecions i en millorar la nutrició dels pacients.

2.1.1. Relació genotip-fenotip

La relació genotip-fenotip en malalts CF és complexa, especialment en la patologia pulmonar, malgrat ser una malaltia monogènica. El fenotip CF pot ser molt variable entre malalts no emparentats, però també entre malalts emparentats. Està ben establert que mutacions als dos al·lels del gen CFTR causen un fenotip CF. I també que són els individus homozigots per a la mutació $\Delta F508$ o heterozigots $\Delta F508$ més una altre mutació greu (com la G551D o la

INTRODUCCIÓ

W1282X), els que cursen amb el fenotip clàssic de fibrosi quística: obstrucció pulmonar, deficiència pancreàtica exocrina, infertilitat masculina i una elevada concentració de clor a la suor (Mickle i Cutting 2000).

2.2. Família de gens transportadors ABC (*ATP-Binding Cassette transporters*)

La comparació de la proteïna CFTR amb altres de conegudes va demostrar gran homologia amb els membres de la superfamília de transportadors ABC, presents tant a procariotes com a eucariotes. Aquesta superfamília consta de P-glicoproteïnes, de la família de resistència a múltiples drogues i de proteïnes implicades en el transport actiu a través de la membrana cel·lular. El nom de transportadors ABC es deu al lloc d'unió d'ATP (*Adenosine TriPhosphate*) altament conservat, que és el tret més característic d'aquesta superfamília (Higgins 1992). Altres característiques són l'existència de quatre dominis, dos dels quals són transmembrana i altament hidrofòbics (generalment formats per sis segments transmembrànics) i dos més d'hidrofílics situats a la cara citoplasmàtica de la membrana. Aquests dominis són els llocs d'unió de l'ATP.

2.2.1. Estructura de CFTR

El gen CFTR (també conegut com ABCC7) (MIM 602421), és el membre número 7 de la sub-família C de la família de gens transportadors ABC (*ATP-Binding Cassette*). Format per 27 exons i un promotor TATA-less, ocupa unes 250 kb al braç llarg del cromosoma 7 a la banda q31.2 (Figura 3). La seqüència del gen és coneguda (Zielenski et al. 1991; Genbank AC000111 i AC000061).

El gen produeix un mRNA (*messenger RiboNucleicAcid*) de 6,5 kb que codifica la proteïna CFTR (EC 3.6.3.49), de 1.480 aminoàcids i un pes molecular de 168.138 Daltons (Da).

La proteïna CFTR està formada per cinc dominis: dos dominis transmembrànics altament hidrofòbics (TMD1 o MSD1 i TMD2 o MSD2), cadascun format per sis segments transmembrana (TM1 a TM6 pel TMD1 i TM7 a TM12 pel TMD2); dos dominis intracel·lulars hidrofílics d'unió a nucleòtids (NBD1 i NBD2), on s'hidrolitza l'ATP; i un domini regulador intracel·lular (R), que conté residus fosforilables per proteïna quinases (en concret per la proteïna quinasa A o

INTRODUCCIÓ

PKA). Cap altre membre conegut de la família ABC posseeix aquest cinquè domini R, domini que regula la funció de CFTR (Higgins 1992).

2.3. Funció de CFTR

Situat a la membrana apical de les cèl·lules epitelials, CFTR és tant un canal de clor, com també un regulador d'altres canals iònics (Sheppard i Welsh 1999; Schwiebert et al. 1999) (Figura 4). Com a canal conductor de clor, CFTR està regulat per cAMP (*cyclic Adenosine MonoPhosphate*). Diferents estudis indiquen que els dominis de membrana són el propi porus del canal i que la fosforilació (per part de proteïna quinases) de residus específics al domini regulador obre el canal després d'hidrolitzar-se l'ATP en els NBD (Sheppard i Welsh 1999). Com a regulador d'altres canals iònics, CFTR interacciona al mateix epiteli respiratori amb canals de Na^+ (ENaC o *epithelial Na⁺ channel*), regulant-los negativament; amb altres canals de clor (ORCC o *outwardly rectifying Cl channel*), regulant-los positivament; i amb canals de K^+ del teixit renal (ROMK o *renal outer medullary K⁺ channel*), modulant la seva sensibilitat a altres membres de la superfamília ABC de transportadors (Schwiebert et al. 1999). Alhora, sembla que CFTR facilita la sortida d'ATP de la cèl·lula.

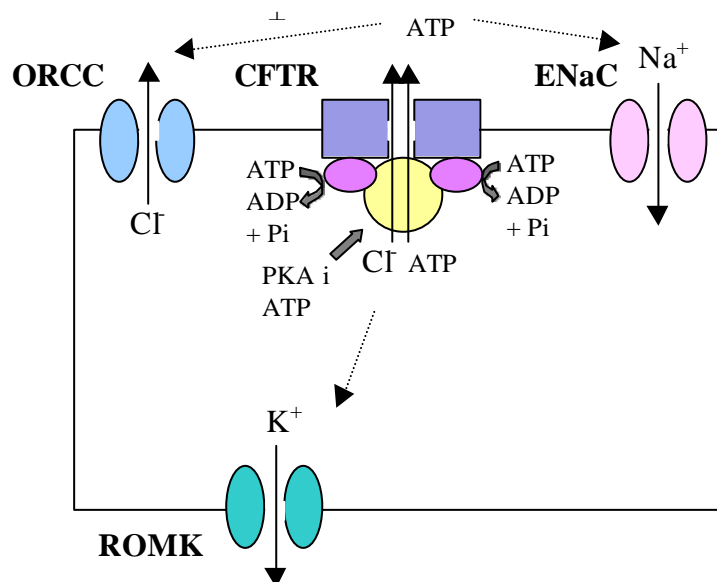


Figura 4. Esquema del funcionament i les interaccions de CFTR amb altres canals iònics en la cèl·lula epitelial.

CFTR es caracteritza per la diferent permeabilitat als anions respecte els altres canals de clor. Així, la seqüència de permeabilitat en CFTR és: $\text{Br}^- \geq \text{Cl}^- > \text{I}^- > \text{F}^-$, mentre que als ORCC hi ha més permeabilitat pel I que pel Cl⁻ (Sheppard i Welsh 1999).

En individus CF, es troba afectat el transport d'electròlits a través de la membrana cel·lular degut a un defecte en la proteïna CFTR. Aquest transport anormal causa alteracions en la viscositat del moc que recobreix l'epiteli, provocant els episodis recurrents d'obstrucció en els òrgans afectats.

2.4. Mutacions al gen CFTR

Hom ha descrit més de 970 mutacions al gen CFTR, nombre que s'amplia diàriament.

Els exons amb major nombre de mutacions són el 4 (70), 7 (60), 13 (86) i 17b (54), mentre que el nombre mínim de mutacions s'ha detectat als exons 6b (12), 14b (4), 23 (11) i 24 (8) (Cystic Fibrosis Mutation Data Base, dades de febrer de 2001). Sembla haver-hi una relació clara entre la llargada de l'exó i el nombre de mutacions que pot acumular, essent els exons més llargs els que n'acumulen més (Figura 5). Hi ha zones que són clarament *hotspots* de mutació, com són els dominis intracel·lulars d'unió a nucleòtids (NBDs) de la proteïna CFTR (Schwiebert et al. 1998). Les mutacions majoritàries (ΔF508 , G542X, G551D, N1303K i W1282X) estan localitzades entre el NBD1 i el NBD2. Hi trobem també, per exemple, codons flanquejants que acumulen diverses mutacions (codons 549 a 553 a l'exó 11, on s'acumulen almenys 13 mutacions, amb el cas particular del codó 549 on es donen 4 canvis diferents). Aquestes zones corresponen a vegades amb dinucleòtids CpG, com és la posició 1172 i també les posicions 1789 i 1790 del codó 553 (Morrall et al. 1994b), que se sap que són *hotspots* de mutació (Cooper i Krawczak 1990). Hom pot trobar també diferents substitucions nucleotídiques per una mateixa base, com per exemple les mutacions R347H (A),

R347L (T), R347P (C) que corresponen a les 3 substitucions possibles en la mateixa posició 1172 (G) del primer domini transmembrànic (TMD1) del gen.

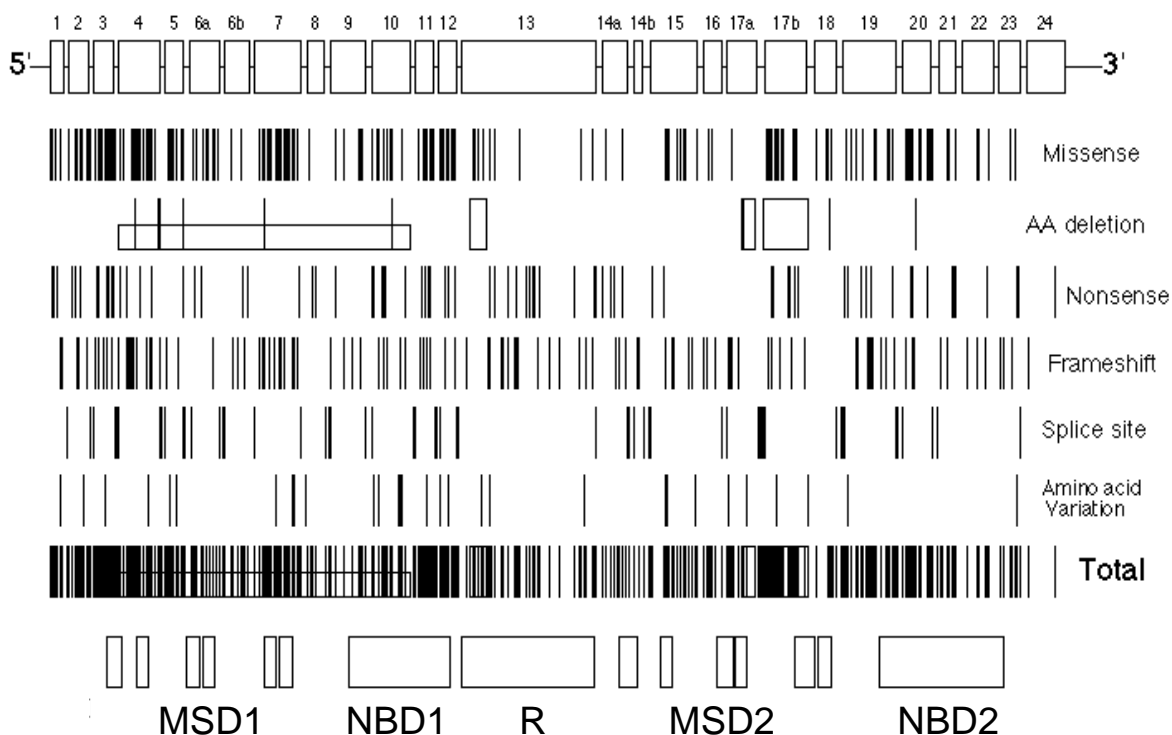


Figura 5. Distribució de mutacions al llarg del gen i la proteïna CFTR.

Les línies verticals indiquen la localització de les diferents mutacions agrupades segons el seu efecte: sentit erroni (Missense), deleció d'aminoàcids (AA deletion), aturada (Nonsense), canvi en la pauta de lectura (Frameshift), efecte sobre l'eliminació d'introns (Splice site) i canvis d'aminoàcid (Amino acid variation). Il·lustració estreta de Zielenski i Tsui (1995).

2.4.1. Tipus moleculars de mutació

Les diferents mutacions poden ser agrupades en diferents classes depenent del seu efecte a nivell molecular (Figura 6). Segons Zielenski (2000), les classes són les següents:

Classe I: Síntesi proteica defectuosa.

Les mutacions que pertanyen a aquesta categoria estan associades amb una falta o defecte en la síntesi de la proteïna, produint variants proteiques anormals (truncades, amb delecions...). Normalment són mutacions de tipus *nonsense* (mutacions d'aturada) com la mutació G542X, *frameshift* (canvi de pauta de lectura degut a delecions o insercions de nucleòtids), de *splicing* (afecten al procés d'eliminació dels introns) o de *missense* (de sentit erroni). El resultat és l'absència de CFTR a la membrana apical de les cèl·lules epitelials. Els efectes fenotípics són severos.

Classe II: Defecte en el processament proteic.

Les mutacions afecten a la correcta maduració de la proteïna i al transport cap a la membrana apical. Les variants proteiques produïdes estan associades amb fenotips severos. El resultat és també l'absència de CFTR a la membrana apical. La mutació $\Delta F508$ pertany a aquest grup. La proteïna CFTR $\Delta F508$ però, té activitat residual com a canal de clor, i podria funcionar com a canal iònic un cop portada a la membrana (Southern 1997). Es per això que la majoria de teràpies gèniques estan encaminades a corregir el processament i facilitar el transport de la proteïna mutada cap a la membrana. Una altra mutació d'aquest grup, que afecta el processament, és la N1303K (*missense*).

Classe III: Alteracions en la regulació del canal de clor.

Les proteïnes mutades es troben correctament situades a la membrana apical però són incapaces de funcionar com a canals. Les mutacions d'aquesta classe afecten els dominis NBD1 i NBD2, alterant la regulació de CFTR en

dificultar la unió i posterior hidròlisi d'ATP, necessària per a l'activació del canal. La mutació *missense* G551D afecta el domini NBD1 i pot afectar també la regulació d'altres canals de clor (ORCC).

Classe IV: Conductància del canal de clor disminuïda

Les proteïnes mutades, situades correctament a la membrana apical, tenen defectes en la conducció del canal de clor. Les mutacions afecten el domini MSD1 que està implicat en la formació del propi canal.

Classe V: Reducció en la síntesi o el transport proteic

Són mutacions que redueixen la quantitat normal de la proteïna CFTR. Poden ser mutacions al promotor del gen que en redueixen la transcripció, mutacions que produeixen *splicing* alternatiu, mutacions que causen una incorrecta maduració de la proteïna, etc... Totes elles produeixen CFTR normal però en menys quantitat.

Classe VI: Disminució en l'estabilitat proteica

La proteïna present a la membrana és funcional però inestable. Són mutacions *nonsense* o *frameshift* que causen una deleció a l'extrem C terminal de CFTR (Haardt et al. 1999). Estan associades amb un fenotip sever.

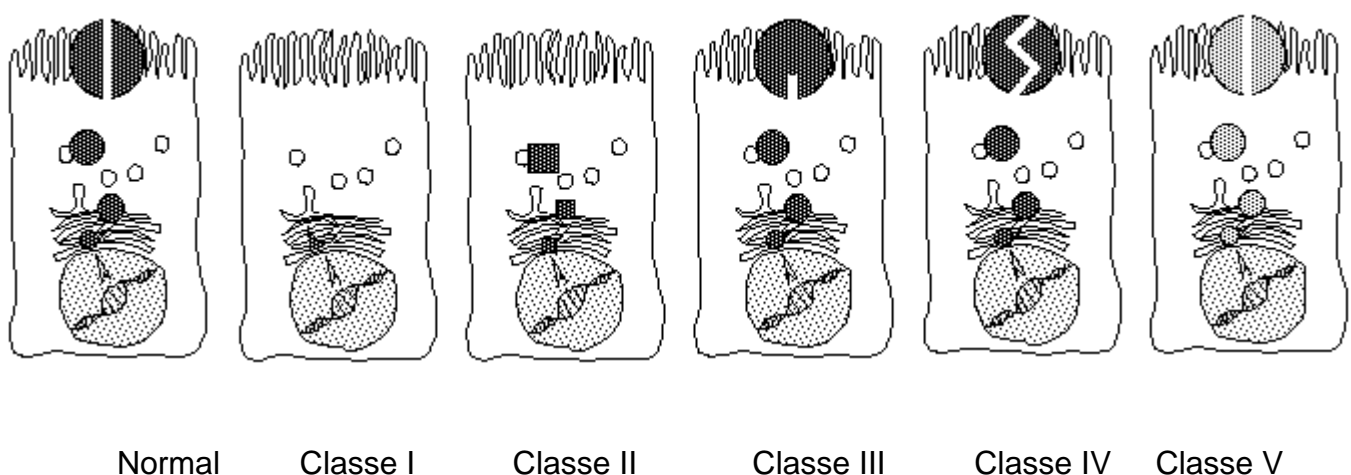


Figura 6. Representació cel·lular de les diferents classes de mutacions (Il·lustració extreta de Zielenski i Tsui, 1995).

2.4.2. La mutació $\Delta F508$ i la seva distribució geogràfica

La principal mutació que causa CF és una deleció de 3 bp al codó 508 de CFTR (entre els nucleòtids 1652 i 1655) que provoca la pèrdua d'una fenilalanina (Phe), a l'exó 10 de la proteïna, afectant el domini NBD1 (Kerem et al. 1989; Riordan et al. 1989; Rommens et al. 1989) i alterant la correcta maduració de la proteïna i el seu transport cap a la membrana apical. La mutació és l'anomenada $\Delta F508$ i es troba gairebé en el 70% dels cromosomes CF d'arreu del món (66% dels 43.849 cromosomes CF analitzats per TCFGAC, 1994) i d'Europa i Nord Àfrica (66,8% dels 27.177 cromosomes analitzats per Estivill et al., 1997), malgrat que la seva incidència varia depenent de la localització geogràfica, essent la més baixa la descrita a Tunísia (17,9%) (Messoud et al. 1996) i la més elevada la descrita a Dinamarca (86,8%) (Schwartz et al. 1990), en el context europeu. En general, i degut a una major heterogeneïtat mutacional a les poblacions mediterrànies, la freqüència de la mutació $\Delta F508$ és menor al sud que al nord d'Europa, formant un gradient del Sud-Est cap al Nord-Oest (Bertranpetit i Calafell 1996; EWGGCF 1990) (Figura 7). La diversitat de les mutacions causants de CF varia depenent de la població. En el cas per exemple, de poblacions de l'àrea mediterrània trobem una baixa prevalença de la mutació $\Delta F508$ i alhora una elevada heterogeneïtat de mutacions. Per exemple, en la població espanyola s'han caracteritzat el 90% dels cromosomes mutants amb 75 mutacions diferents (Casals et al. 1997); en canvi, poblacions del nord d'Europa, presenten una menor heterogeneïtat mutacional; per exemple, en la població irlandesa el 94% dels cromosomes mutants s'expliquen amb només 30 mutacions (Hughes et al. 1996).

2.4.3. Les mutacions **G542X**, **G551D**, **N1303K** i **W1282X** i la seva distribució geogràfica

Altres mutacions, amb freqüències a nivell mundial entre l'1 i el 2,5%, i també presents en pràcticament totes les poblacions estudiades són **G542X** (2,4%), **G551D** (1,6%), **N1303K** (1,3%) i **W1282X** (1,2%) (TCFGAC 1994). Estivill et al. (1997) troben unes freqüències molt similars estudiant només les poblacions europees i del nord d'Àfrica: **G542X** (2,6%), **G551D** (1,5%), **N1303K** (1,6%) i **W1282X** (1,0%).

La mutació **G542X** és un canvi nucleotídic G → T al codó 542 (nucleòtid 1756) de l'exó 11 (Kerem et al. 1990) que provoca el canvi d'una glicina (Gly) per un codó *stop* en la proteïna, al domini NBD1, alterant la síntesi de la proteïna. La mutació és present a la majoria de poblacions europees però es troba a freqüències més elevades a la Mediterrània (Figura 8). La freqüència màxima es dona a les Illes Balears (16,7%) (Estivill et al. 1997).

La mutació **G551D** és un canvi nucleotídic G → A al codó 551 (nucleòtid 1784) també de l'exó 11 (Cutting et al. 1990) que provoca el canvi aminoacídic glicina (Gly) → aspàrtic (Asp), afectant el domini NBD1 de la proteïna i alterant la regulació del canal de clor. La mutació és comuna al nord-oest i centre d'Europa (Figura 9), amb freqüències que varien des de 6,9% a Irlanda a 1,2% a Alemanya (Estivill et al. 1997).

La mutació **N1303K** és un canvi nucleotídic C → G al codó 1303 (nucleòtid 4041) de l'exó 21 (Osborne et al. 1991) que provoca el canvi aminoacídic: asparagina (Asn) → lisina (Lys) en la proteïna, al domini NBD2, afectant el correcte processament de la mateixa. La mutació es troba en la majoria dels països mediterranis (Figura 10), essent la freqüència a Tunísia la més elevada (17,2%) (Estivill et al. 1997).

INTRODUCCIÓ

La mutació **W1282X** és un canvi nucleotídic G → A al codó 1282 (nucleòtid 3978) de l'exó 20 (Vidaud et al. 1990) que provoca el canvi d'un triptòfan (Trp) per un codó *stop* en la proteïna, afectant el domini NBD2. La mutació és comuna a tota la conca mediterrània (Figura 11) (Estivill et al. 1997), essent la principal mutació (60%) descrita entre els jueus asquenasites, on s'associa amb un fenotip sever (Shoshani et al. 1992).

2.4.4. Altres mutacions menys freqüents

Altres 21 mutacions es troben a freqüències globals entre 0,1% i 0,9%, però la seva distribució és variable i la majoria estan restringides a determinades zones geogràfiques (TCFGAC 1994, Estivill et al. 1997) (Taula 1). La resta de mutacions descrites tenen freqüències globals inferiors, malgrat que poden presentar freqüències més elevades en determinades regions.

Mutació	Freqüència relativa (%)	Mutació	Freqüència relativa (%)
1717-1G→A	0,83	R347P	0,20
R553X	0,75	A455E	0,20
621+1G→T	0,54	R560T	0,18
R1162X	0,51	R334W	0,16
2183AA→G	0,36	3849+10kbC→T	0,15
R117H	0,29	3659delC	0,14
394delTT	0,23	1078delT	0,13
ΔI507	0,23	711+1G→T	0,11
1898+1G→A	0,22	2789+5G→A	0,11
G85E	0,21	S549N*	0,10
		2184delA*	0,10

Taula 1. Mutacions al gen CFTR que es troben a freqüències entre 0,1-0,9% a la població europea i mundial. Dades extretes d'Estivill et al. (1997), excepte (*): TCFGAC (1994)

2.5. Per què la CF és una malaltia tan freqüent?

Avantatge dels heterozigots

Amb una freqüència al·lèlica del 2% ($q = 0,02$) i una incidència d'un cada 2.500 naixements, la fibrosi quística és la malaltia autosòmica recessiva més comuna en les poblacions d'origen europeu. Una de cada 25 persones és portadora d'un al·lel mutant. I d'aquests al·lells mutants, gairebé el 70% corresponen a la mutació $\Delta F508$. Durant els últims anys, l'explicació més acceptada per l'elevada freqüència de la fibrosi quística en la població d'origen europeu ha estat aquella que postula un avantatge selectiu dels individus heterozigots (Bertranpetit i Calafell 1996). Aquest avantatge s'ha intentat relacionar especialment amb un efecte protector contra malalties infeccioses com el còlera (Gabriel et al. 1994) o la febre tifoidea (Pier et al. 1998, Pier 2000); però també amb la tuberculosi (Meindl 1987), la sífilis (Hollander 1982), o la grip (Shier 1979); i fins i tot amb una patologia respiratòria com l'asma (Schroeder et al. 1995), encara que altres estudis no han trobat en aquest cas cap associació (Mennie et al. 1995).

En la fibrosi quística, a part d'alteracions en el transport iònic en les cèl·lules epitelials pulmonars, de les glàndules sudorípares i del pàncreas, hom troba també alterat el transport d'ions clor en les cèl·lules epitelials intestinals; de forma que l'intestí dels individus malalts no secreta clor després d'estar exposat a toxines de bacteris com *Vibrio cholerae*, agent causant del còlera, com passaria a l'intestí d'individus sans, provocant diarrees (Romeo et al. 1989). La diarrea és causada per una disminució en l'absorció de sodi o per un augment en la secreció de clor. Així doncs, la fibrosi quística pot tenir un efecte protector en contra de les toxines diarreiques, disminuint la secreció de fluids, i per tant evitant la deshidratació, mitjançant la CFTR intestinal (Guggino 1994, 1999). Si els heterozigots secretessin menys clor com a resposta a enterotoxines, resultarien individus sans resistents a diarrees bacterianes. Ara bé, el *Vibrio cholerae*, que és endèmic a l'Índia, no va aparèixer a Europa fins l'any 1832, i per tant aquest sol agent no ha pogut ser la causa de l'elevada freqüència de la malaltia (Bertranpetit i

Calafell, 1996). L'observació directa de l'activitat CFTR en heterozigots (Högenauer et al. 2000) ha trobat nivells semblants als dels individus no portadors de mutacions a CFTR, i, per tant, contradiu la hipòtesi d'un avantatge dels heterozigots enfront les diarrees bacterianes.

D'altra banda, sembla que CFTR és el receptor de les cèl·lules epitelials gastrointestinals per *Salmonella typhi*, agent causant de la febre tifoidea (Pier et al. 1998; Pier 1999; Pier 2000). Una disminució dels nivells de CFTR en els individus heterozigots pot ser la causa d'una menor susceptibilitat a aquesta malaltia. La resistència a la febre tifoidea pot haver seleccionat l'al·lel CFTR Δ F508 a una elevada freqüència en determinades poblacions.

2.6. Origen de les mutacions més freqüents

Actualment, són ja a prop de 70 els polimorfismes descrits en les regions codificants del gen CFTR i uns 55 els descrits en les no codificants (Cystic Fibrosis Mutation Data Base, dades de febrer de 2001). Entre aquests polimorfismes la gran majoria són substitucions d'un sòl nucleòtid (SNPs), però també hi ha alguns microsatèl·lits (STRPs). Aquests dos tipus de marcadors poden ser utilitzats per estudiar l'origen, l'evolució i la possible recurrència de determinades mutacions. Els STRPs, amb la seva elevada variabilitat, faciliten els estudis evolutius, mentre que els SNPs consten d'al·lels molt més estables evolutivament. Totes les possibles combinacions de diferents tipus de polimorfismes es considera que són haplotips si estan físicament prou propers en el cromosoma com per passar a la següent generació com una unitat (Kidd i Kidd 1996). Existeix una forta associació (desequilibri de lligament) entre mutacions i determinats haplotips de microsatèl·lits, suggerint un origen únic per a tots els cromosomes que porten la mateixa mutació.

Cal remarcar que els termes emprats en genètica mèdica de mutació (variant genètica que provoca malaltia) i polimorfisme (variant neutra, normalment a alta freqüència) no són tan clarament diferenciats en el cas de la CF (ni en la majoria d'altres casos).

2.6.1. Edat de les principals mutacions CF

Morral et al. (1993, 1994a) i Estivill et al. (1994) varen estimar l'edat de la mutació $\Delta F508$ en ~ 2.600 generacions o el que és el mateix, uns 52.000 anys mitjançant l'anàlisi de la variabilitat de tres microsatèl·lits intragènics (IVS8CA, IVS17bTA i IVS17bCA) en cromosomes $\Delta F508$, assumint una determinada taxa de mutació dels mateixos i un determinat nombre de mutacions (el mínim possible) per construir l'arbre de màxima parsimònia pels haplotips dels STRPs. Kaplan et al. (1994) va qüestionar aquesta estima (i va ser respost, Estivill et al., 1994) i, suposant una major taxa de mutació per als microsatèl·lits i un menor nombre de mutacions per llinatge en l'arbre de màxima parsimònia dels haplotips dels tres microsatèl·lits, va donar una edat molt menor (~ 17.000 anys), edat més propera a l'estimada per Slatkin i Rannala (1997), a partir també de la variabilitat dels microsatèl·lits, de ~ 10.000 anys i a l'estimada anys abans per Serre et al. (1990), a partir d'estudis de desequilibri de lligament amb marcadors del gen, que donava una edat entre 3.000 i 6.000 anys.

Morral et al. (1993), a partir també de la variabilitat dels mateixos tres STRPs en cromosomes mutants, van estimar l'edat de les mutacions G542X i N1303K, en ~ 34.000 i ~ 35.000 anys respectivament.

2.6.2. Haplotips per cromosomes mutants i per cromosomes normals

Morral et al. (1996) analitzaren cromosomes normals i cromosomes portadors de mutacions (94 mutacions diferents), d'origen europeu, per diferents polimorfismes, SNPs i STRPs, estudiant la variació de freqüències haplotípiques entre ells. Els marcadors escollits eren J44, IVS6aGATT, T854 i TUB20 com a dialèl·lics i IVS8CA, IVS17bTA i IVS17bCA com a STRPs. L'evolució dels cromosomes CFTR ve marcada primer per l'haplotip definit pels marcadors dialèl·lics, establint un nucli on es van acumulant els canvis als STRPs. Així es veu que la mutació $\Delta F508$ és més freqüent en l'haplotip 1-6-23-1-31-13-2 (J44-

IVS6aGATT-IVS8CA-T854-IVS17bTA-IVS17bCA-TUB20), per la qual cosa s'assumeix que aquest és l'haplotip on s'originà $\Delta F508$. El mateix haplotip, i amb una petita variant (amb l'al·lel 33 a IVS17bTA), 1-6-23-1-31/33-13-2 és també on van sorgir les mutacions G542X i N1303K, haplotip que només es troba en un 1,4% dels cromosomes normals de les poblacions europees actuals. L'elevat nombre d'haplotips per STRPs associats a aquestes tres mutacions suggereix que són les mutacions més antigues de totes. Pel que fa a aquests haplotips, només pels STRPs (IVS8CA-IVS17bTA-IVS17bCA), respecte la mutació $\Delta F508$, els més freqüents són: 23-31/32-13 i 17-31/32-13 (Morral et al. 1994a); la barra (/) indica que per a IVS17bTA, els al·lells 31 i 32 es troben a freqüències altes i semblants. Respecte la mutació G542X els haplotips més freqüents són 23-33/32-13 i 23-33-14; i respecte la mutació N1303K són 23/22/24-31-13 i 23-30-13 (Morral et al. 1996).

Les altres dues mutacions més freqüents (G551D i W1282X) es troben associades cadascuna a un haplotip força diferent respecte als anteriors pel que fa als SNPs; entre elles, difereixen només en una repetició en un STRP. Són 1-7-16-2-7-17-1 per a G551D i 1-7-17-2-7-17-1 per a W1282X. (Morral et al. 1996).

En els cromosomes normals, l'haplotip per marcadors dialèl·lics (J44-IVS6aGATT-T854-TUB20) més freqüent trobat en població europea és 2-7-1-2 (55,4%), i dels 39 haplotips per STRPs (IVS8CA-IVS17bTA-IVS17bCA) diferents que es troben associats a aquest haplotip, els més freqüents són 16-31-13 (14%) i 16-30-13 (13,5%) (Morral et al. 1996).

2.6.3. Població d'origen de les mutacions $\Delta F508$, G542X i N1303K

Morral et al. (1994a) van postular el sorgiment de la mutació $\Delta F508$ en una població diferent de les poblacions europees actuals i la seva primera introducció a Europa durant el Paleolític, a partir de fa 40.000 anys. Una segona entrada durant el Neolític hauria augmentat les freqüències de la mutació i també les freqüències dels haplotips que porten l'al·lel 17, enlloc del 23, pel microsatèl·lit IVS8CA al Centre i Nord d'Europa.

INTRODUCCIÓ

Serre et al. (1990), en canvi, va postular un origen de la mutació $\Delta F508$ a l'Orient Mitjà i una posterior expansió cap a tot Europa durant el Neolític, ara fa entre 3.000 i 6.000 anys.

Les mutacions G542X i N1303K, originades sobre el mateix haplotip que la mutació $\Delta F508$ (1-6-23-1-31/33-13-2), s'haurien expandit per Europa també amb el Paleolític, probablement al mateix temps (Estivill i Morral 1996). D'altra banda, la mutació G542X té una freqüència més elevada a les costes mediterrànies, suggerint una segona entrada d'aquesta mutació, deguda potser a les migracions fenícies ara fa uns 3.000 anys (Casals et al. 1993). Tanmateix, s'ignora quin fou l'impacte demogràfic d'aquesta migració i la seva capacitat d'incidir sobre el *pool* gènic de les zones que l'acolliren.

Essent un avantatge selectiu dels individus heterozigots la principal causa de l'elevada freqüència de la fibrosi quística en la població europea, l'expansió d'unes determinades mutacions i no d'altres seria degut a un efecte fundador d'aquestes mutacions més comunes (Estivill i Morral 1996).

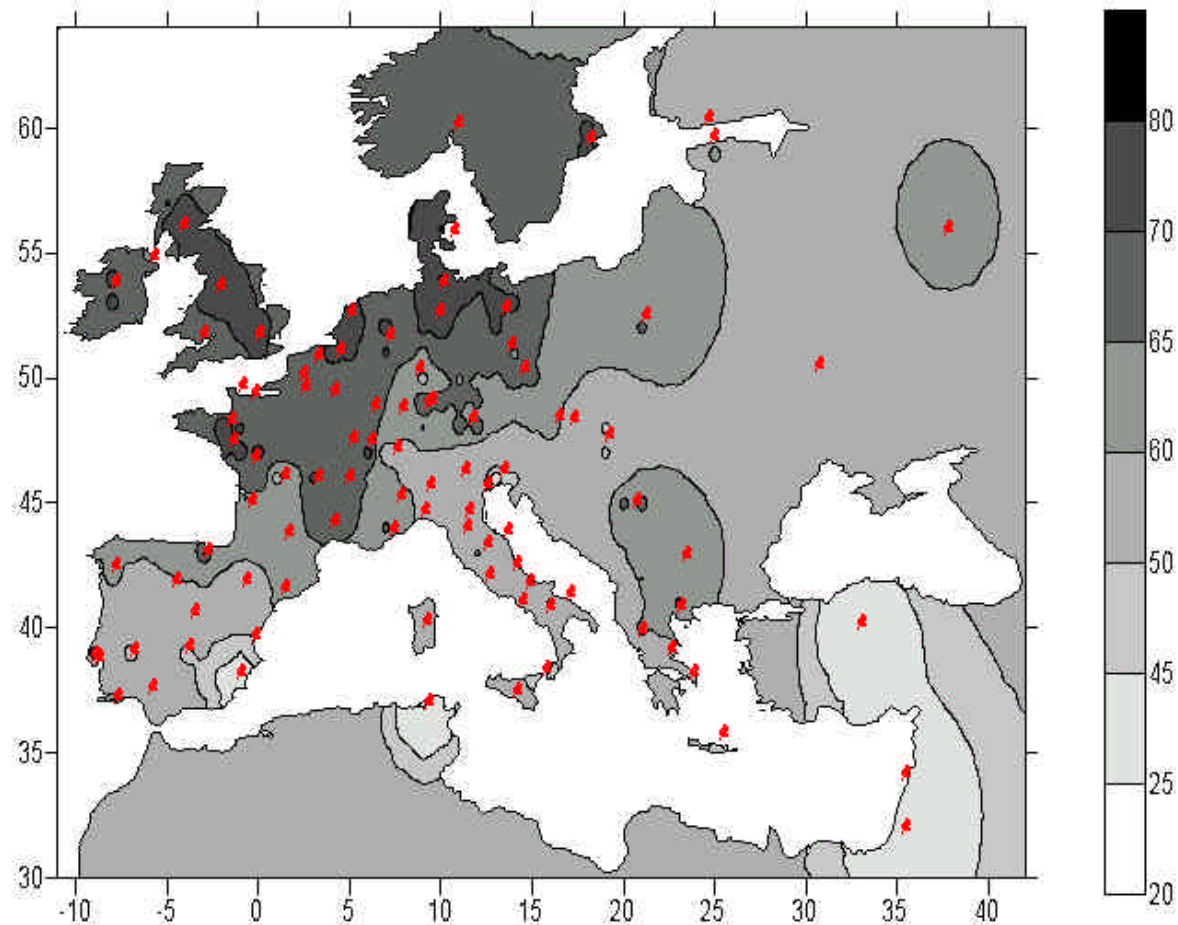


Figura 7. Distribució geogràfica de la mutació $\Delta F508$ a Europa i nord d'Àfrica. L'escala de grisos indica (en %) la freqüència de la mutació. Els punts vermells representen les poblacions de les quals es tenen dades. Només s'han inclòs les poblacions de les quals es disposava de més de 20 cromosomes. Les dades són d'una recopilació feta per Oscar Lao, basant-se en dades bibliogràfiques (Estivill et al. (1997), entre d'altres).

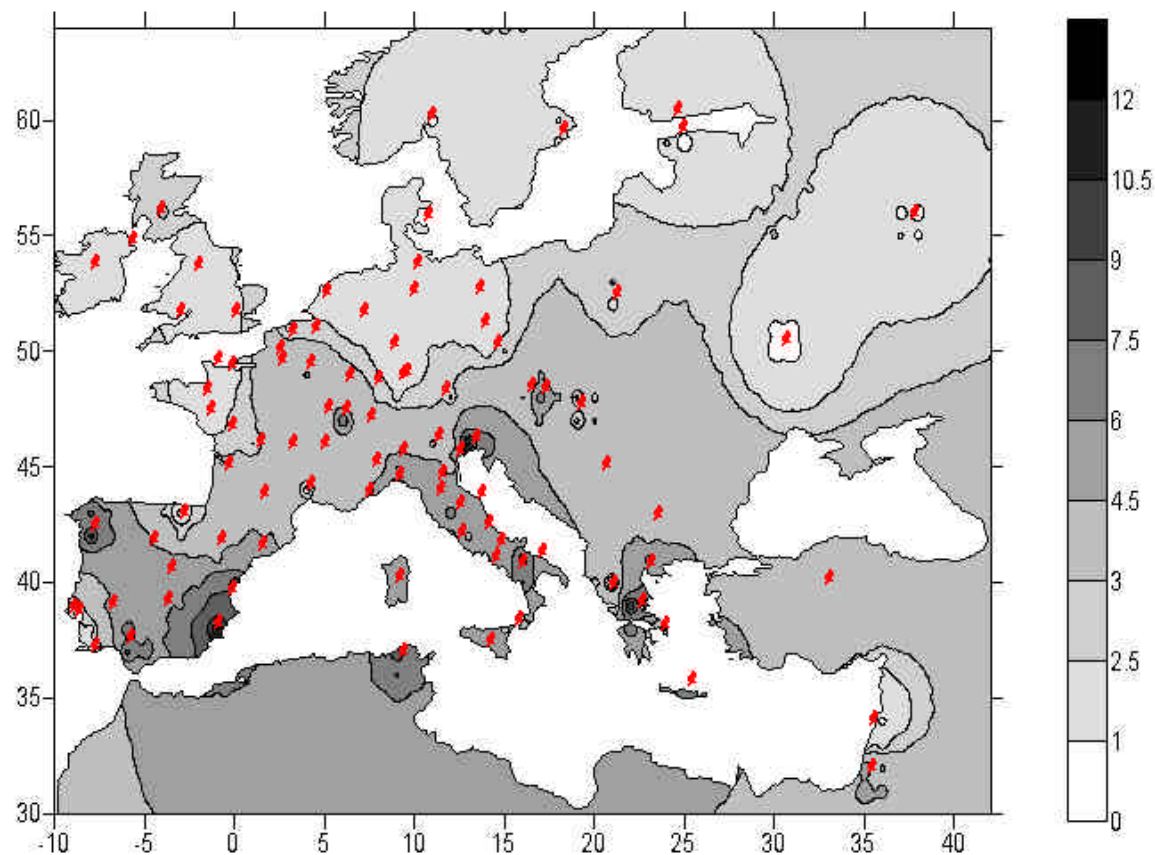


Figura 8. Distribució geogràfica de la mutació G542X a Europa i nord d'Àfrica. L'escala de grisos indica (en %) la freqüència de la mutació. Els punts vermells representen les poblacions de les quals es tenen dades. Només s'han inclòs les poblacions de les quals es disposava de més de 20 cromosomes. Les dades són d'una recopilació feta per Oscar Lao, basant-se en dades bibliogràfiques (Estivill et al. (1997), entre d'altres).

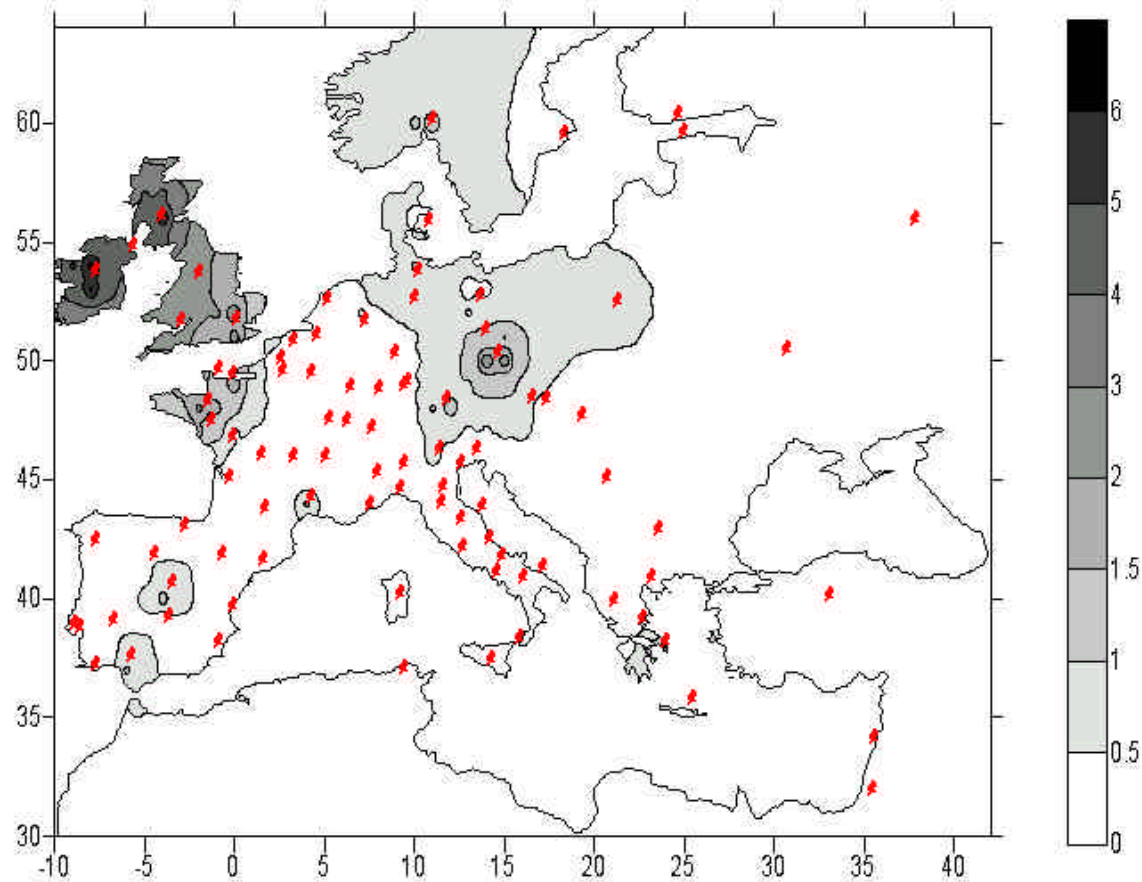


Figura 9. Distribució geogràfica de la mutació G551D a Europa i nord d'Àfrica. L'escala de grisos indica (en %) la freqüència de la mutació. Els punts vermells representen les poblacions de les quals es tenen dades. Només s'han inclòs les poblacions de les quals es disposava de més de 20 cromosomes. Les dades són d'una recopilació feta per Oscar Lao, basant-se en dades bibliogràfiques (Estivill et al. (1997), entre d'altres).

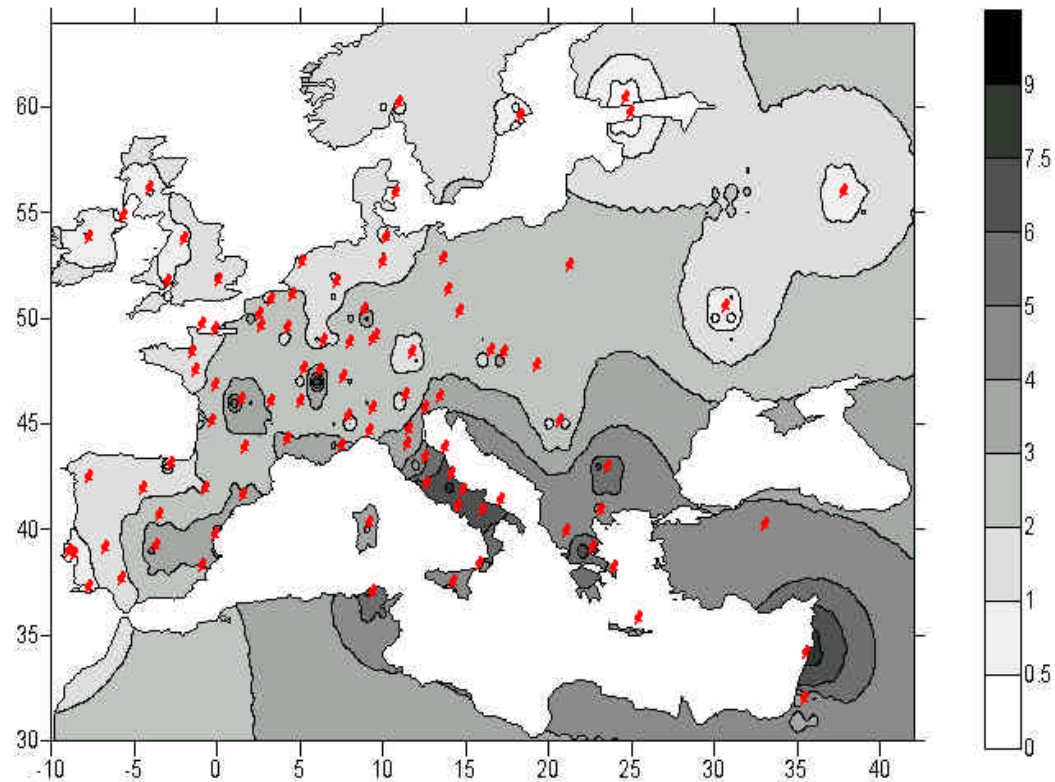


Figura 10. Distribució geogràfica de la mutació N1303K a Europa i nord d'Àfrica. L'escala de grisos indica (en %) la freqüència de la mutació. Els punts vermells representen les poblacions de les quals es tenen dades. Només s'han inclòs les poblacions de les quals es disposava de més de 20 cromosomes. Les dades són d'una recopilació feta per Oscar Lao, basant-se en dades bibliogràfiques (Estivill et al. (1997), entre d'altres).

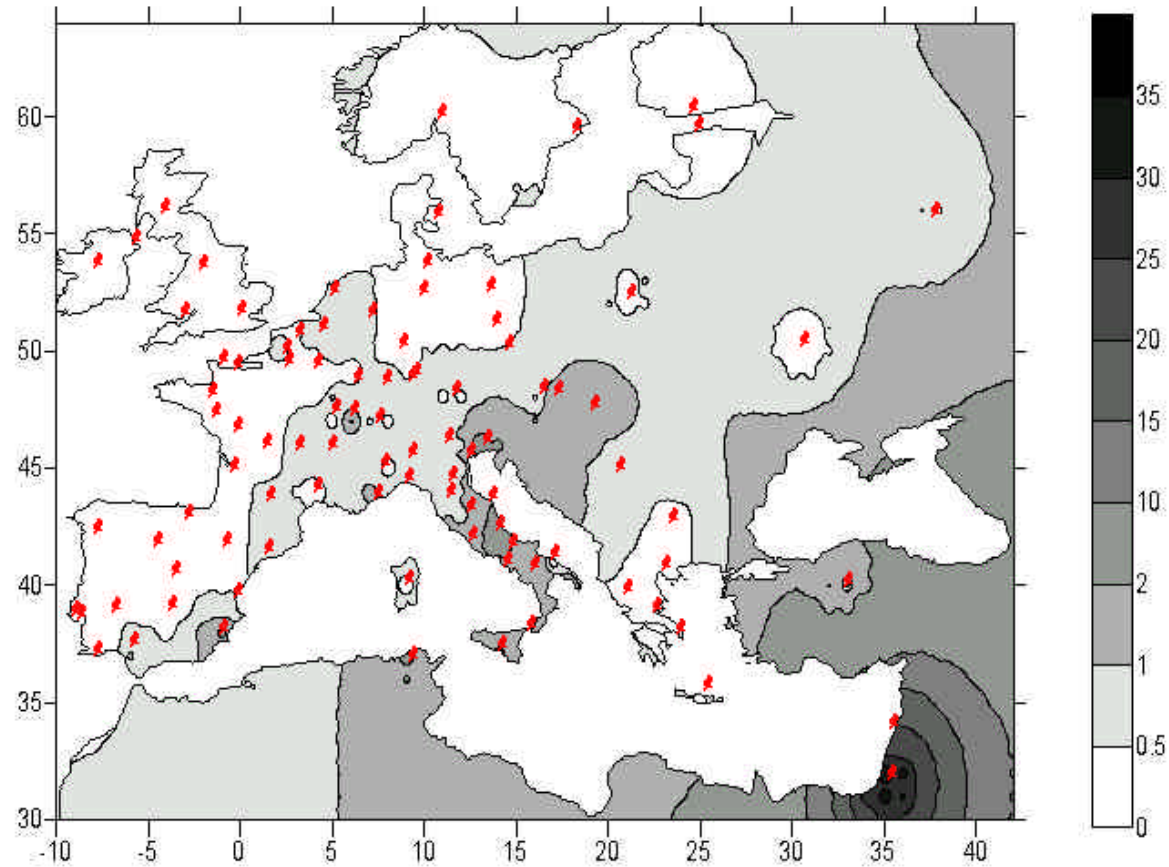


Figura 11. Distribució geogràfica de la mutació W1282X a Europa i nord d'Àfrica. L'escala de grisos indica (en %) la freqüència de la mutació. Els punts vermells representen les poblacions de les quals es tenen dades. Només s'han inclòs les poblacions de les quals es disposava de més de 20 cromosomes. Les dades són d'una recopilació feta per Oscar Lao, basant-se en dades bibliogràfiques (Estivill et al. (1997), entre d'altres).

3. LA REGIÓ GENÒMICA 1q21: ELS GENS GBA I PKLR

Separats per ~70 kb trobem, al braç llarg del cromosoma 1 (1q21), dos gens, les mutacions dels quals causen dues malalties autosòmiques recessives: el gen **GBA** (*Glucosidase, beta; acid*), mutacions en el qual causen la malaltia de Gaucher, per una deficiència de l'enzim β -glucocerebrosidasa; i el gen **PKLR** (*Pyruvate kinase L/R*), mutacions en el qual causen una deficiència de l'enzim piruvat quinasa, produint l'anèmia hemolítica hereditària o HNSHA (*Hereditary NonSpherocytic Hemolytic Anemia*) (Figura 12).

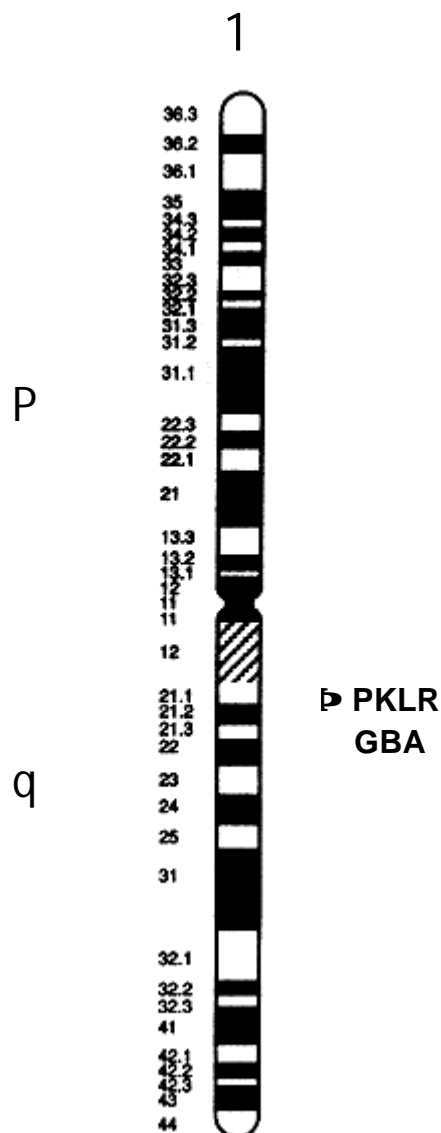


Figura 12. Localització cromosòmica dels gens GBA i PKLR

3.1. Altres gens a la regió genòmica 1q21

A més a més dels gens GBA i PKLR, n'hi ha d'altres de coneguts en aquesta rica regió en gens funcionals (Winfield et al. 1997; Vos et al. 1995) i estan representats a la Figura 13.

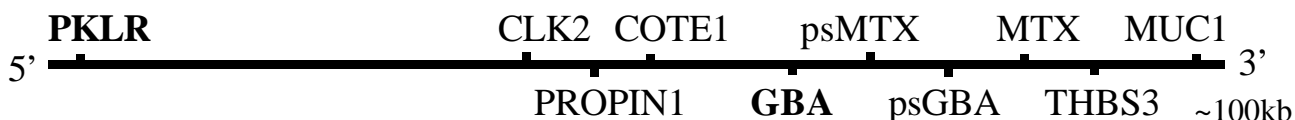


Figura 13. Representació esquemàtica dels gens localitzats a la regió 1q21. La zona representada abasta aproximadament unes 100 kb. (PKLR: piruvat quinasa; CLK2: proteïna quinasa, PROPIN1 o SCAMP3: proteïna secretora-portadora de membrana 3; COTE1 o C1ORF2; GBA: glucocerebrosidasa; psMTX: pseudogèn de la metaxina; psGBA: pseudogèn de la glucocerebrosidasa; MTX: metaxina; THBS3: trombospondina 3; MUC1: mucina epitelial polimòrfica.

Una duplicació a la regió, ara fa uns 40 milions d'anys, va donar lloc al pseudogèn de la glucocerebrosidasa (psGBA) i al pseudogèn de la metaxina (psMTX) (Winfield et al. 1997).

3.2. La regió GBA

L'any 1985 va ser clonat el gen **GBA** (*Glucosidase, beta; acid*) (Sorge et al. 1985a), les mutacions en el qual causen la malaltia de Gaucher. El gen, d'un 7,6 kb, consta d'onze exons i es troba al braç llarg del cromosoma 1 (1q21). La malaltia de Gaucher és la malaltia lisosòmica d'acumulació lipídica més freqüent, amb una elevada prevalença entre els jueus asquenaites (1/1.000). A 16 kb en direcció 3' del gen funcional es troba un pseudogèn (**psGBA**) amb una elevada homologia (96%) (Zimran et al. 1990). S'han descrit ja més de 130 mutacions (Human Gene Mutation Database: GBA) que produeixen la malaltia de Gaucher,

essent les més freqüents les mutacions anomenades N370S, L444P, 84GG i IVS2+1 (Charrow et al. 2000).

3.2.1. Clínica de la malaltia de Gaucher

L'any 1882, Phillipe Charles Ernest Gaucher (Figura 14), en la seva tesi doctoral, va descriure per primera vegada les manifestacions clíniques del que després es va anomenar malaltia de Gaucher. Hom calcula en ~80.000 les persones afectades a tot el món per aquesta malaltia (Brady et al. 1994).



Figura 14. P. C. E. Gaucher (1854-1918)

Les manifestacions clíniques de la malaltia són molt heterogènies. Els trets més característics són: aparició d'hepatomegàlia i esplenomegàlia (augment de la mida del fetge i de la melsa), anèmia i trombopènia (disminució del nombre d'eritròcits i de plaquetes, respectivament), degeneracions òssies i, en una minoria de pacients, es produeix una degeneració progressiva del sistema nerviós central.

L'afectació dels diferents òrgans i teixits es produeix per una acumulació lipídica (de glucocerebròsid) a l'interior dels lisosomes dels macròfags. Els principals òrgans afectats són aquells on trobem normalment elevada quantitat de macròfags, com són el fetge, la melsa i la medul·la òssia. La substitució progressiva de les cèl·lules normals per macròfags carregats de lípids (anomenats

INTRODUCCIÓ

cèl·lules de Gaucher) provoca un augment en la mida del fetge i la melsa i compromet la producció d'eritròcits i plaquetes a la medul·la òssia (Figura 15).

Clínicament, la malaltia de Gaucher es classifica en tres tipus depenent de l'absència (Tipus 1) o presència i gravetat (Tipus 2 i 3) de les alteracions al sistema nerviós central (Beutler i Grabowski 1995).

Malaltia de Gaucher Tipus 1 (MIM 230800)

És la forma més comuna de la malaltia. El sistema nerviós central no està afectat (forma no neuropàtica). L'edat d'aparició de la malaltia pot variar des de la infantesa fins a l'edat adulta. A pesar de l'elevada heterogeneïtat clínica que hi ha entre els pacients, els símptomes predominants són: hepatomegàlia i esplenomegàlia, alteracions òssies (deformacions, fractures), anèmia i trombopènia. La supervivència és variable, des de pocs anys (~ 6) fins a més de 80 anys.

És present en tots els grups ètnics (~1/60.000-1/360.000), però és especialment prevalent entre els jueus asquenasites (~1/1.000) (Grabowski 1997).

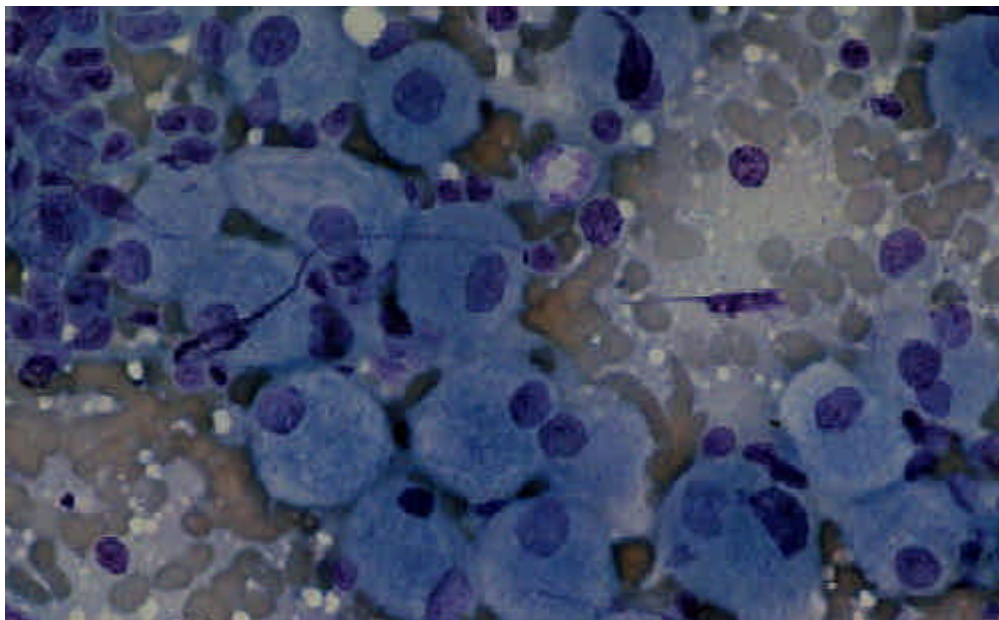


Figura 15. Cèl·lules de Gaucher a la medul·la òssia. (Tinció cel·lular realitzada per Karoly Balogh)

Malaltia de Gaucher Tipus 2 (MIM 230900)

És la forma més greu de la malaltia, amb afectació del sistema nerviós central (forma neuropàtica aguda), a part de les alteracions al fetge (hepatomegàlia) i la melsa (esplenomegàlia). Les anormalitats al moviment de l'ull són sovint les primeres manifestacions de la malaltia, amb l'aparició d'estrabisme bilateral fixat o apràxia oculomotora. La majoria d'infants amb aquest tipus de malaltia moren abans dels dos anys d'edat. La incidència però, a pesar d'estar present a tots els grups ètnics, és molt baixa (~ 1/500.000) (Grabowski 1997).

Malaltia de Gaucher Tipus 3 (MIM 231000)

La gravetat d'aquest tipus de malaltia de Gaucher és intermèdia a la dels tipus 1 i 2, amb una incidència a la població general ~1/100.000 naixements (Grabowski 1997). Hi ha afectació del fetge i de la melsa (hepatomegàlia i esplenomegàlia) i també presenta símptomes neurològics similars als que es donen al tipus 2 (forma neuropàtica sub-aguda), encara que aquests són menys greus i apareixen més tard en la infantesa, permetent la supervivència fins a l'edat adulta. Igual que en el tipus 2, les alteracions oculomotores solen ser els primers símptomes.

El prototip d'aquest tipus és una forma que es troba a elevada freqüència en una zona del nord de Suècia (província de Norrbotten), anomenada tipus de Norrbottenia, on tots els malalts tenen la mateixa mutació en el gen de la β -glucocerebrosidasa (Dahl et al. 1990).

El principal tractament que hi ha actualment per a la malaltia de Gaucher, i que en millora molts dels símptomes, és la teràpia de substitució enzimàtica, és a dir, l'administració de β -glucocerebrosidasa modificada dirigida als macròfags.

3.2.1.1. Relació genotip-fenotip

La malaltia de Gaucher es manifesta de forma diferent en individus amb diferent genotip, però també es pot manifestar de forma diferent en individus amb el mateix genotip. Els genotips més estudiats afecten les mutacions més freqüents en les diferents poblacions. En aquest sentit, hi ha una clara associació entre la presència de la mutació N370S i el tipus 1 de la malaltia (Strasberg et al. 1994). Aquesta mutació en homozigosi (N370S/N370S) representa el fenotip més lleu de la malaltia i només 1/3 dels homozigots la desenvolupen (Beutler et al. 1993). D'altra banda, els al·lels recombinants (al·lels que provenen de la seqüència del pseudogèn i normalment agrupen més d'una mutació) en homozigosi o juntament amb un al·lel L444P, són particularment prevalents entre els pacients amb el tipus 2 de la malaltia (Tayebi et al. 1998). I una homozigositat per a la mutació L444P indica normalment el tipus 3 (Dahl et al. 1990; Ida et al. 1999). Casos més concrets, com la presència en homozigosi de la mutació D409H provoca una variant de la malaltia de Gaucher (*Gaucher-like disease*; MIM 231005) amb calcificacions cardiovasculars (Chabás et al. 1995). De fet la correlació genotip-fenotip, tot i tractant-se d'una malaltia monogènica, és complexa.

3.2.2. Estructura del gen GBA

Localitzat al cromosoma 1, a la banda 1q21, el gen GBA té aproximadament 7,6 kb i conté 11 exons i 10 introns (Figura 16) al costat d'un promotor amb caixes TATA-like i CAAT-like. La seqüència del gen és coneguda (Horowitz et al. 1989, Beutler et al. 1992b; GenBank J03059) i hi trobem diverses insercions de seqüències *Alu* (elements genètics mòbils, Batzer et al. 1996).

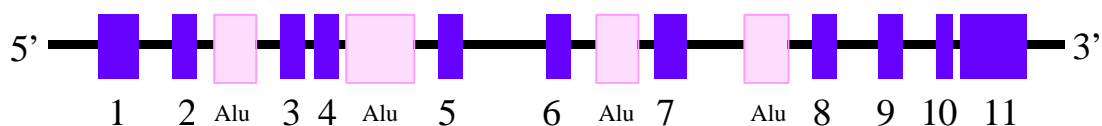


Figura 16. Representació esquemàtica del gen GBA. Les caixes blaves representen els exons i les caixes ratllades roses són insercions de seqüències *Alu*.

El gen produeix un mRNA de 2,6 kb que codifica per l'enzim β -glucocerebrosidasa (EC 3.2.1.45), proteïna de 497 aminoàcids i uns 65 kDa. Hom ha detectat mRNAs de diferents mides, deguts probablement a la utilització de llocs alternatius d'*splicing* (Reiner et al. 1988) o a la presència de mRNA del pseudogèn de GBA (Sorge et al. 1990).

3.2.3. Estructura del pseudogèn GBA (psGBA)

Situat a 16 kb en direcció 3' del gen funcional (Zimran et al. 1990) hom troba el seu pseudogèn. De seqüència també coneguda (Horowitz et al. 1989, GenBank J03060), conserva una gran homologia amb el gen (96%). Conté 11 exons i 10 introns però la mida global és més petita (unes 5,7 kb) degut a la falta de les seqüències *Alu* presents al gen (només en té una) (Horowitz et al. 1989) i a diferents delecions (la més gran és una de 55 bp a l'exó 9) (Figura 17). El pseudogèn es transcriu (Sorge et al. 1990) però no es tradueix a proteïna, degut a diverses mutacions que impossibiliten la síntesi de l'enzim funcional.

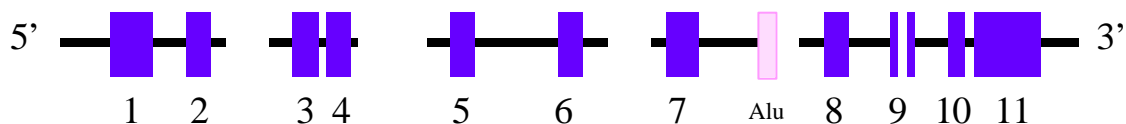


Figura 17. Representació esquemàtica del pseudogèn GBA. Les caixes blaves representen els exons i la caixa ratllada rosa és una inserció de seqüència *Alu*. Les discontinuïtats representen els fragments que li manquen en comparació al gen GBA.

3.2.4. Funció de GBA

El gen GBA codifica l'enzim β -glucocerebrosidasa o àcid β -glucosidasa (EC 3.2.1.45), hidrolasa lisosòmica, que catalitza la següent reacció:



L'enzim β -glucocerebrosidasa necessita d'una proteïna activadora anomenada saposina C o SAP-C. Aquesta reacció es dona a l'interior dels

lisosomes dels macròfags en el procés del catabolisme cel·lular. El mal funcionament de l'enzim provoca l'acumulació del substrat de la reacció (glucocerebròsid) en els lisosomes dels macròfags, afectant principalment els teixits on hi són abundants, com el fetge, la melsa i la medul·la òssia.

3.2.5. La saposina C

La malaltia de Gaucher també pot ser causada per una deficiència en saposina C, proteïna activadora de la β -glucocerebrosidasa. És una proteïna de baix pes molecular (només té 80 aminoàcids) que s'uneix a l'enzim facilitant la degradació del substrat. Aquesta proteïna està codificada pel gen prosaposina (PSAP) localitzat al cromosoma 10, a la banda q21-22 (Gavrieli-Rorman et al. 1992). L'absència de saposina C activa aboleix completament l'activitat de la β -glucocerebrosidasa, produint el fenotip característic de la malaltia de Gaucher (Horowitz i Zimran 1994).

Altres proteïnes descrites, una mala funció de les quals també pot ser responsable de la malaltia de Gaucher, són les proteïnes encarregades del transport de la β -glucocerebrosidasa als lisosomes (LAMP, *Lysosome-Associated Membrane Proteins*) (Zimmer et al. 1999).

3.2.6. Mutacions al gen GBA

Més de 130 mutacions causants de la malaltia de Gaucher han estat descrites fins el dia d'avui (Human Gene Mutation Database: GBA).

Segons la seva base molecular, les mutacions poden ser substitucions nucleotídiques que provoquen canvis d'aminoàcid, mutacions sense sentit, aparició de codons *stop*, alteracions en el procés de *splicing*; insercions o delecions que impliquen canvis en la pauta de lectura; i reordenacions entre el gen funcional i el pseudogèn per processos de recombinació o conversió gènica

que normalment originen al·lels complexos, és a dir, que contenen diverses mutacions.

Segons els seus efectes fenotípics, les mutacions es poden dividir en tres grups: mutacions letals, severes i lleus. Les mutacions letals són aquelles que impedeixen la formació de l'enzim (mutació 84GG i mutació IVS2 + 1); es consideren letals perquè no s'han trobat mai en homozigosi; les mutacions severes poden produir enzim, però aquest sol ésser poc estable i normalment van lligades amb els tipus neuropàtics de la malaltia (Tipus 2 o 3) (mutació L444P); les mutacions lleus estan només associades amb el Tipus 1 (forma no neuropàtica) de la malaltia (mutació N370S) (Beutler i Gelbart 1998).

3.2.6.1. La mutació N370S o 1226G

És la mutació més prevalent en els malalts Gaucher. La mutació N370S és una substitució nucleotídica A → G (nucleòtid 5841 a l'exó 9 del gen; numeració genòmica segons la seqüència del gen de Horowitz et al., 1989), a l'aminoàcid 370 de l'enzim, que provoca el canvi aminoacídic Asparagina (Asn) → Serina (Ser) (Tsuji et al. 1988). La mutació s'anomena també 1226G segons el número corresponent en el cDNA (*complementary DNA*) (Genbank M11080, començant la numeració a partir de la A del primer codó d'iniciació).

3.2.6.2. La mutació L444P o 1448C

La mutació L444P és un canvi nucleotídic T → C (nucleòtid 6433 a l'exó 10 del gen) a l'aminoàcid número 444 de l'enzim, que provoca el canvi aminoacídic Leucina (Leu) → Prolina (Pro) (Tsuji et al. 1987). La mutació s'anomena també 1448C segons el número corresponent en el cDNA. El nucleòtid "C" en aquesta posició correspon al nucleòtid present en la seqüència del pseudogèn.

3.2.6.3. La mutació 84GG

La mutació 84GG és una inserció d'un sol nucleòtid (una G) en la posició 1035 del DNA genòmic o 84 del cDNA en l'exó 2 del gen, que provoca el trencament en la pauta de lectura, acabant prematurament la proteïna (Beutler et al. 1991).

3.2.6.4. La mutació IVS2 + 1

La mutació IVS2 + 1 és un canvi nucleotídic g → a (nucleòtid 1067 a l'intró 2 del gen; els nucleòtids són en minúscula per la seva localització intrònica) que provoca alteracions en el procés de *splicing* produint un mRNA aberrant (Beutler et al. 1992a). El nucleòtid "a" en aquesta posició correspon al nucleòtid present en la seqüència del pseudogèn.

3.2.6.5. Altres mutacions puntuals

Altres mutacions també prevalents en la població de malalts Gaucher són: la mutació D409H, implicada en la variant de la malaltia de Gaucher (*Gaucher-like disease*); les mutacions A456P i V460V, presents en els al·lels recombinants; i les mutacions anomenades V394L, R463C i R496H, entre moltes altres mutacions (fins a més de 100 mutacions puntuals existents al gen GBA) que són molt menys comunes (Horowitz i Zimran 1994; Horowitz et al. 1998).

3.2.6.6. Al·lels complexos o recombinants

La presència del pseudogèn molt a prop del gen funcional (situat a 16 kb) provoca l'aparició de mutacions al gen causades per la transferència de fragments de la seqüència del pseudogèn cap al gen mitjançant processos de recombinació (ja sigui homòloga o desigual) o conversió gènica. S'anomenen al·lels complexos perquè normalment consten de diversos canvis nucleotídics respecte la seqüència normal. Els més comuns són els anomenats RecNcil i RecTL, però són bastants més els descrits (Martínez-Arias et al. 2001). RecNcil consta de tres canvis a l'exó

10 del gen: L444P, A456P i V460V. RecTL consta dels tres canvis anteriors més D409H a l'exó 9 (Eyal et al. 1990; Latham et al. 1990).

3.2.6.7. Distribució poblacional de les mutacions més freqüents

La malaltia de Gaucher és especialment prevalent entre els jueus asquenasites, amb una freqüència al·lèlica del 3,5% i una incidència de ~1/1.000 del tipus 1 de la malaltia. L'espectre de mutacions de la malaltia de Gaucher és molt diferent entre els jueus asquenasites i la població general on la malaltia és més rara.

Els jueus asquenasites representen aproximadament el 80% de la població jueva mundial, estimada en uns 13-14 milions de persones. El terme asquenasita és la paraula hebrea per designar als alemanys. Són els jueus que el segle IX migraren des de l'actual Israel cap a la vora del Rin i des d'allà (segle XIV) es desplaçaren cap a les actuals Polònia, Lituània, Bielorússia, Ucraïna i Rússia. Actualment la gran majoria viuen als Estats Units, a l'antiga Unió Soviètica i a Israel (Motulsky 1995).

Entre els **jueus asquenasites**, només quatre mutacions expliquen el 89-97% dels casos (Beutler et al. 1992a; Beutler et al. 1993; Horowitz i Zimran 1994; Horowitz et al. 1998). Les mutacions són: N370S, 84GG, L444P i IVS2+1. La mutació N370S representa aproximadament el 70-75% dels al·lèls mutats, essent la mutació més comuna. La segona mutació més freqüent és la 84GG amb un 12-15%. Les mutacions L444P i IVS2+1 representen respectivament un 4% i un 3% entre els malalts asquenasites. Altres mutacions prevalents en aquesta població són: V394L (3%), RecTL (2-3%), R496H (1-2%). Entre un 4% i un 10% són mutacions encara desconegudes (Horowitz i Zimran 1994; Horowitz et al. 1998).

En la **població general no jueva**, l'heterogeneïtat al·lèlica és més alta. La principal mutació causant de la malaltia de Gaucher és la mutació L444P, representant un 31-40% dels al·lèls mutats. En segon lloc tenim la mutació N370S amb un 23-27% dels al·lèls. Altres mutacions a més baixa freqüència són: RecNci (8%), D409H (1-5%), R463C (1-4%), RecTL (1-3%), 84GG, IVS2+1 i A456P (1%).

INTRODUCCIÓ

Entre un 23-24% són mutacions desconegudes (Horowitz i Zimran 1994; Horowitz et al. 1998).

Dues poblacions ben estudiades i que representen casos especials són la població japonesa (Iwasawa et al. 1997; Ida et al. 1998, 1999) i la població del Nord de Suècia (província de Norrbotten) (Dahl et al. 1990, 1993). En ambdós casos la mutació predominant es la mutació L444P.

A la **població japonesa**, hom no ha trobat fins ara la mutació N370S. La mutació predominant és la L444P (41%), seguida de la mutació F213I (14%). Les mutacions desconegudes comprenen el 16% del nombre total d'al·lels estudiats. En aquesta població, la correlació genotip-fenotip és molt complexa, havent-se descrit casos on la mutació L444P, tant en homozigosi com en heterozigosi, dona símptomes neurològics i no neurològics (Eto i Ida 1999).

A la **població del Nord de Suècia**, on tots els pacients són homozigots per la mutació L444P, aquesta s'hauria expandit a partir d'un al·lel fundador i un posterior creixement de la població en condicions d'aïllament genètic (Dahl et al. 1990, 1993).

3.2.7. Malaltia de Gaucher i avantatge dels heterozigots en la població jueva

La malaltia de Gaucher no és l'única malaltia especialment prevalent als jueus asquenasites. Entre d'altres, són especialment freqüents la malaltia de Tay Sachs i la malaltia de Niemann-Pick, dues malalties també d'acumulació de lípids als lisosomes (Motulsky 1995). Hom ha postulat que la història de la població asquenasita, amb dramàtiques expansions i contraccions de població, hauria afavorit a la deriva genètica com a causa major de l'elevada freqüència de determinades malalties (Risch et al. 1995). Altres autors però, postulen que seria algun avantatge dels individus heterozigots el que explicaria la seva elevada freqüència en l'actualitat. Kannai et al. (1994), per exemple, intenten relacionar una possible resistència a la tuberculosi com l'avantatge selectiu dels individus heterozigots, sense obtenir, però, resultats concloents en aquest sentit. Donat que

diferents mutacions han assolit una elevada freqüència i que també dues altres malalties lisosomals són comunes en la població jueva, l'avantatge dels heterozigots sembla ser l'explicació més probable: en algun moment de la història jueva, una deficiència parcial en algun enzim del metabolisme glicolipídic hauria estat avantatjosa (Beutler 1995; Boas 2000). Peleg et al. (1998) suggereixen que portadors de dues malalties (Gaucher i Tay Sachs) no tindrien cap avantatge evolutiu extra sobre els portadors d'una sola de les malalties. Boas (2000) a partir de simulacions matemàtiques conclou que la deriva sola no podria haver aconseguit les freqüències actuals de la mutació N370S, essent necessària un avantatge selectiu per aquesta mutació en la població asquenasa.

3.2.8. Polimorfismes i haplotips al gen GBA

Beutler et al. (1992b) varen descriure dotze polimorfismes distribuïts al llarg del gen GBA: una inserció/deleció i onze substitucions nucleotídiques (una d'elles amb tres variants nucleotídiques). Tots aquests polimorfismes es trobaven en fort desequilibri de lligament: dels 6144 haplotips diferents possibles ($2^{11} \times 3$) varen trobar només dos haplotips majoritaris (designats com a haplotip + i haplotip -), tant en població normal no africana com en individus amb la malaltia de Gaucher (Taula 2); en població africana varen descriure un tercer haplotip (anomenat Africà) que només diferia de l'haplotip majoritari – en una posició nucleotídica; un quart haplotip minoritari, trobat en població europea, diferia només de l'haplotip + en una posició nucleotídica. L'existència de dos haplotips majoritaris no es pot explicar amb l'aparició a l'atzar de n mutacions (12 en aquest cas) sobre la seqüència del gen. El nombre mínim d'haplotips que hauríem de trobar si n mutacions ocorren seqüencialment i tots els haplotips queden retinguts és $n + 1$, si no s'han produït recombinacions. Altres polimorfismes en el gen (tres SNPs com a mínim) han estat descrits per altres autors (Martínez-Arias et al., en premsa).

	-802	-725	-614	2128	2834	3297	3747	3854	3931	4644	5135	6144
+	a	c	c	a	c	g	g	t	g	del	c	g
-	g	t	t	g	g	a	g	c	a	a	a	a
Africà	g	t	t	g	a	a	g	c	a	a	a	a
Rar	a	c	c	a	c	g	a	t	g	del	c	g

Taula 2. Polimorfismes i haplotips en el gen GBA descrits per Beutler et al. (1992b). La numeració de les posicions nucleotídiques del gen és segons Horowitz et al. (1989). Els haplotips + i – es poden designar també Pv1.1+ i Pv1.1- basant-se en la posició 3931 segons Sorge et al. (1985b).

Les freqüències haplotípiques varien segons la població analitzada. En poblacions occidentals s’han descrit freqüències gèniques de l’haplotip – de ~0,65 i de l’haplotip + de ~0,35. Mentre que a les poblacions orientals les freqüències estan invertides: ~0,25 per l’haplotip – i ~0,75 per l’haplotip + (Glenn et al. 1994).

3.2.9. Origen de les mutacions més freqüents

Les mutacions més prevalents en població jueva causants de la malaltia de Gaucher es troben associades amb un determinat *background* haplotípic. Així, les mutacions N370S (al·lel G en la posició 1226 del cDNA) i IVS2 + 1 (al·lel a en el nucleòtid genòmic 1067) es troben sempre sobre un haplotip -, mentre que la mutació 84GG (inserció d’una G en la posició 84 del cDNA) es troba associada a l’haplotip +. Aquest fet és compatible amb un origen únic per a aquestes mutacions (Beutler 1993). Díaz et al. (1999, 2000) mitjançant anàlisis de lligament i de desequilibri de lligament estimen l’origen de la mutació N370S en uns 800-1000 anys (segles XI-XIII) en una població europea no jueva d’on hauria passat als jueus asquenassites.

La mutació de distribució mundial L444P (al·lel C en la posició 1448 del cDNA), en canvi, la trobem associada a ambdós haplotips, indicant l’existència de mutacions recurrents en aquesta posició (Beutler 1995). Recordem que l’al·lel C en aquesta posició correspon al nucleòtid present en la seqüència del pseudogèn

de GBA, el qual pot haver passat al gen per processos de recombinació o conversió gènica (Martínez-Arias et al. 2001).

3.3. El gen PKLR

L'any 1987 va ser clonat el gen PKLR (*Pyruvate kinase L/R*) (Tani et al. 1987), mutacions en el qual produeixen una deficiència de piruvat quinasa (PK: *Pyruvate kinase*), que és alhora la causa més comuna de l'anèmia hemolítica hereditària (HNSHA: *Hereditary nonspherocytic hemolytic anemia*). El gen, d'unes 8,4 kb, consta de 12 exons i es troba al braç llarg del cromosoma 1 (1q21). A 71 kb de distància, en direcció 3', trobem el gen GBA (Demina et al. 1998). Hom ha descrit ja més de 120 mutacions (Human Gene Mutation Database: PKLR; Zanella i Bianchi 2000) causants de la deficiència de PK, essent la mutació més freqüent l'anomenada G1529A.

3.3.1. Clínica de la deficiència de piruvat quinasa

La deficiència de PK (MIM 266200) és la causa més comuna de l'anèmia hemolítica hereditària o HNSHA, síndrome que es caracteritza per un escurçament en la vida mitjana dels eritròcits donant lloc a anèmia. La gravetat de l'anèmia és molt variable, des d'asimptomàtica fins als casos més greus on els eritròcits són pràcticament inexistents i la supervivència depèn d'una transfusió sanguínia. Igual que en d'altres anèmies, degudes a altres enzimopaties, apareixen els símptomes característics de processos hemolítics crònics com icterícia, esplenomegàlia i un augment en la incidència de càlculs biliars (Tanaka i Paglia 1995; Zanella i Bianchi 2000).

La prevalença de la deficiència de PK és estimada en 51 casos per milió en població caucasoide (Beutler i Gelbart 2000). La seva distribució geogràfica sembla ser mundial, malgrat que la seva prevalença en moltes regions és

desconeguda. Des que va ser detectat el primer cas l'any 1961 se n'han descrit poc més de 400 casos (Zanella i Bianchi, 2000).

3.3.2. Estructura de PKLR

Localitzat al cromosoma 1, a la banda 1q21, el gen PKLR té aproximadament 8 kb (8.409 nucleòtids) i conté 12 exons i 11 introns (Figura 18). El primer i el segon exó són transcrits específicament per dos promotors diferents específics de teixit, donant lloc a dos tipus de mRNA i posteriorment a dos isoenzims (PK tipus R i PK tipus L), expressats respectivament als eritròcits (PK-R) i als hepatòcits (PK-L). La seqüència del gen és coneguda (Lenzner et al. 1997; GenBank U47654) així com la seqüència d'ambdós cDNAs (Tipus R i Tipus L) (Kanno et al. 1991; Tani et al. 1988; Baronciani i Beutler 1993; GenBank AB015983 i M15465). La PK-R (EC 2.7.1.40) és un isoenzim de 574 aminoàcids i uns 62 kDa, mentre que la PK-L (EC 2.7.1.40) té 543 aminoàcids i uns 60 kDa. Existeixen a més a més dues formes de la PK-R, R₁ i R₂, que representen les formes presents als eritròcits immadurs (reticulòcits) i madurs respectivament.

Dos isoenzims més, PK-M₁ i PK-M₂, presents en altres teixits (PK-M1 a múscul estriat i cervell; PK-M2 a teixit fetal, leucòcits, plaquetes, pulmó, melsa, ronyó i teixit adipós) són producte de *splicing* alternatiu en un gen diferent, anomenat PKM, localitzat en el cromosoma 15, a la banda q22.

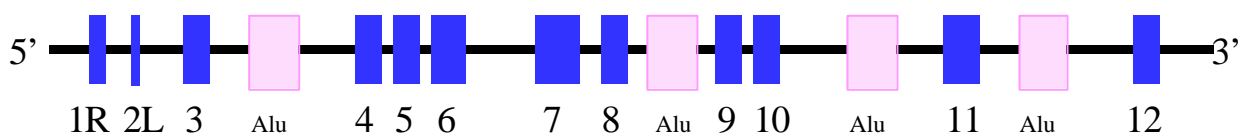


Figura 18. Representació esquemàtica del gen PKLR. Les caixes blaves representen els exons i les caixes ratllades roses són insercions de seqüències *Alu*. 1R és el primer exó del cDNA d'eritròcit i 2L és el primer exó del cDNA de fetge.

3.3.3. Funció de PKLR

La funció de la piruvat quinasa és catalitzar la formació de piruvat i ATP a partir de fosfoenolpiruvat (PEP) i ADP en la glucòlisi anaeròbica (degradació dels glúcids, en el citoplasma cel·lular, que té com a funció obtenir energia per a incorporar àcid fosfòric en forma d'ATP). L'enzim necessita magnesi (Mg^{2+}) i potassi (K^+) per a la seva correcta activitat.

Als eritròcits, en la deficiència de PK, augmenten els nivells intracel·lulars dels metabòlits de la via glucolítica anteriors al PEP, especialment els de 2,3-DPG (2,3-difosfoglicerat). Els nivells elevats de 2,3-DPG poden inhibir altres enzims de la mateixa via glucolítica, com l'hexoquinasa o la fosfofructoquinasa, i també inhibiria el pas cap a la via oxidativa de degradació glucolítica (via de les pentoses), contribuint a l'anèmia hemolítica. Al mateix temps disminueix la concentració d'ATP, que altera la permeabilitat de la membrana, contribuint a una progressiva deshidratació cel·lular, iniciant una prematura destrucció dels eritròcits (Tanaka i Paglia 1995).

3.3.4. Mutacions al gen PKLR i distribució geogràfica

Hom ha descrit ja més de 120 mutacions al gen PKLR (Human Gene Mutation Database: PKLR; Zanella i Bianchi 2000) que inclouen mutacions *missense*, *nonsense*, d'*splicing*, delecions i insercions. Són les mutacions del gen PKLR, que afecten als isoenzims R i L, les que estan implicades en l'anèmia hemolítica hereditària per dèficit de PK. El gen PKM, localitzat en un altre cromosoma, no està afectat en l'HNSHA per dèficit de PK (Beutler i Baronciani 1996).

Hom anomena les mutacions a partir de la seqüència del cDNA del tipus R, a partir de la A del codó d'iniciació ATG com a número 1. A partir d'aquí, hom numera els aminoàcids a partir de la primera metionina de l'enzim de tipus R.

Malgrat que la majoria de les mutacions només s'han trobat una vegada, per algunes d'elles existeix una clara distribució geogràfica i predilecció poblacional. La mutació G1529A és majoritària en **població d'origen europeu**, on representa entre el 40-45% del total de mutacions (Baronciani et al. 1995; Lenzner et al. 1997). És una substitució nucleotídica G → A, a l'exó 11 del gen, que implica un canvi d'aminoàcid Arg → Gln, a la posició 510 de l'enzim. La segona mutació més freqüent en població europea és la C1456T (11% del total) (Lenzner et al. 1997), una altra substitució nucleotídica C → T, també a l'exó 11 del gen, que implica el canvi de l'aminoàcid en posició 486 de l'enzim, Arg → Trp. A la **població d'origen asiàtic**, l'espectre de mutacions sembla ser ben diferent, no havent-se trobat la mutació G1529A en cap pacient japonès (Baronciani et al. 1995) i essent la mutació més comuna la C1468T (substitució nucleotídica C → T, també a l'exó 11 del gen, que implica el canvi de l'aminoàcid en posició 490 de l'enzim, Arg → Trp) (Zanella i Bianchi, 2000). En **població gitana** hom ha descrit una de les mutacions conegudes més greus del gen PKLR: una deleció de 182 nucleòtids a l'exó 11 (Baronciani i Beutler 1995).

3.3.5. Polimorfismes, haplotips i mutacions associades al gen PKLR

Són diversos els polimorfismes descrits al gen PKLR (Lenzner et al. 1997; Kugler et al. 2000) i inclouen fins al moment tres SNPs, un microsatèl·lit i una seqüència de poli-Ts (T_n) amb dos al·lels descrits. Ja Glenn et al. (1994) analitzant un SNP del gen GBA i un SNP del gen PKLR en diverses poblacions varen trobar que ambdós polimorfismes (i per extensió ambdós gens) estaven fortament lligats, ampliant els haplotips descrits pel gen GBA (haplotips + i -) al gen PKLR. Aquest anàlisi es va confirmar amb treballs posteriors d'altres autors (Demina et al. 1998; Rockah et al. 1998) que varen trobar un fort desequilibri de lligament entre les mutacions més comunes del gen GBA i determinats al·lels en polimorfismes del gen PKLR. Han estat descrits també (Lenzner et al. 1997; Kugler et al. 2000) els haplotips (pels polimorfismes de PKLR) associats amb les principals mutacions

causants de la deficiència de PK i es veu que cada mutació es troba associada, en general, a un haplotip concret, fet compatible amb un origen únic de les mutacions.

OBJECTIUS

Els objectius del present treball han estat:

1. Descriure les freqüències al·lèliques en poblacions humanes de tot el món de marcadors intragènics coneguts per CFTR i per GBA-PKLR. És d'especial interès (en el cas de CFTR) la descripció de la variació existent en un nou microsàtel·lit (IVS1CA), ja que fins al moment són molt poques les dades poblacionals que hi ha del mateix. A partir dels mateixos marcadors, reconstruir-ne els haplotips i descriure'n la seva distribució geogràfica.
2. Analitzar el desequilibri de lligament, entre *loci*, existent a cadascuna de les dues regions (CFTR i GBA-PKLR) i intentar entendre els factors genòmics i poblacionals que hi contribueixen. Pel cas del gen GBA (i per extensió la regió GBA-PKLR), és difícil d'entendre el fort desequilibri de lligament que hi ha descrit en poblacions d'origen europeu; es pot pensar que l'existència de només dos haplotips majoritaris, per un conjunt de dotze marcadors, pot ser deguda a raons poblacionals per un pur problema de la mostra analitzada i, en aquest cas, serà interessant analitzar aquesta regió genòmica en poblacions de tot el món. Aportar noves eines metodològiques en l'estudi del LD en haplotips complexos.
3. Coneixent el *background* haplotípic, en el gen CFTR, de les mutacions més importants causants de fibrosi quística (*background* que sabem que és diferent al dels cromosomes no portadors de mutacions); cercar i descriure la seva distribució geogràfica actual a nivell mundial, pot permetre reconstruir i entendre la història de les mutacions i de la pròpia malaltia, és a dir, els seus orígens i la seva dispersió. De fet un objectiu inicial era trobar alguna població no europea on els haplotips CF estiguessin presents entre els cromosomes normals per tal d'inferir el lloc d'origen de les mutacions CF.

MATERIALS I MÈTODES

1. POBLACIONS ESTUDIADES

Per a tenir una bona representació mundial, s'han estudiat un total de 18 poblacions humanes, que cobreixen tots els continents. Les poblacions estudiades són les següents:

Poblacions de l'Àfrica sud sahariana: Dues mostres de pigmeus (els *mbuti* i els *biaka*) i una mostra de població tanzana; els pigmeus **mbuti** són del NE de l'antic Zaire, concretament d'una zona selvàtica anomenada *Ituri*; els pigmeus **biaka** són del poble *Bagandu*, al SO de la República Central Africana; i els **tanzans** són d'*Ifakara*, al districte *Kilombero* de la regió *Morogoro* al SE de Tanzània. Població del nord d'Àfrica: els **sahrauís**, del Sàhara Occidental, però actualment localitzats en diversos camps de refugiats al sud d'Algèria. Poblacions de l'Orient Mitjà: drusos i jueus iemenites. Els **drusos**, que són una comunitat islàmica, de Galilea, al nord d'Israel; i els jueus **iemenites** que són immigrants del Iemen a Israel. Poblacions europees: **bascos** de la província de *Gipuzkoa*, dins del País Basc espanyol; **catalans** de les comarques gironines de Catalunya, en territori espanyol; **finlandesos** que no fossin d'origen suec, de Finlàndia; **russos** del districte *Zuevsky* al NE de Moscú, a Rússia; i la població **adygei**, del nord del Caucas, a la regió de *Krasnodar*, al SE de Rússia. Poblacions asiàtiques: **kazakhs** del poble *Aktasty*, a la vall del *Kegen*, a la regió *Almaty* del Kazakhstan, a l'Àsia central; **iacuts** de la República Autònoma de Iacútia de Rússia, a l'est de Sibèria; **xinesos** Han, ètnia majoritària de la Xina, del sud de la Xina, però que viuen a la badia de San Francisco, als Estats Units; i **japonesos** també de l'àrea de San Francisco. Població de l'àrea del Pacífic: els **nasioi** melanesis, de *Bougainville* a les Illes Salomon, a Melanèsia. I finalment les poblacions americanes: els indis **maia**, de l'estat de *Campeche* al Yucatán, Mèxic; i els **surui** de *Rondônia* al SO de l'Amazònia, al Brasil.

Per a una part del treball s'han inclòs en la mostra dos goril·les (*Gorilla gorilla*), un orangutan (*Pongo pygmaeus*) i tres ximpanzés (*Pan troglodytes*).

1.1. Origen i obtenció del DNA de les diferents poblacions

La major part de les mostres (mbuti, biaka, drusos, iemenites, finlandesos, russos, adygei, iacuts, xinesos, japonesos, nasioi, maia i surui) provenen de línies cel·lulars de limfoblasts mantingudes al laboratori de J. R. Kidd i K. K. Kidd, a la Universitat de Yale. Les mostres de sahrauís, bascos i catalans formen part de la col·lecció de mostres del laboratori de J. Bertranpetit i han estat recollides per membres de l'equip. La mostra de tanzans prové de la Unitat d'Epidemiologia i Bioestadística de l'Hospital Clínic de Barcelona, cedida per C. Menéndez. La mostra de kazakhs va ser recollida per membres del Departament d'Antropologia de la Universitat de Bolonya i cedida per D. Pettener. Les mostres de primats no humans són del zoològic de Barcelona.

El DNA de les mostres de sahrauís, bascos i catalans s'ha obtingut a partir de mostres de sang extretes amb *vacutainers* de 5 ml que contenen citrat sòdic. El DNA de la mostra de tanzans s'ha obtingut a partir de pèl·let d'eritròcits de mostres de cordó umbilical. En tots els casos l'extracció de DNA es realitzà mitjançant un mètode estàndard d'extracció amb fenol i cloroform (Sambrook et al. 1989), a partir de la lisi dels limfòcits i posterior digestió de les cèl·lules amb proteïnasa K, en presència d'EDTA i SDS (aquest darrer com a detergent), per tal d'eliminar les proteïnes. El DNA així extret es precipita seguidament amb etanol.

Per tal de determinar la quantitat de DNA obtinguda per a cada mostra es va mesurar l'absorbància en un espectrofotòmetre a 260 nm (una unitat d'absorbància a 260 nm equival a 50 ng/μl de DNA de doble cadena) i es va calcular la relació d'absorbàncies a 260 nm i 280 nm ($D.O_{.260} / D.O_{.280}$) per tal de determinar la relació d'àcids nucleics i de proteïnes a cada mostra. El valor òptim es troba al voltant de 1,8 (un valor inferior indica contaminació de proteïnes o restes de fenol). Per últim, les mostres es van dissoldre a 100 ng/μl com a solució de treball.

2. ANÀLISI DE MICROSATÈL·LITS (STRPs) I SUBSTITUCIONS NUCLEOTÍDIQUES (SNPs)

Els mètodes emprats per l'anàlisi dels dos tipus de marcadors estan descrits en els corresponents apartats dels capítols de Resultats. En aquest apartat només es pretén fer una síntesi dels marcadors utilitzats en les regions genòmiques estudiades (Taula 3).

Regió	Localització cromosòmica	Locus	Tipus marcador	Nombre d'al·lels	Poblacions analitzades
CFTR	7q31.2	IVS1CA	STRP	17	1-18
		IVS6aGATT	STRP	5	1-18
		IVS8CA	STRP	12	1-18
		T854	SNP	2	1-18
		IVS17bTA	STRP	36	1-18
		TUB20	SNP	2	1-18
PKLR	1q21	PKLR(ATT) _n	STRP	12	1-11, 13-18
		PKLR1705 c/a	SNP	2	1-11, 13-18
GBA	1q21	GBA2834 g/c/a	SNP	3	1-3
		GBA3931 a/g	SNP	2	1-11, 13-18
		GBA5135 a/c	SNP	2	1-3
		GBA6144 a/g	SNP	2	1-11, 13-18

Taula 3. Marcadors emprats i regió genòmica on es localitzen. Dins de cada regió es troben ordenats segons l'ordre 5'-3' en que es troben al gen. S'indiquen també les poblacions on s'ha analitzat cada polimorfisme (1: Biaka; 2: Mbuti; 3: Tanzans; 4: Sahrauí; 5: Drusos; 6: Iemenites; 7: Bascos; 8: Catalans; 9: Finlandesos; 10: Russos; 11: Adygei; 12: Kazakhs; 13: Iacuts; 14: Xinesos; 15: Japonesos; 16: Nasioi; 17: Maia; 18: Surui)

3. ANÀLISI ESTADÍSTICA DE LES DADES

Els tractaments estadístics utilitzats es troben detallats també en els corresponents apartats, de materials i mètodes, dels capítols de Resultats (Capítol I a Capítol IV).