

3.1. DISSENY EXPERIMENTAL

Per a estudiar la possible implicació del gen *DDR1* en l'esquizofrènia ens vam plantejar un disseny d'estudi d'associació cas-control. Com ja s'ha comentat, l'anàlisi d'un únic SNP o variant genètica en aquest tipus d'estudi pot conduir a resultats esbiaixats, per la qual cosa ens vam plantejar l'anàlisi de diversos SNPs i haplotips resultants en la regió del gen *DDR1*. D'altra banda, en el moment de començar aquest treball, no s'havien descrit, ni en la bibliografia ni en base dades d'accés públic, SNPs ni d'altres variants genètiques en el gen *DDR1*, a excepció d'un microsatèl·lit intrònic (Playford *et al*, 1996). Era necessari, doncs, fer un primer estudi d'identificació de polimorfismes o variants genètiques en aquest gen. En concret, vam analitzar les regions codificants, incloent les seqüències flanquejants intró-exó, en una submostra de malalts d'esquizofrènia. A continuació, es van triar els SNPs que causaven un canvi d'aminoàcid i els més distants en els extrems 5' i 3' del gen per estimar les freqüències haplotípiques i comparar la seva distribució entre malalts i individus controls. A la vegada, era necessari demostrar que ambdues mostres, de malalts i controls, tenien un mateix bagatge o *background* genètic. La presència d'estratificació en la població estudiada esbiaixaria els resultats i podria resultar en associacions espúries. Aquest plantejament teòric, a més, va estar condicionat pel fet que la doctoranda poc després d'iniciar aquest treball es va traslladar a treballar a ACLARA BioSciences Inc.

El present treball s'ha dividit en tres estudis i són els següents:

- i. Estudi 1. Identificació de variants genètiques en el gen *DDR1* en una mostra de malalts d'esquizofrènia. Treball realitzat a ACLARA BioSciences Inc.
- ii. Estudi 2: Anàlisi d'associació dels haplotips de la regió del gen *DDR1* en una mostra de malalts d'esquizofrènia i individus controls. Treball realitzat a ACLARA BioSciences Inc. Les anàlisis estadístiques es van fer en col.laboració amb l'Albert Ameijide (Fundació Privada IRCIS, Reus), el Dr. Francesc Calafell (Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona) i el Dr. Javier Costas (Fundación Pública Galega de Medicina Xenómica, Complejo Hospitalario Universitario de Santiago). En aquest estudi es van incloure 3 de les variants identificades en l'Estudi 1 i 2 SNPs de la dbSNP; Estudi de l'estratificació de la mostra. Treball realitzat conjuntament al Centre de Recerca Biomèdica de l'Hospital Universitari Sant Joan (HUSJ) de Reus i a ACLARA BioSciences Inc. Les anàlisis estadístiques es van fer en col.laboració amb l'Albert Ameijide.
- iii. Estudi 3: Desenvolupament i validació d'un nou mètode de genotipatge. Treball realitzat a ACLARA BioSciences Inc. Aquest mètode es va emprar pel genotipatge de 2 SNPs inclosos en l'estudi de l'estratificació de la mostra.

3.2. DESCRIPCIÓ DE LA MOSTRA

Aspectes ètics

Tots els procediments descrits en aquesta tesi han seguit les pautes de la Declaració de Helsinki i les Normes de Bona Pràctica Clínica. Cada un dels subprojectes va ser aprovat per la Comissió d'Investigació de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM) i pel Comitè Ètic d'Investigació Clínica de l'HUSJ. Els participants van ser informats de l'estudi i si acceptaven participar-hi signaven el full de Consentiment Informat. La custòdia del Consentiment Informat és responsabilitat de l'equip de recollida de la mostra. En el cas dels pacients, va ser l'equip clínic de l'HPUIPM i, pels participants controls, l'equip de la Unitat de Medicina Preventiva (Facultat de Medicina i Ciències de la Salut, URV).

Per a mantenir la confidencialitat respecte a la identitat dels participants es va fer servir un sistema de codificació de les mostres. A més, es va eliminar qualsevol variable identificadora quan es van traspasar les dades de la història clínica informatitzada o en paper a la base de dades de recerca. Les bases de dades que mantenen la relació d'identificació de cada malalt o individu control es guarden a l'HPUIPM i la Unitat de Medicina Preventiva (Facultat de Medicina i Ciències de la Salut, URV), respectivament. Aquestes són d'accés restringit als investigadors principals del projecte.

Grup de pacients d'esquizofrènia

Durant el període gener-1996/desembre-1998, es van seleccionar tots aquells pacients amb diagnòstics pertanyents al codi 295 (esquizofrènia) de la classificació ICD-9, en les unitats de mitja i llarga estada de l'HPUIPM. Un membre de l'equip d'investigació, el Dr. Joaquín Valero, es va posar en contacte amb l'equip assistencial a fi de tenir informació sobre la capacitat del pacient per a poder entendre la seva participació en l'estudi. En aquest primer pas, es van excloure pacients amb demència, analfabetisme o inestabilitat clínica. En presència d'un infermer, el Dr. Joaquín Valero va explicar a cadascun dels malalts la naturalesa de l'estudi i sol·licitava la seva participació. Un 10% dels malalts va refusar de participar-hi. Els 372 restants va accedir i així queda constància en el document de Consentiment Informat. D'aquests, se'n van escollir 297. La mida teòrica de la mostra havia de ser de més de 225 individus en base a la probabilitat d'estimar una diferència en les freqüències de l'al·lel mutant entre els grups a comparar amb un nivell de confiança del 95%, un poder estadístic del 80% i un risc mínim associat de 1.5 per a variants amb una freqüència al·lèlica al voltant de 10-15% (Epi Info versió 6). En la Taula X es mostren les característiques clíniques d'aquest grup.

Els diagnòstics i les variables clíniques considerades en aquest estudi (edat d'inici, tipus d'inici, curs i evolució de la malaltia i antecedents familiars de patologia psiquiàtrica) es van obtenir de la història clínica modular de l'hospital que està informatitzada.

Per dur a terme el rastreig de mutacions en el gen *DDR1* (Estudi 1), dels 297 individus, es van escollir els 100 malalts amb una edat d'inici més jove (Taula X). En aquest treball, l'edat d'inici es va definir com l'edat dels primers símptomes psicòtics prominents.

Taula X. Individus estudiats.

	Esquizofrènia (n=297)	Esquizofrènia (n=100)	Controls (n=298)
Relació homes/dones	176 (59.3%)/121 (40.7%)	53/47	158 (54.9%) /130 (45.1)
Edat^a			
Tots	57.5 ± 16	46.5 ± 16	43.7 ± 15
Homes	54.9 ± 15	52.9 ± 18	43.6 ± 14
Dones	61.6 ± 16	49.5 ± 17	43.8 ± 15
Edat d'inici^b			
Tots	24.4 ± 8.7	20.51 ± 5	
Homes	22.72 ± 7.8	19.9 ± 5	
Dones	26.9 ± 9.5	21.2 ± 4	
Subtipus esquizofrènia^c (Negativa/Positiva)			
Tots	119 (40.1%)/178 (59.9%)	40 (39.6%)/59 (58.4)	
Homes	72 (40.9%)/104 (59.1%)	24 (46.2%)/28 (53.8%)	
Dones	47 (38.4%)/74 (61.2%)	16 (34%)/31 (66%)	
Antecedent familiars^d			
Tots	75 (25.2%)	31 (31%)	
Homes	40 (13.5%)	18 (34%)	
Dones	35 (11.8%)	13 (28%)	

^a Edat en el moment en que es va reclutar al pacient, en concret del dia en què es va fer la extracció de sang; ^b Edat d'inici en la que apareixen els primers símptomes psicòtics (referit pel propi pacient, pels familiars o per l'entorn assistencial); ^c Subtipus d'esquizofrènia, d'acord amb la subclassificació DSM-III-R, agrupada en positiva (inclou esquizofrènia paranoide, episodi esquizofrènic i esquizofrènia esquizoafectiva) i negativa (inclou tots els altres subtipus); ^d Antecedents familiars d'esquizofrènia, psicosis delirants o trastorns afectius o depressions.

Grup control

Durant els anys 1999 i 2000 es va recollir una mostra d'individus aparentment sans per a emprar-los com controls en estudis d'associació. La iniciativa va ser de tres grups de recerca biomèdica vinculats a Facultat de Medicina de la URV (Unitat de Medicina Preventiva de l'URV, el Centre de Recerca Biomèdica de l'HUSJ i el Grup d'Investigació en Psiquiatria de l'HPUIPM) que varen comptar amb la col·laboració de l'ajuntament i de l'àrea bàsica de salut del municipi escollit al Camp de Tarragona. A partir del cens municipal es van escollir a l'atzar 1.200 individus d'acord amb les piràmides de població catalana (dades obtingudes del Instituto Nacional de Estadística). D'aquests, un total de 409 van participar en l'estudi de forma voluntària i van signar el full del Consentiment Informat. Per aquest estudi, es van incloure els participants no emparentats, sense antecedents personals per patologia psiquiàtrica major i amb una puntuació menor a 6 pel qüestionari Goldberg de salut mental (Goldberg i Williams, 1988).

D'aquest grup control es van escollir 298 individus (Taula X) per la raó esmentada pel grup de malalts.

La recollida d'un grup control planteja la pregunta de com aparellar els individus controls i casos. L'aparellament geogràfic és un dels criteris més importants a tenir en compte ja que proporciona un context o *background* genètic similar per ambdós grups. D'aquesta manera, es pot evitar el biaix introduït per marcadors genètics neutrals per la malaltia, però específics de la regió geogràfica. Per aquest motiu, tots els individus inclosos en l'estudi tenen nacionalitat espanyola i són d'origen caucàsic. Per ambdós grups, casos i controls, es van excloure aquells individus pertanyents a ètnies minoritàries (com la gitana o magribi) o amb un o dos progenitors de nacionalitat no espanyola, amb la finalitat d'obtenir una mostra amb un *background* genètic similar.

DNA genòmic control

Pel desenvolupament i optimització del mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs es van utilitzar les següents 12 mostres de DNA genòmic:

- i. Vuit mostres cedides pel Dr. Kwiatkowski (Third Wave Technologies, Inc). Aquestes havien estat caracteritzades prèviament per alguns dels SNPs d'interès.
- ii. Mostra K562 DNA High Molecular Weight (Promega, Madison, WI). Utilitzada freqüentment com a control positiu en anàlisis genètiques.
- iii. DNA genòmic obtingut de tres línies cel·lulars humanes del Coriell Institute for Medical Research (Camden, NJ).

DNA control de cadena senzilla

L'especificitat de les sondes al·lel específiques del mètode *eTag Multiplex Invader* es va determinar fent ús de DNA sintètic de cadena senzilla el qual contenia les seqüències complementàries a les sondes i oligonucleòtid Invader (Taula XI). Aquests oligonucleòtids van ser sintetitzats per Integrated DNA Technologies (Coralville, IA, EUA).

3.3. AÏLLAMENT I QUANTIFICACIÓ DE DNA GENÒMIC

Obtenció de DNA genòmic

Protocol d'aïllament del DNA genòmic a partir de sang recollida amb l'anticoagulant EDTA K₃

Les mostres de sang es van extreure en dejú. L'extracció de sang es va realitzar mitjançant una punció en la vena antecubital. Es va extreure un tub de 10 mL de sang recollida amb l'anticoagulant EDTA K₃ amb la finalitat d'obtenir una mostra de DNA.

Després de centrifugar la mostra a 2.500 rpm durant 15 minuts a temperatura ambient s'obtenen 3 fraccions diferents: 1. El plasma sobrenadant, se'n van recollir tres al·lquotes d'1 mL i es van congelar a -20°C, per possibles posteriors anàlisis bioquímiques; 2. La capa intermitja formada per les cèl·lules blanques, a partir de la qual se'n va extreure el DNA mitjançant el mètode de la precipitació per sals o *salting-out* (Miller *et al*, 1988) per les mostres del grup de malalts i amb reactius

comercials per a les mostres provinents del grup control; 3. La capa inferior de concentrat d'hematies que es va descartar.

Totes les mostres, de casos i controls, es van processar al Centre de Recerca Biomèdica de l'HUSJ, d'acord amb els procediments normalitzats de treball pertinent.

Taula XI. Disseny dels oligonucleòtids del mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs (Estudi 3).

SNP	C ^a	Oligonucleòtid	Seqüència (5'-3') ^{b,c}
APOE C/T Cys112Arg	S Sonda 112 Cys		ACACGTCCTCCATGTC
	Sonda 112 Arg		GCACGTCCTCCATGT
	Invader		CGGTACTGCACCAGGCGGCCGCT
	Control positiu		GCCCCGGCTGGGCGCGGACATGGAGGACGTG[T/C]GCGGCCGCTGGTGCAGTACCGCGGCCA GGTGCAGGC
APOE C/T Arg158Cys	S Sonda 158 Arg		ACTTC TGCAGGTCATCG
	Sonda 158 Cys		GCTTC TGCAGGTCATCG
	Invader		CCCCGGCCTGGTACACTGCCAGGCT
	Control positiu		CTCCTCCGCGATGCCGATGACCTGCAGAAG[T/C]GCCTGGCAGTGTACCAGGCCGGGGCCCCGG AGGGCGC
F5 A/G Gln596Arg	AS Sonda 506 Gln		AAGGAA TACAGGTATTTTGTC
	Sonda 506 Arg		GAGGAA TACAGGTATTTTGTC
	Invader		TCTAATCTGTAAGAGCAGATCCCTGGACAGACC
	Control positiu		TCAAGGACAAAATACCTGTATTCTCT[T/C]GCCTGTCCAGGGATCTGCTCTTACAGATTAGAAGTGA TTT
MTHFR T/C Val222Ala	S Sonda 222 Val		AC TCCCCGAGACAC
	Sonda 222 Ala		GC TCCCCGAGACAC
	Invader		CAAAGAAAAGCTGCGTGATGATGAAATCGC
	Control positiu		GAAGGTGTCTGCGGGAG[C/T]CGATTTCATCATCACGCAGCTTTTCTTTGAGG
MTHFR A/C Glu429Ala	S Sonda 429 Glu		GCTTCAC TGGTCAGCT
	Sonda 429 Ala		TCTTCAC TGGTCAGCT
	Invader		CCCCGAGAGGTAAGAACAAGACTTCAAAGACACTTA
	Control positiu		GGAGCTGACCAGTGAAG[A/C]AAGTGTCTTTGAAGTCTTTGTTCTTTACCTCTCGGGATT
F2 G/C G20210A	AS Sonda G 20210		GAGCC TCAATGCTCCC
	Sonda A 20210		AAGCC TCAATGCTCCC
	Invader		TATGGTTCCCAATAAAAGTGACTCTCAGCT
	Control positiu		TAGCACTGGGAGCATTGAGGCT[C/T]GCTGAGAGTCACTTTTATTGGGAACCATAGTTTTAGAA CACAAAAAT
ADRB2 A/G Arg16Gly	S Sonda 16 Arg		TAT TGGGTGCCAGCAA
	Sonda 16 Gly		CAT TGGGTGCCAGCAA
	Invader		TCGTGGTCCGGCGCATGGCTTCA
	Control positiu		CAGCGCCATTCTTGTGGCACCCAAT[A/G]GAAGCCATGCGCCGACCACGACGTACGCAGCA AAGGGACGAGGTGT
ADRB2 C/T Thr164Ile	AS Sonda 164 Ile		TCTCC TCTTGCCCAT
	Sonda 164 Thr		CCTCC TCTTGCCCAT
	Invader		GTGATCATTCTGATGGTGTGGATTGTGTCAGGCCTTAA
	Control positiu		GCATCTGAATGGCAAGAAGGAG[G/A]TAAGGCCTGACACAATCCACACCATCAGAATGATCAC CCGGG
ACAA1 C/T Glu255Glu	AS Sonda T 854		TTCCA TGGTGGTGCTG
	Sonda C 854		CTCCA TGGTGGTGCTA
	Invader		GGCAGGCTTCAGTTTGGCCAGGCCA
	Control positiu		AGGGTATCCGCCCCAGCACCATGGA[G/A]GGCCTGGCCAACTGAAGCCTGCCTTCAAGA
DDAH1 A/G rs761601	S Sonda al.llel T		TGGTA TGTGTGTCAGGCTC
	Sonda al.llel C		CGGTA TGTGTGTCAGGCTC
	Invader		CTCAGCCTTAAAAGACCTCCAGGCTTGATGCA
	Control positiu		AGCCTGACAACATACC[A/G]GCATCAAGCCCTGGAGGTCTTTTTAAGGCTGAGCCAATATAGCTA TGGATAACATTCTAA

^a C, cadena interrogada: AS, *antisense*, S, *sense*; ^b Per les sondes al.llel específiques, en negreta s'indica la posició polimòrfica i en cursiva la del nucleòtid biotinitat; ^c Per cada SNP, es va disposar de dos controls positius, un per cada al.llel. En parèntesis s'indica el punt polimòrfic.

Protocol d'aïllament del DNA genòmic de línies cel.lulars humanes derivades de limfòcits

Les línies cel.lulars humanes derivades de limfòcits AG12729, AG09714 i NG11755 es van obtenir del Coriell Institute for Medical Research (Camden, NJ). Aquestes es van fer créixer d'acord amb els corresponents protocols. El DNA genòmic es va extreure mitjançant la utilització del sistema comercial GeneElute Mammalian Genomic DNA Kit (Sigma-Aldrich, St Louis, MO, EUA).

Quantificació de DNA genòmic

Quantificació del DNA genòmic per espectrofotometria: absorbància a 260 nm

La concentració del DNA de totes les mostres de DNA genòmic es va determinar per absorbància a 260 nm (A260) i seguint la relació 1 OD/50 ng/μL. La qualitat del DNA es va estimar amb la relació d'absorbàncies A260/A280. Les mesures es van realitzar amb un espectrofotòmetre GeneQuantPro (APBiotech, Uppsala, Suècia). Per tal de facilitar les posteriors anàlisis, es van fer alíquotes de totes les mostres de DNA a una concentració de treball de 50 ng/μL que es guardaren a 4°C. Les solucions mare es van guardar a -20°C.

Quantificació del DNA genòmic per fluorometria: PicoGreen™ Kit

En la construcció dels pools o barreges de mostres de DNA és necessari un mètode de quantificació de DNA acurat, ja que hi ha d'haver la mateixa quantitat de cadascuna de les mostres en el pool. El sistema comercial PicoGreen Kit (Molecular Probes Inc., Eugene, OR, EUA) és un mètode molt sensible i específic per la quantificació de DNA de cadena doble (dsDNA). PicoGreen només emet fluorescència quan interacciona amb dsDNA i no interacciona amb DNA de cadena senzilla (ssDNA) o RNA. Aquest mètode es va emprar per la quantificació i construcció de *pools* de DNA de les 100 mostres incloses en l'Estudi 1.

Es va seguir el protocol indicat per Molecular Probes Inc. Les mesures es van realitzar a 485/535 nm (filtres d'excitació/emissió, respectivament) utilitzant un lector de plaques de 96 pous (fMax, Molecular Devices, Sunnyvale, CA, EUA). Es van utilitzar plaques de 96 pous negres i amb fons pla Microfluor 2 (Dynex Technologies, Chantilly, VA, EUA).

Cada mostra es va quantificar per triplicat. La mesura es va acceptar com a vàlida si la desviació estàndard era menor o igual al 10%. Si aquesta era major al 10%, la mostra es mesurava tres vegades més. La desviació estàndard d'aquesta segona mesura havia de ser menor o igual al 10% per acceptar el valor com a vàlid. Després de la quantificació, es van construir *pools* de 10 mostres a una concentració final de 2.5 ng/μL.

3.4. SEQÜENCIACIÓ AUTOMÀTICA DE LES REGIONS CODIFICANT I EXÓ-INTRÓ DEL GEN *DDR1*

La detecció de noves variants genètiques en el gen *DDR1* es realitzà per seqüenciació automàtica. Aquesta consta de les següents fases:

Amplificació mitjançant la tècnica de la reacció en cadena de la polimerasa (PCR) de les regions codificants i exó-intró del gen *DDR1*

La seqüència del gen *DDR1* es va obtenir de la base de dades GenBank (AC004211.1). Els *primers* o encebadors es van dissenyar amb el programa Primer3 (Rozen i Skaletsky, 2000) i es va comprovar la seva especificitat de seqüència mitjançant la utilització del programa BLAST (de l'anglès, *Basic Local Alignment Search Tool*). Es van dissenyar un total de 16 parelles d'encebadors, els quals s'uneixen a les regions intròniques de cada exó (Taula XII). Els encebadors van ser sintetitzats per Integrated DNA Technologies. Les condicions de PCR per cadascun dels fragments o amplicons, enumerades de A a D, foren les següents:

Protocol A. Fragment Ex1i2

La reacció es va dur a terme en un volum final de 50 µL que contenia 50 ng de DNA genòmic, 2 mM MgCl₂, 1x tampó de reacció de PCR (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 0.2 mM dNTPs (Applied Biosystems), 0.5 µM de cada encebador i 2 U AmpliTaqGold DNA polimerasa (Applied Biosystems). La reacció es va amplificar en un termociclador (MJ Research Tetrad, Waltham, MA, EUA) i el programa emprat va ser: 95°C per 10 min, 35 cicles de 94°C per 30 seg, 56°C per 20 seg i 72°C per 45 seg i una extensió final de 10 min a 72°C. Les mostres es van guardar a 4-10°C fins la purificació dels productes de la PCR.

Protocol B. Fragments Ex10 i Ex11

Protocol A, amb una concentració final d'1.5 mM MgCl₂.

Protocol C. Fragment Ex14

Protocol B, afegint 1 M betaina (Sigma-Aldrich, St Louis, MO, EUA). Betaina és un additiu de PCR, que com el DMSO o la formamida, millora l'eficiència de l'amplificació per PCR de seqüències amb estructura secundàries (Henke *et al*, 1997).

Protocol D. Fragments Ex3, 4, 5, 6, 7, 8, 9, 12, 13, 15, 16, 17

Protocol B però amb una T_m de 60°C.

La detecció i quantificació dels productes de PCR es va realitzar mitjançant electroforesi en microcapil·lars utilitzant el sistema comercial DNA 500 LabChip Kit i l'instrument Agilent 2100 BioAnalyzer (Agilent Technologies, Palo Alto, CA, EUA).

Purificació del producte de PCR

Per tal d'eliminar l'excés de deoxinucleòtids i encebadors abans de fer la reacció de seqüenciació, els productes de PCR es van incubar amb un còctel comercial

d'exonucleasa I i fosfatasa alcalina de gamba, ExoSAP-IT (USB Corporation, Cleveland, OH, EUA). A 5 µL de producte de PCR se li van afegir 2 µL d'ExoSAP-IT. Aquesta barreja es va incubar a 37°C per 20 min i 85°C per 20 min. Les mostres es van guardar a 4-10°C per un màxim de 8 hores abans de la reacció de seqüenciació.

Taula XII. Disseny dels oligonucleòtids emprats per amplificar per PCR les regions codificants i exó-intró del gen *DDR1* (Estudi 1).

Fragment	Encebadors ^a	Seqüència (5'-3')	L ^b	Comentaris
Ex1i2	F34623DDR1	ACTTCATTTGCCCCCTCAG	458	Exons 1 i 2 i regions exó-intró
	R35080DDR1	TTATCCCCTGCAGAGATGCT		
Ex3	F34241DDR1	TCGTCAGCGTATAGGTGGTG	452	Exó 3 i regions exó-intró
	R34692DDR1	CCTCTACTTCCCCTCCAACC		
Ex4	F32482DDR1	CTCATTGGCATGGAGGAGAC	418	Exó 4 i regions exó-intró
	R32899DDR1	CACCCCTCATGGGTCTCTAA		
Ex5	F32240DDR1	ACCCCAAACCTCCATATCCTG	339	Exó 5 i regions exó-intró
	R32578DDR1	GAGTGTGGAGAATGGGCATC		
Ex6	F31478DDR1	CACAGGGTCAAGGCTAGGAG	428	Exó 6 i regions exó-intró
	R31905DDR1	CAGCTGCATGAGTGTGAGGT		
Ex7	F31086DDR1	AGAAGTTTCGGTGGGTGATG	412	Exó 7 i regions exó-intró
	R31497DDR1	CTCCTAGCCTTGACCCTGTG		
Ex8	F30496DDR1	AGGGTCAGCACACAGGA	320	Exó 8 i regions exó-intró
	R30815DDR1	CACTGGTCAGTGGTTGCCTA		
Ex9	F30172DDR1	GTGAGCCCCATACCCTCTTT	393	Exó 9 i regions exó-intró
	R30564DDR1	GACCCTCTGCACCCTTCTC		
Ex10	F28956DDR1	CAGGGGCACAGAAAGAAGAA	403	Exó 10 i regions exó-intró
	R29358DDR1	GCTGTAGGGCTCTTGTGAGG		
Ex11	F28098DDR1	CACAAGAGCTTCTCCTTGG	401	Exó 11 i regions exó-intró
	R28498DDR1	CGTCACTCTGCAGATCCTTG		
Ex12	F26855DDR1	TAGGGCCAGACTGGGTAG	449	Exó 12 i regions exó-intró
	R27303DDR1	TCTCCTGGTTTGAGGTTGG		
Ex13	F26490DDR1	CATGTAATCCCCTCCAGA	442	Exó 13 i regions exó-intró
	R26931DDR1	GACTCCGCTTCAAGGAGAAG		
Ex14	F26561DDR1	TGGGGAGCCATCTAGAGAGA	568	Exó 14 i regions exó-intró
	R25994DDR1	CCAGCCTTCTGTCCATGTT		
Ex15	F25269DDR1	ACACTGCAGCTCCTGCTTTC	490	Exó 15 i regions exó-intró
	R25758DDR1	GGAGTGGGCTCTCTCTCCTC		
Ex16	F24589DDR1	CAGGTACACCTGCATTGTGG	447	Exó 16 i regions exó-intró
	R25035DDR1	GGGAGAGGAGTTGGAAAAGG		
Ex17	F24268DDR1	GTGTCCAGGAAGGGGAGAAG	452	Exó 17 i regions exó-intró
	R24719DDR1	GCAGGTCAGAGTGGAGGAGA		

^aF indica encebador *forward*, R encebador *reverse* i el número indica la posició del primer respecte la seqüència AC004211.1; ^bLongitud en parelles de bases del fragment de PCR.

Reacció de seqüenciació i separació per electroforesi capil·lar

Els encebadors utilitzats van ser els mateixos que els emprats en la reacció de PCR. Es va fer servir el sistema comercial de seqüenciació DYEnamic ET Dye Terminator Cycle Sequencing Kit (APBiotech), i els productes de la seqüenciació es van purificar mitjançant la precipitació per etanol d'acord amb els protocols del *kit*. Els productes purificats de seqüenciació es van separar amb l'aparell d'electroforesi capil·lar MegaBACE500 (APBiotech) seguint les instruccions del fabricant. La qualitat de les seqüències es va avaluar amb el programa MegaBACE Sequencing Scorecard d'acord al *base caller* Cimarron 2.19.12.

Les dues cadenes de DNA de cada fragment de PCR van ser seqüenciades per duplicat. Les reaccions de seqüenciació que no passaren els criteris de qualitat (10%), es van tornar a repetir per duplicat.

Anàlisi de les seqüències

Els cromatogrames es van analitzar amb el programa comercial Sequencher™ (GeneCodes Corporation, Ann Arbor, MI, EUA).

Les posicions de les variants genètiques identificades es van enumerar d'acord amb la seqüència de la isoforma DDR1b dipositada a la base de dades Vega amb identificació OTTHUMG00000016356 (http://vega.sanger.ac.uk/Homo_sapiens/geneview?gene=OTTHUMG00000016356&db=core).

3.5. CLONATGE DE FRAGMENTS DE PCR EN UN VECTOR PLASMIDI

Per comprovar una de les variants genètiques identificades per seqüenciació en el fragment Ex4 (deleció IVS3 -18delCT), es van clonar els fragments de PCR de dues mostres diferents. Els productes de PCR es van purificar abans de ser clonats amb el sistema comercial QIAquick PCR purification Kit (Qiagen, Valencia, CA, EUA). La clonació es va fer emprant el sistema comercial pGEM®-T Easy Vector System II (Promega) seguint el protocol recomanat, excepte que la ligació del producte de PCR i el vector va realitzar-se a 14°C i durant 10 hores.

Per cada fragment de PCR, es van seleccionar 6 colònies transformants i d'aquestes es va aïllar i seqüenciar l'insert per tal d'identificar la variant genètica. Les colònies seleccionades es van fer créixer seguint el protocol recomanat. La purificació del plasmidi es va fer utilitzant el sistema comercial QIAprep Spin Miniprep Kit (Qiagen), d'acord amb els protocols del *kit*. La restricció i aïllament de l'insert es va fer per digestió amb l'enzim de restricció EcoRI (New England Biolabs, Beverly, MA, EUA) seguint el protocol indicat. Els inserts es van purificar mitjançant el mètode QIAquick PCR purification Kit (Qiagen) i es van seqüenciar amb els primers de PCR corresponents, d'acord amb el protocol de seqüenciació mencionat anteriorment.

3.6. MARCADORS I GENOTIPATGE MITJANÇANT EL MÈTODE PCR-RFLP (POLYMERASE CHAIN REACTION – RESTRICTION FRAGMENT LENGTH POLYMORPHISM)

El protocol de genotipatge pel mètode de PCR-RFLP desenvolupat en aquest treball consta de 4 fases i són les següents: *i.* Amplificació del fragment de DNA que conté el SNP d'interès per PCR; *ii.* Discriminació per digestió amb l'enzim de restricció; *iii.* Separació de la reacció de restricció per electroforesi en gels d'agarosa o poliacrilamida; *iv.* Lectura i anotació dels resultats als fulls de treball a fi de què quedi constància i que els resultats puguin ser revisats. Posteriorment, aquests s'introdueixen a la base de dades corresponent.

A continuació es descriuen els protocols emprats en els Estudis 2 i 3.

Estudi 2: Anàlisi de l'associació del gen *DDR1* amb l'esquizofrènia

Donat que l'anàlisi de tots els polimorfismes identificats en l'Estudi 1 no va ser possible per motius econòmics, es van estudiar inicialment 7 SNPs, cobrint una extensió d'unes 20 kb (Taula XIII). D'aquests, 2 van resultar ser no polimòrfics. Els 4 SNPs de la regió del gen *DDR1* inclosos en l'estudi d'associació van ser:

- Els dos SNPs identificats en l'Estudi 1 i que comporten un canvi d'aminoàcid (tots dos es localitzen a l'exó 10) i un SNP silenciós a l'exó 3 (rs2229933).
- El quart es va extreure de la dbSNP i es localitza en l'extrem 3'UTR (rs8408).

A més, es va incloure un cinquè SNP de la regió 5' del gen *GTF2H4* (localitzat a unes 10 Kb *downstream* de *DDR1*). La descripció dels SNPs, seqüències dels encebadors de PCR, característiques del fragment amplificat i enzim de restricció utilitzat es recullen en la Taula XIII.

El mateix protocol de PCR es va utilitzar per l'anàlisi dels 7 SNPs. La reacció es va dur a terme seguint les següents condicions: 25 ng de DNA genòmic, 2 mM MgCl₂, 1x tampó de reacció de PCR (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 0.2 mM dNTPs (Applied Biosystems), 0.5 µM de cada primer, 0.04 U/µL AmpliTaqGold DNA polimerasa (Applied Biosystems) en un volum final de 10 µL. La reacció es va amplificar en un termociclador MJ Research Tetrad i el programa emprat va ser: 95°C per 10 min, 35 cicles de 94°C per 30 seg, 61°C per 20 seg i 72°C per 45 seg i una extensió final de 10 min a 72°C. Les mostres es van guardar a 4-10°C fins la restricció enzimàtica. Cinc µL de la PCR es van digerir amb 3 U del corresponent enzim de restricció, d'acord amb el protocol especificat pel proveïdor de l'enzim (New England BioLabs). Les restriccions es van separar per electroforesi en gels d'agarosa al 2 o 4% amb bromur d'etidi incorporat (Invitrogen, Carlsbad, CA, EUA).

Com a control de qualitat, totes les mostres van ser genotipades pels SNPs de l'exó 3 (rs2229933) i regió 3'UTR (rs8408) a Reus per la tècnica PCR-RFLP. No es van trobar discrepàncies entre els resultats obtinguts a ACLARA BioSciences Inc. i al laboratori del Centre de Recerca Biomèdica.

Taula XIII. SNPs explorats en l'estudi d'associació de casos i controls (Estudi 2).

SNP	Primers ^a	Seqüència (5'-3')	L ^b	Enzim de restricció ^c
DDR1-exó 3 Leu 94 CTG/CTC rs2229933 ^d	F34241DDR1 R34692DDR1	TCGTCAGCGTATAGGTGGTG CCTCTACTTCCCCTCCAACC	452	Hinf I al.lel G: 452 bp; al.lel C: 229+223 bp
DDR1-exó 10 Ser 496 TCC/TCT rs1264319	F29202DDR1 R29051DDR1	GTCCTGGGGACACTATCCT AGAGGTGCGACTGGAACAAG	152	BsrB I al.lel C: 100+52 bp; al.lel T: 152 bp
DDR1-exó 10 Asn 502 Ser AAT/ AGT	F29202DDR1 R29051DDR1	GTCCTGGGGACACTATCCT AGAGGTGCGACTGGAACAAG	152	Tsrp I al.lel C: 115+37 bp; al.lel A: 152 bp
DDR1-intró 17 C/T -18357 rs1049628	F24268DDR1 R24719DDR1	GTGTCCAGGAAGGGGAGAAG GCAGGTCAGAGTGGAGGAGA	452	BspM I al.lel G: 6+118+282+52 bp; al.lel A: 6+394+52 bp
DDR1-3'UTR A/G - 18778 rs8408	F24163DDR1 R23647DDR1	GGAAGGGTGGGGAGAAATATAG ATTGCCTCAAGCTAGGTCCA	517	Pst I al.lel C: 304+223 bp; al.lel T: 527 bp
GTF2H4 G/A rs1052693	F15268DDr1 R15461DDr1	GCCTTCCGATCTCCTAGCTT CTCAATCTCCAGGAGCCAAT	194	Hinf I al.lel G: 136+58 bp ; al.lel A: 194 bp
GTF2H4 A/G rs2230121	F11466DDR1 R11166DDR1	CTCACACCCAAAGCCTTCAT CTATCCCTGGCTCAGAGTCG	301	Mfe I al.lel A: 181+120 bp; al.lel G: 301 bp

^aEl número indica la posició del primer respecte la seqüència AC004211.1; ^bLongitud en parelles de bases del fragment de PCR; ^cIndica la longitud dels fragments digerits per l'enzim corresponent per cadascun dels al.lels; ^dNúmero identificador de l'SNP en la dbSNP.

Estudi 2: Estudi de l'estratificació de la mostra

Es van escollir un total de 25 SNPs localitzats en diferents cromosomes autosòmics, per assegurar la seva independència (no lligament) seguint les recomanacions de Pritchard i col.laboradors (Pritchard *et al*, 2000). La descripció dels SNPs, seqüències dels encebadors de PCR, característiques del fragment amplificat, enzim de restricció i condicions d'electroforesi es recullen en la Taula XIV.

Dels 25 SNPs, 20 van ser analitzats al Centre de Recerca Biomèdica per PCR-RFLP i 5 a ACLARA BioSciences Inc. D'aquests últims, 3 es van analitzar pel mètode PCR-RFLP (SNPs ESR2 A1730G, SERPINB2 Asn120Asp i IL-B1 C-514T) i 2 pel mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs (SNPs F5-Leiden, ADRB2 Arg16Gly). Pels SNPs IL-B1 C-511T i ESR2 A1730G es van seguir els protocols descrits per Katila *et al*, 1999 i Kealey *et al*, 2001, respectivament. Per SERPINB2 Asn120Asp, es va seguir el mateix protocol de PCR que en l'Estudi 2. Les restriccions es van separar per electroforesi en gels d'agarosa 4% amb bromur d'etidi incorporat (Invitrogen).

Taula XIV. SNPs estudiats en l'estudi de l'estratificació de la mostra (Estudi 2).

Gen SNP	C ^a	Seqüència (5'-3') (forward i reverse)	L ^b	Enzim restricció	Referència
MTHFR Ala222Val	1p36	AGGACGGTGCGGTGAGAGTG TGAAGGAGAAGGTGTCTGCGGGA	196	Hinf I	Frost <i>et al</i> , 1995
F5-Leiden	1q23	GGAAGGTTACTTCAAGGACAAAATACCT CAGGGGAAACCTATACTTATAAGTGAACATC	200	Mnl I	Bertina <i>et al</i> , 1994
IL1B -511C>T	2q14	TGGCATTGATCTGGTTCATC GTTTAGGAATCTTCCCCTT	304	Ava I	Katila <i>et al</i> , 1999
PPAR γ Pro12Ala	3p25	GCCAATTCAAGCCCAGTC GATATGTTTGACAGACAGTGTATCAGTGAAGGATCGCTTCCG	270	BstU I	Yen <i>et al</i> , 1997
TF Pro570Ser	3q21	GCTGTGCCTGATGGTACCAGGTAA GGACGCAAGCTTCCCTTATCT	110	BstE II	Namekata <i>et al</i> , 1997
FABP-2 Ala54Thr	4q28-q31	ACAGGTGGTAATATAGTGAAAAG TACCTGAGTTTCAGTTCCGTC	180	Hha I	Chiu <i>et al</i> , 2001
ADRB2 Arg16Gly	5q32-q34	ATGGGGCAACCCGGGAACGG ATGGCCAGGACGATGAGAGAC	131	Bbv I	Turki <i>et al</i> , 1995
HFE Cys282Tyr	6p21.3	ACATGGTTAAGGCCTGTTGC GCCACATCTGGCTTGAAATT	208	Mbo I	Jouanolle <i>et al</i> , 1997
ESR1	6q25.1	CTGCCACCCTATCTGTATCTTTTCTATTCTCC TCTTTCTGCCCACCCTGGCGTCGATTATCTGA	1371	Pvu II	Kobayashi <i>et al</i> , 1996
IL16 -285T>C	7p21	CTCCACTCAAAGCCTTTTGTTCCTATGA CCATGTCAAAAACGGTAGCCTCAAGC	280	AhdI	Nakayama <i>et al</i> , 2000
MDR1 3435C>T	7p21.1	TGTTTTAGCTGCTTGTATGG GCATGTATGTTGGCTCCTT	193	Mbo I	Hoffmeyer <i>et al</i> , 2000
LPL Ser2447Thr	8p22	CATCCATTTTCTCCACAGGG TAGCCAGAATGCTCACCAGACT	137	Hinf I	Fisher <i>et al</i> , 1997
RXRA -25A>G	9q34.3	CAGTGCCAGGGCAGAACT CTGAGAAGAACAGCTGGCGT	279	Nci I	Hegele <i>et al</i> , 2001
ADRB1 Gly389Arg	10q24-q26	GGGCTACGCCAACTCGGCCTTCAACCCCATC GCCGGCGGGCAGCCCTGCGCG	108	Bst NI	Dionne <i>et al</i> , 2002
APOA5-A4 IntergenicT>C	11q23	GTGCCTGTCACCACCGTTTGG ATGCATTAGCCTCTGCTGTTT	162	Bst NI	Talmud <i>et al</i> . 2002
A2M Val1000Ile	12p13.3- p12.3	CTGCTTAATGACTTTGATAGAT ACTGAAACCTACTGGGAAAT	312	Mbo I	Comunicació personal del Dr. Perez-Tur
BRCA2 203G>A	13q12.3	CTGTTTTCAGACTTATTACCAT ACACTGTGACGTACTGGGTTTT	206	Nsi I	Breast Cancer Information Core
ESR2 1730A>G	14q	GACCTGCTGCTGGAGATGCT AATGAGGGACCACACAGCA	234	Alu I	Kaeley <i>et al</i> , 2001
SULT1A1 Arg213His	16p12.1	GGTTGAGGAGTTGGCTCTGC ATGAACCTCTGGGGACGGT	281	Hae II	Engelke <i>et al</i> , 2000
MCP-1 -2518G>A	17q11.2- q12	TCTCTACGCCAGCACTGACC GAGTGTTACATAGGCTTCTG	234	Pvu II	Rovin <i>et al</i> , 1999
SERPINB2 Asn120Asp	18q21.3	CGCAGACTTCTACCAACA GGCCTTCTCTAGGCTGT	368	Tsp509 I	rs6098 ^c
APOE Cys112 Arg Arg 158 Cys	19q13.2	ACAGAATTGCCCCGGCCTGGTACAC TAAGCTTGGCAGGCTGTCCAAGGA	274	Hha I	Hixon i Vernier, 1990
PRNP Met129Val	20pter-p12	TTTTGCAGAGCAGTCATTATG CGTGTGCTGCTTATTGTG	480	Nsp I	Tsai <i>et al</i> , 2001
CBS 1080C>T	21q22.3	GTGGCAGTCTGGCAGCACG ATGTAGTCCGCACTGAGTC	106	BstU I	de Stefano <i>et al</i> , 1998
PPAR α Leu162Val	22q13.31	GACTCAAGCTGGTGTATGACAAGT CGTTGTGTGACATCCCAGAGAAT	117	Hinf I	Vohl <i>et a</i> , 2000

^aC, localització cromosòmica; ^bLongitud en parelles de bases del fragment de PCR; ^cNúmero identificador del SNP en la dbSNP.

Estudi 3: Validació del mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs

Donada la simplicitat i el relativament baix cost del mètode PCR-RFLP, aquest es va emprar com a mètode de referència per comparar els genotips obtinguts amb el nou protocol *eTag Multiplex Invader* per l'anàlisi d'SNPs. La descripció dels 10 SNPs

evaluats, seqüències dels encebadors de PCR, característiques del fragment amplificat i enzim de restricció es recullen en la Taula XV.

El protocol de PCR emprat en l'Estudi 2 es va seguir per l'amplificació de 8 dels SNPs, a excepció de l'amplificació del fragment del gen *APOE* (SNPs *APOE* 112, *APOE* 158). En aquest cas, es va afegir DMSO (Sigma-Aldrich) a una concentració final del 10%. Les restriccions es van separar per electroforesi en gels de poliacrilamida al 10% (Invitrogen). La visualització dels resultats es va fer per tenyiment amb SybrGold® (Molecular Probes).

Taula XV. SNPs inclosos en l'estudi de validació del mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs (Estudi 3).

SNP	Seqüència primer (5'-3') (forward i reverse)	L ^a	Enzim restricció	Referència
<i>APOE</i> 112 ^b Cys112Arg <i>APOE</i> 158 Arg158Cys	ACAGAATTCGCCCCGGCCTGGTACAC TAAGCTTGGCACGGCTGTCCAAGGA	274	Hha I	Hixson i Vernier, 1990
F5 Leiden Arg506Gln	GGAAGGTTACTTCAAGGACAAAATACCT CAGGGGAAACCTATACTTATAAGTGGAAACATC	200	Mnl I	Bertina <i>et al</i> , 1994
MTHFR C677T Ala222Val	AGGACGGTGC GG T GAGAGTG TGAAGGAGAAGGTGTCTGCGGGA	196	Hinf I	Frosst <i>et al</i> , 1995
ADRB2 1 Arg16Gly	ATGGGGCAACCCGGAACGG ATGGCCAGGACGATGAGAGAC	131	Bbv I	Turki <i>et al</i> , 1995
F2 G20210A 3'UTR	GCACAGACGGCTGTTCTCTT ATAGCACTGGGAGCATTGAAGC	506	Hind III	Poort <i>et al</i> , 1996
MTHFR A1298C Glu429Ala	CTTTGGGGAGCTGAAGGACTACTAC CACTTTGTGACCATTCCGGTTTG	163	Mbo II	Weisberg <i>et al</i> , 1999
ADRB2 2 Thr164Ile	TTACTTCACCTTTCAAGTACCAGAGC CATAGCAGTTGATGGCTTCCTG	154	Mnl I	Busjahn <i>et al</i> , 2000
ACAA1 Glu255Glu	CACCAGCTGTGGTAGAACCAT AGCTGAGATTGTGCCTGTGA	161	Dra II	rs2011003 ^c
DDAH1 Intrònic	TGCCCAAGTACACTCTTCATTG CCCTCTCAGGCAGTGTTCAT	226	BsrF 1	rs761601

^aLongitud en parelles de bases del fragment de PCR; ^bEls SNPs *APOE* 112 i *APOE* 158 s'amplifiquen en el mateix fragment de PCR i poden ser discriminats fent ús d'un únic enzim de restricció; ^cNúmero identificador de l'SNP en la dbSNP.

3.7. MARCADORS I GENOTIPATGE MITJANÇANT EL MÈTODE *eTAG MULTIPLEX INVADER* PER L'ANÀLISI D'SNPS

Un total de 10 SNPs es van analitzar simultàniament mitjançant el mètode *eTag Multiplex Invader*. Aquest protocol consta de 4 fases: i. Coamplificació per PCR dels fragments de DNA que contenen els SNPs d'interès; ii. Discriminació al.lèlica pel mètode *eTag Multiplex Invader*; iii. Separació de la reacció de discriminació al.lèlica per EC; iv. Anàlisi dels electroforogrames mitjançant el programa *eTag Informer™*

(ACLARA BioSciences Inc.). Els genotips resultants es van entrar en la base de dades corresponent.

Coamplificació de 10 fragments de DNA per PCR

La descripció dels SNPs inclosos, seqüències dels encebadors de PCR i característiques dels fragments amplificats es recullen en la Taula XV.

La composició de la PCR va ser: 15 ng de DNA genòmic, 2 mM MgCl₂, 1x tampó de reacció de PCR (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 0.2 mM dNTPs (Applied Biosystems), 0.3 µM de cadascun dels 22 encebadors, 5% DMSO i 0.04 U/µL AmpliTaqGold DNA polimerasa (Applied Biosystems) en un volum final de 2.5 µL. La reacció es va amplificar en un termociclador MJ Research Tetrad i el programa emprat va ser: 95°C per 10 min, 25 cicles de 94°C per 30 seg, 57°C per 20 seg i 72°C per 45 seg i una extensió final de 10 min a 72°C. Les mostres es van guardar a 4-10°C fins la reacció de discriminació al.lèlica.

Discriminació al.lèlica pel mètode *eTag Multiplex Invader*

Les seqüències dels oligonucleòtids al.lèl específics i Invader es recullen en la Taula XI. Totes es van dissenyar fent ús del programa Invader Creator perquè tinguessin unes T_m teòriques de 77°C i 63°C per l'oligonucleòtid Invader i sondes al.lèl específiques, respectivament. Cada sonda al.lèl específica es va marcar en el seu extrem 5' amb un *eTag reporter* concret i amb una molècula de biotina en el nucleòtid timina més proper a l'extrem 5' però, que no fos aquest ni el segon ni el tercer nucleòtid començant per aquest extrem. Havíem observat prèviament que la presència de biotina en aquestes posicions afecta l'eficiència del reconeixement i tall de l'*eTag reporter*. El marcatge amb biotina permet capturar l'excés de sonda intacta (o no tallada) amb estreptavidina al final de la reacció i, prèviament a la separació per EC. D'aquesta manera, només els *eTag reporters* en solució són separats i detectats per EC. Totes les sondes al.lèl específiques es van sintetitzar a ACLARA BioSciences, Inc. Els oligonucleòtids Invader van ser sintetitzats per Integrated DNA Technologies.

A l'anterior reacció de PCR se li va afegir 1 µL del tampó de reacció E (50 mM MOPS (Sigma-Aldrich), 16% PEG (Sigma-Aldrich), 37.5 mM Cl₂Mg (Sigma-Aldrich)), 0.08 µL dels 11 oligonucleòtids Invader 5 µM (barreja equimolar), 0.295 µL dels 21 oligonucleòtids al.lèl específic 3.5 µM (barreja equimolar), 0.2 µL de l'enzim Cleavase X a 100 ng/µL (Third Wave Technologies Inc.) i 2.425 µL d'aigua estèril i lliure de nucleases. Aquesta barreja es va incubar en un termociclador MJ Research Tetrad a 94°C per 1 min 30 s, 63°C per 30 min. Les mostres es van guardar a 4-10°C fins la separació dels *eTag reporters* per EC.

En cada placa de 96, es van incloure 2 controls negatius (no DNA o mostra) i 2 controls positius (10⁷ còpies de les seqüències control per cadascun dels diferents al.lèls analitzats). El control negatiu informava sobre el soroll de fons de la tècnica; el

control positiu confirmava la presència de totes les sondes al.llel específiques incloses en la reacció.

Separació de la reacció *eTag Multiplex Invader* per electroforesi capil·lar

La separació per EC dels *eTag reporter* es va realitzar fent ús de l'instrument MegaBACE1000 (APBiotech). Abans de la separació, a cadascuna de les mostres, se'ls hi va afegir 10 µL de la solució de separació electroforètica (estreptavidina a 1 mg/mL (Molecular Probes), eTag1-marcador electroforètic a 18 nM, eTag2-marcador electroforètic a 8 nM dissolts en aigua). Les mostres es van guardar a 4-10°C fins la separació electroforètica dels *eTag reporters*.

La configuració de l'instrument MegaBACE1000 i paràmetres electroforètics van ser els mateixos que els emprats en l'aplicació de seqüenciació. Les mostres es van injectar durant 15 segons a 15 kV i separar durant 30 min a 15 kV. En aquest aparell, la detecció dels *eTag reporter* va tenir lloc per *Laser-Induced Fluorescence* (LIF), amb una excitació i captació de l'emissió de fluorescència a 488 nm i 520 nm, respectivament.

Anàlisi dels resultats: programa eTag Informer

Els electroforogrames es van analitzar amb el programa eTag Informer (ACLARA BioSciences Inc.). Aquest està dissenyat per la identificació i quantificació dels pics electroforètics corresponents als *eTag reporters*, cadascun dels quals codifica per un al.llel diferent. La variabilitat deguda a la injecció per EC fou controlada normalitzant l'altura del pic de cada *eTag reporter* amb un dels marcadors electroforètics. Aquesta variable es va designar com alçada relativa del pic electroforètic (ARPE). El criteri per acceptar l'anotació dels *eTag reporter* fou que la relació senyal:soroll de fons fos superior a 10:1. Per cada *eTag reporter*, aquesta es va calcular com el valor ARPE de la mostra respecte a la del control negatiu.

A continuació, a cada *eTag reporter* identificat se li va sostroure el valor ARPE del control negatiu, corresponent al soroll de fons o senyal inespecífic. Aquestes dades van ser transferides a un full de càlcul de Microsoft® Excel2000 on es va calcular la relació entre al.llels per cadascun dels SNPs. Per un determinat SNP amb els al.llels A i a, la relació al.lèlica es va calcular d'acord a,

$$\text{Relació al.lèlica} = \frac{\text{ARPE eTag1 al.llel A}}{\text{ARPE eTag1 al.llel A} + \text{ARPE eTag2 al.llel a}}$$

Segons aquesta relació, els valors s'agruparien al voltant dels valors 1, 0.5 i 0 per mostres homozigotes per l'al.llel A, heterozigotes i homozigotes per l'al.llel a, respectivament.

3.8. ANÀLISIS DE LES DADES

Anàlisis estadístiques

Els programes d'anàlisis estadístiques utilitzats foren:

SPPS/PC+ Versió 11 (Chicago, IL, EUA)

Programa comercial emprat per les següents anàlisis estadístiques:

Test χ^2 . Test estadístic utilitzat per comparar si una distribució observada correspon a una distribució teòrica. S'ha utilitzat per la comparació de freqüències genotípiques entre els grups dels malalts i controls. I per la comprovació de l'acompliment de la llei de Hardy-Weinberg per la distribució de freqüències al·lèliques i genotípiques en els grups estudiats.

Anàlisi de regressió. Test estadístic que permet descriure quina relació existeix entre dues variables, comprovar la hipòtesi d'independència de les variables, avaluar la intensitat de l'associació entre les dues variables i efectuar prediccions. Es va utilitzar per analitzar l'associació entre l'edat d'inici de la malaltia i els genotips.

Arlequin versió 2.00 (Schneider *et al*, 2000)

Es pot trobar a la direcció URL: <http://lgb.unige.ch/arlequin/>

Programa que permet realitzar anàlisis específiques de l'estudi de genètica de poblacions. En concret, es va emprar en l'Estudi 2 per:

- Estimació de les freqüències haplotípiques mitjançant l'algoritme *expectation-maximization* (E-M).
- Valoració del desequilibri de lligament (LD) entre parelles de *loci* mitjançant el coeficient de desequilibri D' i r^2 .
- Valoració de l'equilibri de Hardy-Weinberg (H-W). La presència d'equilibri genotípic en la mostra control fou un requisit indispensable per a l'anàlisi de la mostra.

Phase versió 2.0.2

Es pot trobar a la direcció URL: <http://www.stat.washington.edu/stephens>

Programa que permet l'estimació de les freqüències haplotípiques mitjançant l'algoritme bayesià Phase.

Structure versió 1.00 (Pritchard *et al*, 2000)

Es pot trobar a la direcció URL: <http://pritch.bsd.uchicago.edu>

Programa emprat per investigar la presència d'estratificació de la mostra a partir de les dades de múltiples *loci* (Estudi 2). L'algoritme Markov Chain Monte Carlo implementat en aquest programa calcula la probabilitat de què les dades s'ajustin a un número predeterminat de subpoblacions (K) i estima per cada individu la probabilitat de pertànyer a k. En aquest estudi, les anàlisis es van fer assumint els models de mescla i no mescla de poblacions (*admixture/no admixture models*) i considerant la presència de 1, 2, 3 o 4 subpoblacions (k).

Anàlisi de seqüències

Els programes d'anàlisi de seqüències utilitzats van ser els següents:

Primer3 (Rozen i Skaletsky, 2000)

Es pot trobar a la direcció URL: http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

Programa emprat pel disseny de primers de PCR.

BLAST (*Basic Local Alignment Search Tool*)

Es pot trobar a la direcció URL: <http://www.ncbi.nlm.nih.gov/BLAST/>

Programa emprat per la cerca de similituds en seqüències de DNA i proteïnes.

Sequencher™ (GeneCodes Corporation, Ann Arbor, MI, EUA)

Programa comercial emprat per l'alineament i detecció de discrepàncies en cromatogrames.

WebCutter2.0

Es pot trobar a la direcció URL: <http://www.firstmarket.com/cutter/cut2.html>

Programa emprat per la identificació d'enzims de restricció capaços de discriminar SNPs per poder ser analitzats pel mètode PCR-RFLP.

InvaderCreator® (Third Wave Technologies)

Es pot trobar a la direcció URL: <http://www.invadercreator.com>

Programa comercial que permet el disseny simultani dels oligonucleòtids al·lels específics i *Invader*. Aquests també es poden dissenyar manualment fent ús del programa Primer3 o similar.