

## ESTUDI 2: Anàlisi de l'associació del gen *DDR1* amb l'esquizofrènia: estudi de casos i controls

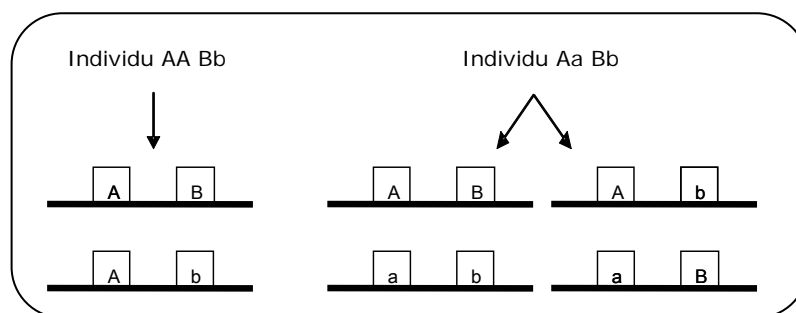
### INTRODUCCIÓ

Els estudis d'associació de casos i controls utilitzen individus no emparentats com a controls que provenen de la mateixa població que els individus afectes del fenotip d'estudi (casos). Aquest disseny és, certament, més convenient pel reclutament d'individus i facilita l'obtenció d'un major número de mostres en comparació amb els estudis basats en famílies. A més, permet l'estudi de la freqüència al·lèlica, penetrança i risc atribuïble de la variant genètica en la població estudiada.

Els aspectes limitants més importants a tenir en compte en aquest disseny experimental són el possible biaix introduït per l'estratificació de la població, i la incapacitat de determinar els haplotips individuals sense el tipatge de familiars o bé, sense la utilització de mètodes d'haplotipatge extremadament costosos. Si ambdós aspectes són superats, l'estudi de casos i controls és un disseny experimental molt atractiu per la identificació i caracterització de variants gèniques de susceptibilitat en malalties complexes.

Actualment, l'aproximació més emprada per determinar els haplotips és mitjançant el genotipatge de familiars i inferir l'haplotip en base a la segregació Mendeliana dels al·lèls en la família estudiada. La dificultat d'aquesta aproximació és el reclutament de familiars i el fet de què no sempre són informatius respecte la transmissió (només els pares heterozigots són inclosos), (Morton i Collins, 1998). D'altra banda, les tècniques per establir haplotips de mostres individuals són molt costoses, tant en temps com en diners, i inclouen la separació de cromosomes o de regions cromosòmiques haploides (Douglas *et al*, 2001) o bé, fer PCR al·lèl específiques de fragments grans (Michalatos-Beloin *et al*, 1996; Clark *et al*, 1998). Per aquests motius, quan no es disposa de nuclis familiars, s'ha de recórrer a l'estimació probabilística de les freqüències dels haplotips. En aquest cas, assumint equilibri de Hardy-Weinberg a tots els *loci* estudiats en una població, es disposa de diferents mètodes per l'estima de les freqüències haplotípiques: mètode parsimònia (Clark, 1990), l'algoritme E-M (*Expectation-Maximization*) (Excoffier i Slatkin, 1995; Hawley i Kidd, 1995) i l'algoritme bayesià Phase (Stephens *et al*, 2001).

L'assignació dels haplotips d'un individu pot fer-se de forma inequívoca si tots els genotips dels loci inclosos en l'haplotip són homozigots o bé només un dels locus és heterozigot. Per exemple, en la Figura 15, a l'individu AA Bb li corresponen els haplotips AB/Ab. La incertesa sorgeix per aquells individus heterozigots en dos o més *loci* de l'haplotip, aleshores hi ha més d'una parella d'haplotips (diplotips) consistent amb els genotips observats. Per exemple, a l'individu Aa Bb li poden correspondre els següents diplotips AB/ab o Ab/aB. El repte, doncs, per qualsevol dels mètodes d'estimació d'haplotips anteriors és establir quin és l'haplotip més probable entre els possibles per unes dades determinades.



**Figura 15.** Assignació d'haplotips. Situació més simple en què l'haplotip està constituït per 2 loci  $\alpha$  i  $\beta$ . Per l'individu amb genotip AA Bb només una parella d'haplotips és possible. Per l'individu Aa Bb, dos parelles d'haplotips són consistents amb els genotip observat.

L'estratificació fa referència a la mescla recent entre poblacions que dona lloc a associacions espúries entre al·lels no lligats per diferències en les freqüències al·leliques en les poblacions ancestrals i no perquè aquests es trobin propers en el cromosoma (Lander i Schork, 1994). Així, per exemple, suposem que una mostra de casos i controls és triada a partir d'una població constituïda de diverses subpoblacions. Si la malaltia o tret fenotípic d'interès es troba en una major freqüència en una de les subpoblacions, és d'esperar que aquesta estigui més representada en el grup de casos. Tanmateix, qualsevol al·lel que es trobi en una major freqüència en aquesta subpoblació que en la resta, presentarà una associació amb la malaltia, independentment de la seva localització cromosòmica. Com s'ha mencionat anteriorment, aquest fenomen és una de les causes d'associacions espúries o falsos positius (Knowler *et al*, 1988, Kittles *et al*, 2002) i explicaria, en part, la incapacitat de replicar alguns dels estudis d'associació de casos i controls (Terwilliger i Weiss, 1998).

S'han proposat diferents mètodes per validar els estudis de casos i controls, fins i tot, en presència d'estratificació de la mostra. Aquests es basen en l'anàlisi de genotips de múltiples *loci* independents escampats pel genoma i assumeixen la premissa que l'estratificació opera globalment en el genoma, afectant de manera similar a tots els *loci*. Mentre que l'associació per desequilibri de lligament amb l'al·lel causant de la malaltia o tret d'estudi opera localment, afectant tan sols marcadors propers en el cromosoma. Aquests mètodes es poden classificar en dues categories: *i*. Mètode de control genòmic (Devlin i Roeder, 1999; Pritchard i Rosenberg, 1999). Calcula la distribució estadística dels múltiples marcadors genètics no lligats en casos i controls. En presència d'estratificació, és d'esperar una major proporció d'associacions pels múltiples marcadors estudiats; *ii*. Mètode de valoració d'estratificació (Pritchard *et al*, 2000). Aquest assumeix que la mostra estudiada, tot i ser heterogènia, està composta de subpoblacions homogènies i assigna probabilitats als individus estudiats de pertànyer a les subpoblacions latents. Ambdues aproximacions no són mútuament

excloents. De fet, s'ha suggerit que el mètode de valoració d'estratificació pot emprar-se inicialment per identificar i excloure les mostres responsables de l'estratificació. I, a continuació, aplicar el mètode de control genòmic en l'estudi final (Freedman *et al*, 2004).

En aquest treball ens vam plantejar d'analitzar la possible associació entre el gen *DDR1* i l'esquizofrènia mitjançant un estudi del tipus cas-control. L'estratègia emprada va ser l'anàlisi de cinc SNPs localitzats en la regió del gen *DDR1*. De les variants genètiques identificades en l'Estudi 1, es van analitzar els SNP2, SNP6 i SNP11. A més, es va incloure una variant localitzada en l'extrem 3'UTR del gen *DDR1* (rs8408) i un cinquè SNP en la regió 5'UTR del gen *GTF2H4*, el qual es troba a unes 10 kb *downstream* de *DDR1* (rs1052693). Els SNP6 (S496S, P469L isoforma e) i SNP11 (N502S, M475V isoforma e) van ser les úniques variants incloses que implicaven un canvi d'aminoàcid en almenys una de les isoformes.

Les freqüències al·lèliques i genotípiques dels cinc polimorfismes estudiats es van comparar independentment entre ambdós grups, de casos i controls. A continuació, es va fer un estudi d'associació d'haplotips mitjançant l'estimació dels mateixos pels algorismes E-M i Phase. Paral·lelament i amb l'objectiu de descartar la presència d'estratificació en la mostra estudiada, es van analitzar 25 marcadors no lligats d'acord amb el mètode de valoració d'estratificació.

## RESULTATS

### SNPs analitzats

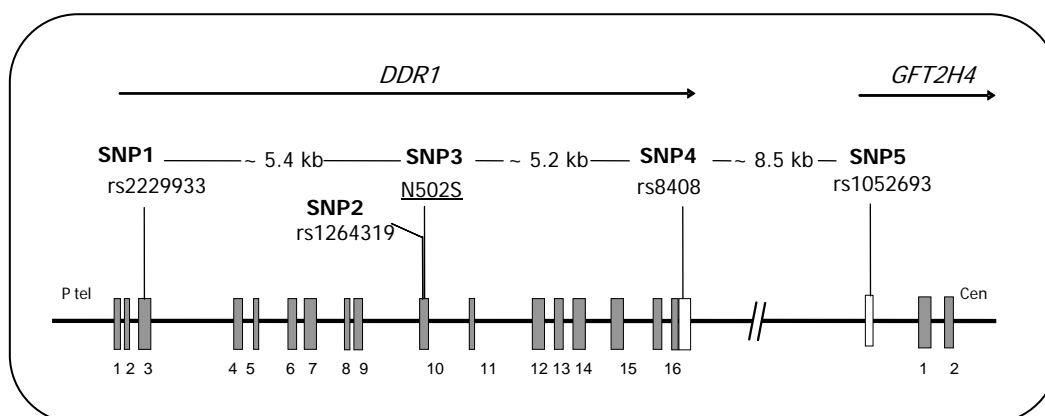
Per l'estudi d'associació, es van analitzar els dos SNPs identificats prèviament en l'Estudi 1 que comportaven un canvi d'aminoàcid. Aquests foren la variant SNP11 g13691A>G que resulta en el canvi d'una asparagina per una serina en la posició 502 i la variant SNP6 g13674C>T que resulta en el canvi d'una prolina per una leucina en la posició 469 en la isoforma e. A més, es van incloure dos SNPs addicionals, propers als extrems 5' i 3' del gen i amb una freqüència al·lèlica major a 0.1 per l'al·lel minoritari. Les variants que van complir amb aquest requisit, després d'analitzar una mostra de 48 controls agafats a l'atzar, van ser l'SNP2 g8312A>G, a l'exó 3, i l'SNP g18917T>C (rs8404), a la regió 3'UTR de *DDR1*. La informació per aquest últim es va extreure de la dbSNP.

Addicionalment, es va voler estudiar un SNP en el gen *GTF2H4*. Es van triar 2 SNPs localitzats en *GTF2H4* (rs1052963 i rs2230121) i es van analitzar en les 48 mostres control per tal de determinar si eren polimòrfics en la població estudiada. Dels dos SNPs analitzats, només rs1052963 va resultar polimòrfic i amb una freqüència al·lèlica superior a 0.1.

En la Taula XVI es mostra la correspondència entre la nomenclatura dels SNPs identificats en l'Estudi 1 i l'emprada en aquest estudi. La raó d'aquesta recodificació no és altra que la de simplificar el text i fer-lo més entenedor. En la Figura 16 es mostra la regió compresa pels 5 SNPs analitzats.

**Taula XVI.** Correspondència entre la nomenclatura dels SNPs entre l'Estudi 1 i l'Estudi 2.

SNP	Estudi 1	Estudi 2
rs2229933	2	1
rs1264319	6	2
N502S	11	3
rs8408	-	4
rs1052693	-	5



**Figura 16.** Esquema de les localitzacions dels 5 SNPs estudiats en les regions *DDR1* i *GTF2H4*. Les fletxes indiquen el sentit de la transcripció. Els rectangles blancs i grisos corresponen a les regions 5' i 3' UTR i exons, respectivament. Les línies verticals indiquen les posicions relatives pels diferents SNPs. Nomenclatura dels SNPs d'acord amb la dbSNP. El nou SNP no-sinònim identificat en l'Estudi 1 es mostra subratllat. P tel, p telòmer; Cen, centròmer.

### **Anàlisi d'associació per al·lels i genotips entre el grup control i el grup de malalts d'esquizofrènia**

Cap de les 5 variants estudiades en els gens *DDR1* i *GTF2H4* s'allunyava de l'equilibri de Hardy-Weinberg ( $p > 0.12$  per tots els *loci*), tant en el grup de malalts com de controls.

Els resultats obtinguts no van indicar la presència de diferències estadísticament significatives en la distribució dels genotips o al·lels entre la mostra control i la mostra de malalts d'esquizofrènia (Taula XVII). L'anàlisi de la distribució de les freqüències genotípiques es va realitzar sota tres supòsits diferents corresponents a tres models d'acció: 1. codominant, amb els tres genotips amb un efecte fenotípic dominant; 2. recessiu, amb el genotip homozigot per l'al·lel mutant o salvatge; 3. dominant, en el cas que només la presència d'un al·lel sigui suficient per a la manifestació de la malaltia (heterozigot i homozigot per l'al·lel mutant o salvatge). En cap cas es van observar diferències significatives.

Mitjançant la regressió logística es va estudiar l'associació de les 5 variants genotipades. Cap d'elles es va associar a cap de les variables clíniques considerades (subtipus d'esquizofrènia, tipus d'inici, tipus i curs de l'evolució de la malaltia, presència d'antecedents familiars i sexe). Agrupant els subtipus d'esquizofrènia, segons la simptomatologia positiva i negativa tampoc no es va detectar cap associació. Per últim, cap de les variants genètiques estudiades va tenir efecte sobre l'edat d'inici de la malaltia.

**Taula XVII.** Distribució de les freqüències al·lèliques i genotípiques de les variants analitzades en els gens *DDR1* i *GTF2H4* en el grup de malalts d'esquizofrènia i individus controls.

SNP	Genotip/ Al·lel	Casos N(%)	Controls N(%)	P
SNP 1 rs2229933	GG	189 (63.6)	204 (68.5)	ns
	GC	98 (33)	90 (30.2)	
	CC	10 (3.4)	4 (1.3)	
	G	0.801	0.836	ns
	C	0.199	0.164	
SNP 2 rs1264319	CC	232 (78.1)	226 (75.8)	ns
	CT	59 (19.9)	70 (23.5)	
	TT	6 (2.0)	2 (0.7)	
	C	0.880	0.876	ns
	T	0.120	0.124	
SNP 3 N502S	AA	285 (96)	289 (97)	ns
	AG	12 (4)	9 (3)	
	GG	0	0	
	A	0.980	0.980	ns
	G	0.020	0.020	
SNP 4 rs8408	TT	194 (65.3)	204 (68.5)	ns
	TC	93 (31.3)	90 (30.2)	
	CC	10 (3.4)	4 (1.3)	
	T	0.810	0.836	ns
	C	0.190	0.164	
SNP 5 rs1052693	AA	190 (64)	205 (68.8)	ns
	AG	95 (32)	88 (29.5)	
	GG	12 (4)	5 (1.7)	
	A	0.799	0.835	ns
	G	0.201	0.164	

P, significació considerada si  $p < 0.05$  per la prova  $\chi^2$ ; ns, no significatiu.

### Estimació d'haplotips en el grup control i el grup de malalts d'esquizofrènia i anàlisi d'associació entre ambdós grups

Les freqüències haplotípiques per les combinacions dels diferents marcadors analitzats es van estimar, en casos i controls per separat, fent ús de l'algoritme E-M, implementat en el programa Arlequin. Malgrat no es van detectar diferències significatives de les distribucions haplotípiques entre els dos grups d'estudi ( $p=0.775$ , 10 000 Markov iteracions), es va observar una major diversitat d'haplotips en el grup de malalts (25) que en el de controls (18) (Taula XVIII). En ambdós grups, es van detectar tres haplotips majoritaris (haplotips 1, 8 i 16), amb una freqüència major al 10% i els quals representaven més del 93% dels cromosomes estudiats. La resta dels haplotips es presentaven amb freqüències menors a l'1.5%.

Atès que varis autors apunten que l'algoritme bayesià Phase és més acurat que l'algoritme E-M per l'estimació d'haplotips (Stephens *et al*, 2001), les dades es van analitzar amb el programa Phase. Sota aquest model, els resultats evidenciaven la presència de diferències significatives en la distribució d'haplotips entre els dos grups d'estudi ( $p=0.002$ , 500 permutacions) (Taula XVIII). A l'igual que en els anteriors resultats, es van observar tres haplotips majoritaris (haplotips 1, 8 i 16), els quals

representaven també més del 93% dels cromosomes estudiats. Es continuava detectant un major nombre d'haplotips minoritaris o rars, de freqüència menor a 1.5%, en els malalts (18) que en els controls (13).

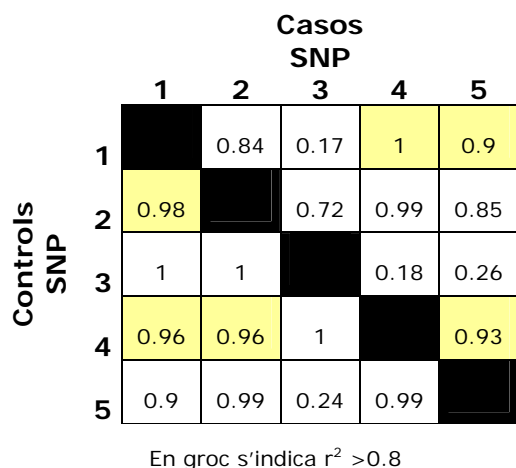
**Taula XVIII.** Distribució de les freqüències dels haplotips estimats mitjançant els algorismes E-M i Phase implementats en els programes Arlequin i Phase, respectivament.

Haplotype <sup>a</sup>	SNP					Mètode			
	1	2	3	4	5	Phase <sup>b</sup>		Arlequin <sup>c</sup>	
	G/C	C/T	A/G	T/C	A/G	Controls	Casos	Controls	Casos
1	G	C	A	T	A	0.688768	0.651686	0.69306	0.66082
2	.	.	.	.	G	0.004943	0.015259	0.00337	0.01216
3	.	.	.	C	.	0.000067	0.001088	0	0
4	.	.	.	C	G	0.002946	0.001841	0.00169	0
5	.	.	G	.	.	0.014990	0.013944	0.01426	0.01000
6	.	.	G	.	G	0	0.000744	0	0.00151
7	.	.	G	C	G	0	0.000084	0	0
8	.	T	.	.	.	0.123410	0.114043	0.12236	0.11351
9	.	T	.	.	G	0.000218	0.002248	0	0.00170
10	.	T	.	C	.	0	0.00005	0	0
11	.	T	.	C	G	0.000117	0.000101	0	0
12	.	T	G	.	.	0.000111	0.000117	0.00084	0.00165
13	C	.	.	.	.	0.001732	0.005242	0	0.00171
14	C	.	.	.	G	0.001298	0.002975	0.00169	0.00403
15	C	.	.	C	.	0.006358	0.011916	0.00409	0.01026
16	C	.	.	C	G	0.154736	0.170797	0.15769	0.17293
17	C	.	G	C	G	0	0.005228	0	0.00704
18	C	T	.	.	.	0.000101	0.002198	0	0.00172
19	C	T	.	.	G	0	0.000268	0	0.00095
20	C	T	.	C	.	0.000034	0.000034	0.00096	0
21	C	T	.	C	G	0.000171	0.000119	0	0

<sup>a</sup> Amb un punt s'indica l'al·lel corresponent a l'haplotip majoritari; <sup>b</sup>  $p=0.002$ , 500 permutacions; <sup>c</sup>  $p=0.775$ , 10 000 Markov iteracions.

### Anàlisi de lligament entre els al·lells dels diferents loci dels haplotips estimats

Les relacions d'associació al·lèlica entre els diferents parells d'SNPs es van estimar amb el test de desequilibri de lligament (LD) implementat en el programa Arlequin. El grau d'aquesta associació es va quantificar amb el coeficient  $D'$ . En els dos grups d'estudi, els SNP1, SNP2, SNP4 i SNP5 es trobaren en un fort LD entre ells (Figura 17). Mentre que el patró de lligament observat per l'SNP3 fou lleugerament diferent en el grup de malalts comparat amb el grup de controls. En controls, es mantenia un fort LD entre l'SNP3 i la resta dels SNPs, excepte per l'SNP5. Mentre que en casos, aquest LD es trobava debilitat.

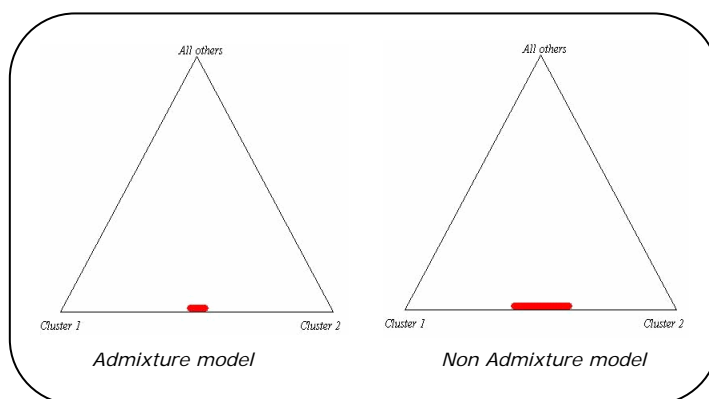


**Figura 17.** Sumari del test de desequilibri de lligament entre les parelles de SNPs.

### Anàlisi de l'estratificació de la mostra

Amb l'objectiu de descartar la presència d'estratificació en la mostra estudiada, es van genotipar un total de 25 marcadors bial·lèlics escampats pels diferents cromosomes i les dades es van analitzar d'acord amb el mètode de valoració d'estratificació, implementat en el programa Structure. Els resultats obtinguts no van suggerir la presència de subpoblacions en la mostra estudiada. De fet, tal i com s'observa en la Figura 18, tots els individus es trobaren agrupats en únic *cluster* o grup, independentment del model de partida (vegeu Participants, materials i mètodes).

Els 25 marcadors analitzats no es van allunyar de l'equilibri de Hardy-Weinberg en cap del dos grups d'estudi. No es van observar diferències significatives entre les distribucions de les freqüències al·lèliques i genotípiques entre casos i controls. La distribució de les freqüències al·lèliques pels 25 marcadors analitzats es mostren en la Taula XIX.



**Figura 18.** Diagrama corresponent als resultats obtinguts en l'estudi de valoració de l'estratificació amb el programa Structure assumint la presència de 2 poblacions (K=2).



**Taula XIX.** Distribució de les freqüències al·lèliques dels marcadors estudiats en l'anàlisi de l'estratificació de la mostra.

SNP	Al·lel	Casos	Controls
<i>MTHFR</i> Ala222Val	Ala 222	0.597	0.614
	Val 222	0.403	0.386
<i>F5</i> Arg506Gln	Arg 506	0.976	0.976
	Gln 506	0.024	0.024
<i>IL1B</i> -511C>T	-551 C	0.676	0.705
	-551 T	0.324	0.295
<i>PPAR<math>\gamma</math></i> Pro12Ala	Pro 12	0.922	0.934
	Ala 12	0.078	0.066
<i>TF</i> Pro570Ser	Pro 570	0.840	0.813
	Ser 570	0.160	0.187
<i>FABP-2</i> Ala54Thr	Ala 54	0.730	0.712
	Thr 54	0.270	0.288
<i>ADRB2</i> Arg16Gly	Arg 16	0.598	0.600
	Gly 16	0.402	0.400
<i>HFE</i> Cys282Tyr	Cys 282	0.772	0.771
	Tyr 282	0.228	0.229
<i>ESR1</i> IVS1 -401T>C	-401 T	0.533	0.542
	-401 C	0.467	0.458
<i>IL16</i> -285T>C	-285T	0.803	0.786
	-285C	0.197	0.214
<i>MDR1</i> 3435C>T	3435 C	0.526	0.519
	3435 T	0.474	0.481
<i>LPL</i> Ser244Ter	Ser244	0.909	0.892
	Thr244	0.091	0.108
<i>RXRA</i> -25A>G	-25A	0.635	0.618
	-25G	0.375	0.382
<i>ADRB1</i> Gly389Arg	Gly 389	0.737	0.743
	Arg 389	0.263	0.257
<i>APOA5-A4</i> Intergenic T>C	T	0.629	0.628
	C	0.371	0.368
<i>A2M</i> Val1000Ile	Val 1000	0.686	0.686
	Ile 1000	0.314	0.314
<i>BRCA2</i> 203G>A	203G	0.963	0.962
	203A	0.037	0.038
<i>ESR2</i> 1730A>G	1730 A	0.564	0.571
	1730 G	0.436	0.429
<i>SULT1A1</i> Arg213His	Arg 213	0.732	0.727
	His 213	0.268	0.273
<i>MCP-1</i> -2518G>A	-2518 G	0.805	0.764
	-2518 A	0.195	0.236
<i>SERPINB2</i> Asn120Asp	Asn 120	0.753	0.759
	Asp 120	0.247	0.231
<i>APOE</i> Cys112 Arg	Cys 112	0.901	0.902
	Arg 112	0.099	0.098
<i>PRNP</i> Met129Val	Met 129	0.66	0.648
	Val 129	0.34	0.352
<i>CBS</i> 1080C>T	1080 C	0.654	0.639
	1080 T	0.346	0.361
<i>PPAR<math>\alpha</math></i> Leu162Val	Leu 162	0.925	0.934
	Val 162	0.075	0.066

## DISCUSSIÓ

Aquest és el primer treball en què el *DDR1* s'estudia com a gen candidat per l'esquizofrènia. Dins del marc de la hipòtesi del neurodesenvolupament per aquesta malaltia, el raonament que ens van empènyer a suggerir *DDR1* com a gen candidat en l'esquizofrènia va raure en la seva localització cromosòmica a 6p21.3 (Valent *et al*, 1996), regió candidata a contenir un gen de susceptibilitat per aquesta malaltia, el seu patró d'expressió diferencial durant el neurodesenvolupament (Zerlin *et al*, 1993; Sanchez *et al*, 1994) i el fet que es tractés d'un RPTK de resposta a col.lagen (Shrivastava *et al*, 1997; Vogel *et al*, 1997), el qual s'havia demostrat que modula la diferenciació i proliferació de cèl.lules neuroepitel.lials i astroglia (Goetschy *et al*, 1987; Ali *et al*, 1998). En el moment de plantejar-nos aquesta hipòtesi de treball, *DDR1* no havia estat implicat en cap patologia en humans i la seva expressió gènica i proteica en SNC no estava caracteritzada.

En concret, en aquest treball, ens vam plantejar de valorar l'associació entre el gen *DDR1* i l'esquizofrènia a nivell genètic, mitjançant un estudi de casos i controls. Per això, es van estudiar 5 marcadors cobrint part de les regions del gen *DDR1* i *GTF2H4* (Figura 16). Aquest últim és el gen més proper a *DDR1* i es localitza a unes 10 kb *downstream*. La variant estudiada a *GTF2H4* es va incloure amb l'objectiu d'expandir la zona d'anàlisi genètica.

L'anàlisi de la distribució de les freqüències haplotípiques va evidenciar diferències estadísticament significatives entre el grup de malalts d'esquizofrènia i el d'individus control (Taula XVIII). En el grup de malalts es va detectar una major diversitat haplotípica causada per la presència d'haplotips rars, de freqüència menor a 0.5%. Tant en malalts com en controls, es va observar que més d'un 93% de la distribució haplotípica es concentrava en tan sols 3 combinacions al.lèliques (haplotips 1, 8, 16), de les quals l'haplotip majoritari estava compost per l'al.lèl més comú de cada locus. La resta dels haplotips es trobaren amb freqüències menors a 1.5%. Les freqüències pels haplotips majoritaris foren molt similars en els dos grups d'estudi.

El fet que els resultats de l'anàlisi individual dels SNPs no mostressin cap contribució al.lèlica o genotípica amb la malaltia o amb cap de les variables clíniques considerades, posava de manifest quelcom ben establert que és que l'avaluació de la combinació d'al.lèls o haplotips és més informativa que l'anàlisi independent dels marcadors.

Els resultats del desequilibri de lligament entre parelles de *loci*, indicaven la presència d'un fort LD entre els marcadors SNP1, SNP2, SNP4 i SNP5, tant en el grup de casos com el de controls, la qual cosa no era d'estranyar tenint en compte la seva proximitat física. De manera interessant, es van detectar diferències en l'SNP3 entre els dos grups d'estudi. En malalts, l'associació d'aquesta variant amb la resta dels SNPs analitzats es trobava debilitada. Mentre que en controls, l'SNP3 es mantenia amb un fort LD amb els SNP1, SNP2 i SNP4 i moderat amb l'SNP5. Aquest resultat

explicava, en part, l'aparició d'un major nombre d'haplotips rars en el grup de malalts.

L'SNP3 és una variant no-sinònima que comporta el canvi d'una asparagina a serina en la posició 502 de la proteïna. Tot i que aquest canvi pugui tenir un efecte en la funcionalitat de DDR1 (de moment, en desconeixem el seu impacte), la mutació per si mateixa no s'ha pogut associar amb l'esquizofrènia, sinó el conjunt del seu entorn haplotípic. En ambdós grups d'estudi, els dos al·lels de l'SNP3 foren equipfregüents, amb una freqüència de 0.02 per l'al·lel mutat S502.

L'anàlisi de les variables clíniques dels malalts d'esquizofrènia identificats amb l'al·lel mutat S502 (12 dels 297 malalts) no va evidenciar la presència d'una característica comuna en aquest grup de pacients. D'aquests, sis eren homes (amb una edat d'inici entre 18 i 36) i sis eren dones (amb una edat d'inici entre 18 i 29). Ens va cridar l'atenció, però, que 8 d'aquests malalts tinguessin antecedents familiars per esquizofrènia (66.6% versus 25.3% en la mostra total de malalts estudiats) i, en concret, 4 (3 dones i 1 home) tenien una mare malalta d'esquizofrènia. L'al·lel S502 es distribueix de manera similar en el grup de controls, on es presenta de forma heterozigota en 9 participants (7 dones i 2 homes d'edats compreses entre els 28 i els 53 anys). Aquests no presentaven antecedents per cap malaltia major i tampoc no es va observar la presència d'una característica comuna entre ells.

Malgrat no s'ha pogut trobar una correlació directa entre cap dels haplotips estimats i la malaltia, els resultats d'aquest treball semblen apuntar a que d'altres variants en el gen *DDR1* o en regions flanquejants podrien tenir un paper en l'etiologia de l'esquizofrènia, en un grup dels malalts. No es pot excloure la possibilitat que un o més al·lels de susceptibilitat estiguin en LD amb algunes de les variants haplotípiques observades per una submostra dels malalts.

La interpretació d'aquestes observacions s'ha de fer amb cautela, ja que com s'ha comentat anteriorment, són diversos els factors que poden esbiaixar els resultats d'un estudi de casos i controls. La diversitat haplotípica observada en el grup de malalts d'esquizofrènia podria ser reflex de la presència d'heterogeneïtat genètica o subpoblacions en la mostra. En el present treball, les mesures de control de l'estratificació aplicades foren dues: una a nivell demogràfic i l'altra a nivell genètic. L'aparellament geogràfic i ètnic dels individus casos i controls fou un requisit per la seva inclusió en l'estudi. Tots els participants eren de nacionalitat espanyola i d'origen caucàsic mediterrani. Per altra banda, d'acord amb el mètode de valoració de l'estratificació (Pritchard *et al*, 2000), es van genotipar 25 marcadors escampats pel genoma i que no havien estat associats a l'esquizofrènia. Els resultats obtinguts no van indicar la presència de subpoblacions en la mostra estudiada (Fig 17). S'ha proposat que l'anàlisi de 20 a 30 marcadors és suficient per la detecció d'estratificació (Pritchard i Rosenberg, 1999). No obstant, estudis recents apunten que aquest nombre no seria suficient per detectar la presència subtil d'estratificació

poblacional (Freedman *et al*, 2004). Així, doncs, tot i haver considerat dues mesures de control de l'estratificació, no es pot excloure una certa presència de la mateixa en la mostra estudiada.

Malgrat els protocols i procediments de treball establerts en l'estudi, hi havia la possibilitat que errors en el genotipatge fossin els responsables directes dels resultats observats. Per tal de descartar aquesta sospita, les mostres es van genotipar per duplicat pels SNP1 i SNP4 en els laboratoris d'ACLARA BioSciences Inc. i del Centre de Recerca Biomèdica. Els resultats obtinguts van presentar una concordança del 100%. Aquest control també es va aplicar per 4 dels SNPs analitzats en l'estudi d'estratificació. En aquest cas, els genotips pels polimorfismes en els gens *MTHFR*, *F5*, *APOE* i *ADRB2* es van analitzar pel mètode *eTag Multiplex Invader SNP Assay* a ACLARA BioSciences Inc. i per PCR-RFLP al Centre de Recerca Biomèdica. Tampoc no es van observar discrepàncies.

Un dels valors d'aquest estudi rau en l'àmplia mostra de malalts i controls analitzada, així com també en els procediments seguits pel reclutament de tots els participants, controls i malalts. Pel grup control, és d'importància no tan sols la disposició de la mostra biològica, sinó també de la història clínica detallada. Aquesta permet de tenir en compte, per exemple, els antecedents personals i familiars per patologia psiquiàtrica i origen geogràfic, variables de rellevància a tenir en compte per tal de minimitzar el fenomen d'estratificació.

Com ja s'ha comentat en la introducció d'aquesta tesi, l'etiologia de l'esquizofrènia és encara confosa. Durant els últims anys, diversos estudis han aportat noves dades que apunten a una disfunció de la glia i del procés de mielinització en l'etiopatogènia de l'esquizofrènia (Bartzokis, 2002; Hof *et al*, 2002). En aquest context, és interessant apuntar que els estudis d'expressió en ratolins, duts a terme pel nostre grup, demostren que *DDR1* té un patró d'expressió gènica molt particular durant el neurodesenvolupament i en l'animal adult, localitzant-se específicament en oligodendròcits (treballs en fase de publicació). Es desconeix quina o quines poden ser les funcions de *DDR1* en aquest tipus cel.lular i quins factors afectarien la seva expressió. Fins el present moment, les úniques dades publicades que posen de manifest la importància de *DDR1* en SNC són les de Bhatt i col.laboradors. En aquest treball es demostra que la interacció de *DDR1* amb el col.lagen i la corresponent via de senyalització activada és essencial per la formació axonal de les cèl.lules granulars durant el desenvolupament del cerebel (Bhatt *et al*, 2000).

D'altra banda, s'ha observat recentment que l'activació de *DDR1* inhibeix l'expressió del gen *agrin*, el qual codifica per una proteïna de MEC que s'encarrega de l'agregació de receptors d'acetilcolina durant el desenvolupament de la unió neuromuscular en l'embriogènesi (Faraci *et al*, 2004). Durant aquest procés de maduració neuromuscular és clau la participació de neuregulin 1 (*NRG1*), la principal funció de la qual, en aquest procés, seria induir l'expressió de les subunitats del receptor

d'acetilcolina (Martinou *et al*, 1991). De manera interessant, diversos estudis recents suggereixen *NRG1* com a gen de susceptibilitat per l'esquizofrènia (Stefansson *et al*, 2002; Stefansson *et al*, 2003; Williams *et al*, 2003).

És clar que els resultats d'aquest treball han de ser replicats per tal d'adreçar la incertesa de si *DDR1* està associat a l'esquizofrènia. La replicació en d'altres mostres, ja siguin poblacionals o nuclis familiars i amb un bagatge genètic diferent, serà una prova de confiança. D'altra banda, a la llum dels resultats d'aquest estudi, l'anàlisi de la resta de variants identificades en l'Estudi 1 pot ser molt informativa i aportar noves dades. Sis d'aquestes variants es localitzaren en regions intròniques i es desconeix si tenen cap efecte en els mecanismes de *splicing* o en la maduració del transcrit. L'anàlisi mutacional realitzat es va centrar en l'estudi de les regions codificants, excloent regions reguladores de l'expressió gènica, com regions promotores i 5' i 3'UTR. Serà també interessant explorar la variabilitat d'aquestes regions i estudiar el seu efecte en l'expressió de *DDR1*.

És cert que el repte de la investigació genètica en psiquiatria i, en concret, en l'esquizofrènia, és el fet que aquesta malaltia continua essent un diagnòstic d'exclusió, ja que cap de les característiques clíniques són patognòmiques de la malaltia. Les categories de diagnòstic es basen en la identificació d'una sèrie de signes i símptomes, els quals podrien ser resultat de múltiples processos patològics amb bases biològiques diferents. La mancança d'una eina de diagnòstic acurat i el seu patró d'herència no mendeliana havien posat en dubte la utilitat dels estudis genètics. Però ha estat durant els darrers dos anys que treballs independents evidencien un efecte moderat dels gens *DTNBP1* (Straub *et al*, 2002; Schwab *et al*, 2003; Williams *et al*, 2004), *NRG1* (Stefansson *et al*, 2002; Stefansson *et al*, 2003; Williams *et al*, 2003) i molt probablement també *G30/G70* (Chumakov *et al*, 2002) en l'etiologia de l'esquizofrènia. Fins el present moment, no s'han pogut identificar al·lels o haplotips concrets de risc i les seves implicacions funcionals en l'esquizofrènia són encara desconegudes. Malgrat l'evident complexitat de l'esquizofrènia, d'aquests resultats se'n desprèn que d'altres gens de susceptibilitat seran molt probablement identificats durant els propers anys, obrint noves línies de recerca amb les perspectives d'entendre la patogènia d'aquesta malaltia i aportar unes millors eines de diagnòstic i tractament.

## ESTUDI 3: Validació del mètode *eTag™ Multiplex Invader®* per l'anàlisi d'SNPs

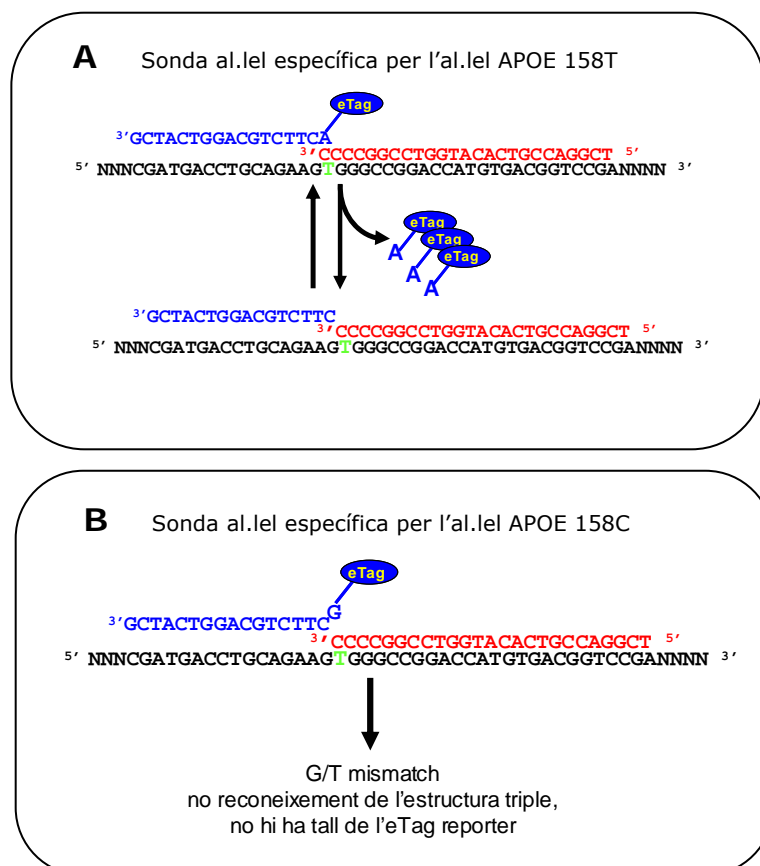
### INTRODUCCIÓ

La capacitat de genotipar múltiples SNPs ha esdevingut una eina indispensable en els laboratoris de genètica humana. Els SNPs són el tipus de variant genètica més abundant i un dels principals responsables de la diversitat genètica entre individus. És per això que molts d'ells estan involucrats en condicions amb base genètica, com malalties o la resposta diferencial a fàrmacs. Al mateix temps, donada la seva abundància, estabilitat i simplicitat (la gran majoria d'SNPs són bial·lèlics), fa que aquests siguin molt valuosos pels estudis de diversitat genètica, ja siguin estudis d'associació global de genoma implicant centenars d'SNPs o bé d'associació i lligament centrats en determinades regions cromosòmiques. Com ja s'ha descrit en l'apartat de la Introducció, existeix un gran ventall de tècniques per l'anàlisi d'SNPs, de les quals moltes permeten la detecció simultània de més d'un SNP per mostra. Ara bé, no es disposa d'un mètode de genotipatge estàndard o de referència (Kwok i Chen, 2003). La utilització d'una o altra metodologia en un laboratori determinat depèn sovint de la quantitat de mostres a processar per dia, número d'SNPs a genotipar per mostra, infraestructura disponible i cost final del procés.

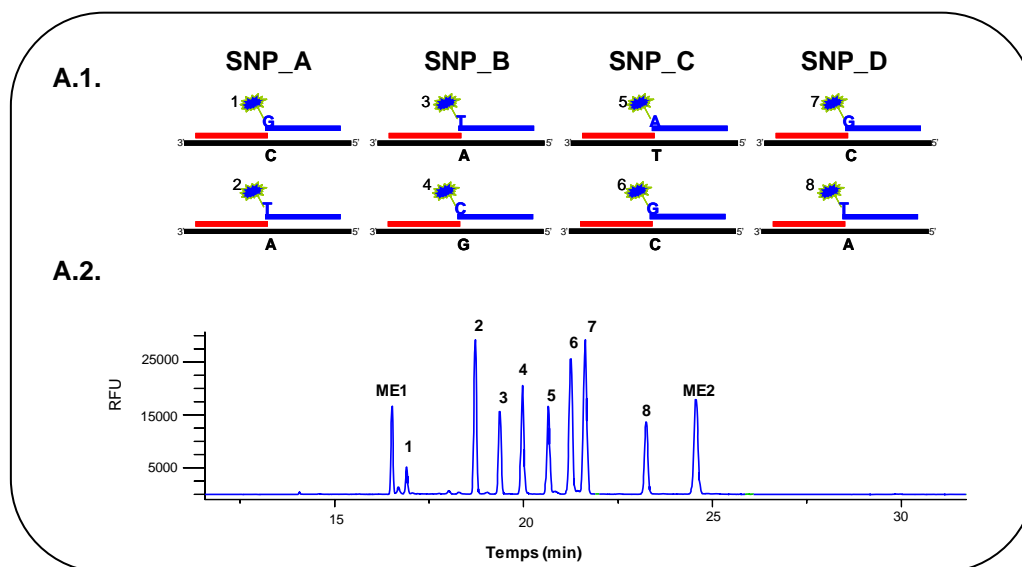
El mètode Invader ha estat extensament validat i estudiat per la detecció d'SNPs (Lyamichev *et al*, 1999; Hall *et al*, 2000; Fors *et al*, 2000). Es tracta d'un mètode homogeni, isoterm, quantitatiu, altament específic i que prescindeix de l'amplificació del DNA genòmic. La reacció té lloc a única temperatura en plaques de 96 o 384 pouets les quals es llegeixen, al final de la reacció, amb un lector de plaques de fluorescència. És per això que és un mètode atractiu ja que no requereix d'instruments específics i pot transferir-se ràpidament a qualsevol laboratori. No obstant, com també ja s'ha comentat, es veu limitat pel fet que només un SNP pot analitzar-se per pouet i mostra. El motiu rau en el sistema de marcatge i detecció de les sondes al·lèl específiques amb fluorocroms convencionals, com per exemple, carboxifluoresceïna (FAM) i RedMond Red™. D'altra banda, requereix de grans quantitats de DNA genòmic, generalment de >100 ng per SNP (Lyamichev i Neri, 2003). Per tal de superar aquesta limitació, ja que molt sovint es disposa de quantitats limitades de mostra, el DNA genòmic pot ser amplificat per PCR prèviament a la reacció Invader (PCR-Invader) (Mein *et al*, 2000).

En el present estudi es va explorar la possibilitat de combinar el mètode Invader i la capacitat de detecció de múltiples SNPs mitjançant el marcatge al·lèl específic de les sondes amb els fluorocroms *eTag reporters*. Aquesta nova aproximació és anomenada reacció *eTag Multiplex Invader* per l'anàlisi d'SNPs. A l'igual que en el mètode Invader original, l'especificitat es basa en el reconeixement i tall per l'enzim Cleavase de l'estructura triple formada per superposició entre l'olignucleòtid Invader i una sonda al·lèl específica, marcada amb un *eTag reporter*, en hibridar amb el DNA diana. De manera que, l'*eTag reporter* i el nucleòtid en l'extrem 5' són tallats i

alliberats a la solució (Figura 19). El canvi del nucleòtid en el punt de superposició elimina la hibridació entre la sonda al.lel específica i el DNA diana. Això impedeix la formació de l'estructura triple necessària per la plena activitat de l'enzim i la sonda al.lel específica marcada es manté intacta. Com s'ha descrit anteriorment, els *eTag reporter* són molècules fluorescents amb mobilitats electroforètiques definides que es poden resoldre i detectar per LIF-CE en seqüenciadors automàtics. A mode d'exemple, en la Figura 20, es mostra un esquema de l'anàlisi de 4 SNPs pel mètode *eTag Multiplex Invader*.



**Figura 19.** Esquema de la reacció *eTag Multiplex Invader*. A. L'estructura triple per superposició de la sonda APOE 158T i l'oligonucleòtid Invader en 1 nucleòtid amb la seqüència diana és reconeguda i tallada per l'enzim Cleavase alliberant-se l'*eTag reporter* a la solució. B. La no complementarietat entre la sonda APOE 158C i la seqüència diana impedeix la formació de l'estructura triple, de manera que la sonda roman intacta. En blau s'indiquen les sondes al.lel específiques APOE 158T i APOE 158C marcades en l'extrem 5' amb un *eTag reporter*. En vermell i negre, s'indiquen l'oligonucleòtid Invader i la seqüència diana, respectivament. El punt polimòrfic per l'SNP APOE 158 en la seqüència diana se senyala en verd.



**Figura 20.** Il·lustració esquemàtica de la reacció *eTag Multiplex Invader* per la detecció de 4 SNPs. A.1. Esquema de les sondes i oligonucleòtids necessaris per la detecció de 4 SNPs. Totes les sondes estan marcades amb un *eTag reporter* de diferent mobilitat electroforètica. Els *eTag reporters* s'enumeren de l'1 al 8. A.2. Electroforograma amb 8 *eTag reporters*, corresponents als 4 SNPs, i 2 marcadors electroforètics (ME1 i ME2). Separació electroforètica realitzada amb un ABI3100 Genetic Analyzer (Applied Biosystems). En blau s'indiquen les sondes al·lel específiques i sonda control marcades en l'extrem 5' amb un *eTag reporter*, en vermell els oligonucleòtids Invader i en negre les seqüències diana.

L'objectiu d'aquest treball va ser desenvolupar i optimitzar els protocols pel mètode *eTag Multiplex Invader* per l'anàlisi simultània d'SNPs. En primer lloc, es va observar que l'amplificació del DNA genòmic per PCR era imprescindible per poder dur a terme la reacció *eTag Multiplex Invader*. A continuació, es va realitzar un estudi preliminar on es van optimitzar els protocols i es va avaluar l'especificitat de la tècnica per un total de 10 SNPs i 60 mostres, algunes de les quals havien estat prèviament caracteritzades per alguns dels SNPs. Els genotips resultants del mètode *eTag Multiplex Invader* es van comparar amb els obtinguts pel mètode PCR-RFLP. En un segon estudi de validació de la tècnica, es van analitzar 480 mostres. Els genotips resultants es van comparar amb els obtinguts al Centre de Recerca Biomèdica.



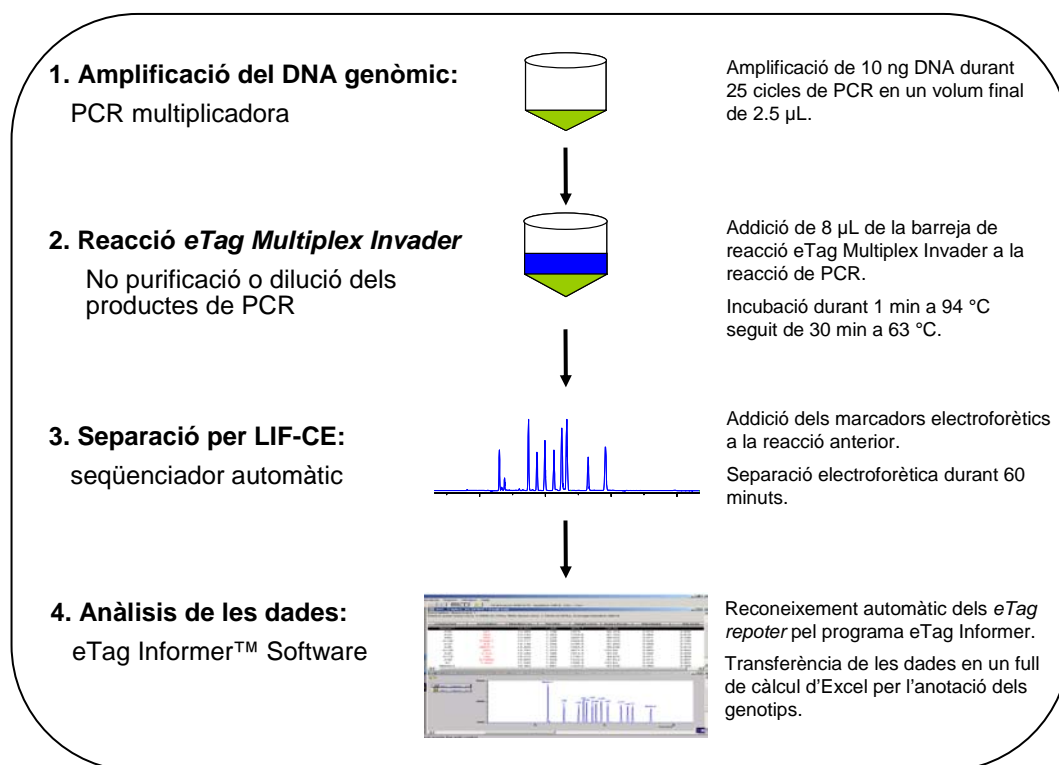
## RESULTATS

### Primera avaluació del mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs

Els 10 SNPs analitzats van ser seleccionats a partir d'una llista generada a Third Wave Technologies Inc., i es van triar sense tenir referència dels resultats del mètode Invader. Totes les seqüències de les sondes i oligonuclèotids van ser dissenyades amb el programa Invader Creator d'acord amb el mètode Invader Square i per una temperatura de reacció d'uns 60-63°C. Les sondes al·lel específiques es van modificar d'acord amb el protocol *eTag Multiplex Invader*, i es van marcar amb un *eTag reporter* en l'extrem 5' i una molècula de biotina en un dels nucleòtids timina. El marcatge amb biotina va permetre 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 eren separats i detectats per LIF-CE.

El mètode Invader es caracteritza per poder genotipar DNA genòmic directament sense la necessitat d'amplificar-lo per PCR o altres mètodes alternatius. Ara bé, el cas particular de la utilització d'*eTag reporters* en DNA genòmic, ja fos seguint el disseny Invader Square o marcant directament les sondes al·lel específiques (vegeu Figura 9), va resultar amb genotips inespecífics i, molt sovint, amb una baixa intensitat del senyal. En aquest estudi inicial, només es van incloure 4 dels 10 SNPs (MTHFR Ala222Val, MTHFR Glu429Ala, ApoE Cys112Arg, ApoE Arg158Cys). Amb l'objectiu de facilitar l'accés dels oligonuclèotids específics a les seqüències dianes, es van avaluar les següents estratègies: 1. Trencament del DNA genòmic (per tractament amb DNases, vòrtex, elevades temperatures i ultrasons); 2. Increment del temps d'incubació de la reacció d'1 fins a 16 hores; 3. Valoració de la quantitat de DNA genòmic de 10 ng fins a 1 µg; 4. Valoració de la concentració dels oligonuclèotids Invader i sondes al·lel específiques de 0.1 fins a 1 µM. En cap d'aquests casos es va detectar un genotipatge específic o un increment del senyal pels SNPs analitzats. Seguidament, es va avaluar la tècnica d'amplificació total del genoma mitjançant la utilització d'encebadors degenerats, d'acord amb el protocol de Sanchez-Cspedes i col.laboradors. D'aquesta manera es disposaria d'un únic protocol d'amplificació i s'evitava el disseny d'encebadors específics de PCR. Ara bé, els resultats van tornar a ser negatius i no es va detectar un senyal específic.

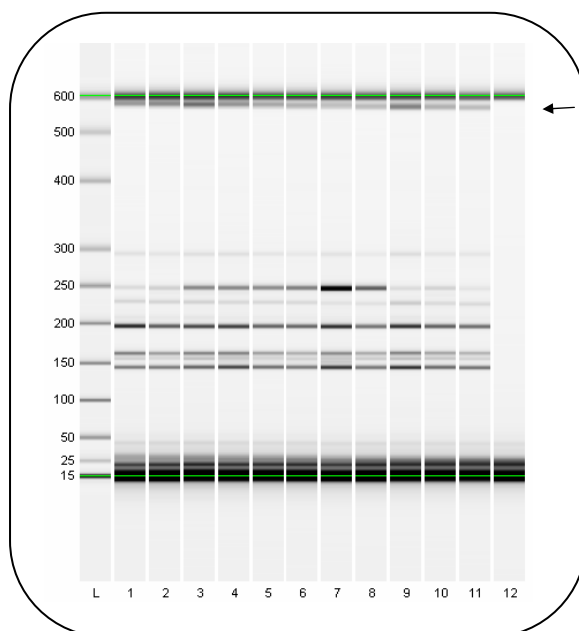
Finalment, les regions amb els SNPs d'interès es van amplificar per PCR multiplicadora. En aquest cas, els genotips van ser específics pels 4 SNPs analitzats. A partir d'aquests resultats, es van analitzar diferents paràmetres de la reacció *eTag Multiplex Invader* (requeriment o no de la purificació del producte de PCR, temperatura d'incubació, temps d'incubació, tampó de reacció, concentració de l'enzim Cleavase X i dels diferents oligonuclèotids) i es van obtenir les condicions del protocol estàndard que es descriuen en l'apartat de Participants, materials i mètodes. El protocol establert es resumeix en la Figura 21.



**Figura 21.** Esquema del protocol seguit en l'estudi de validació de la tècnica *eTag Multiplex Invader*.

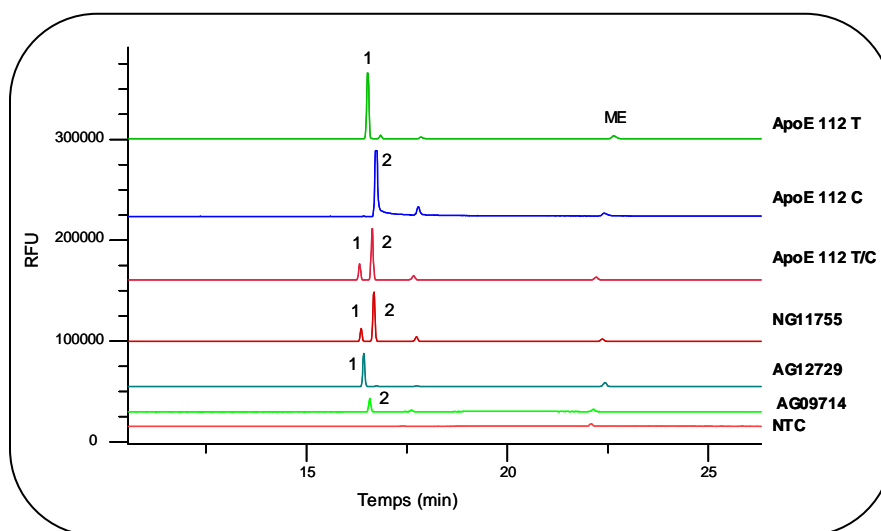
La purificació, dilució o transferència dels productes de PCR no va ser necessària. La reacció de genotipatge es va afegir directament al pouet de PCR sense la necessitat d'eliminar els dNTPs i primers en excés, la Taq DNA polimerasa o sals del tampó de la reacció de PCR. És a dir, la mateixa placa de 96 pouets podia emprar-se per la reacció de PCR, reacció *eTag Multiplex Invader* i injecció en l'aparell d'EC, ja que en aquest la injecció de les mostres es fa a partir de plaques de 96.

Es va observar que la presència de més de  $10^{e5}$  còpies de la seqüència d'interès era necessària per tal d'obtenir uns resultats acceptables. D'altra banda, productes de PCR inespecífics i el fet d'obtenir concentracions variables entre els productes de PCR no van interferir amb els resultats finals. D'acord amb el protocol de PCR establert, l'amplificació de 5-10 ng DNA genòmic va resultar en unes  $10^{e7}$ - $10^{e11}$  còpies per cadascun dels fragments de PCR d'interès. En la Figura 22 es mostra la separació electroforètica de la reacció de PCR multiplicadora.



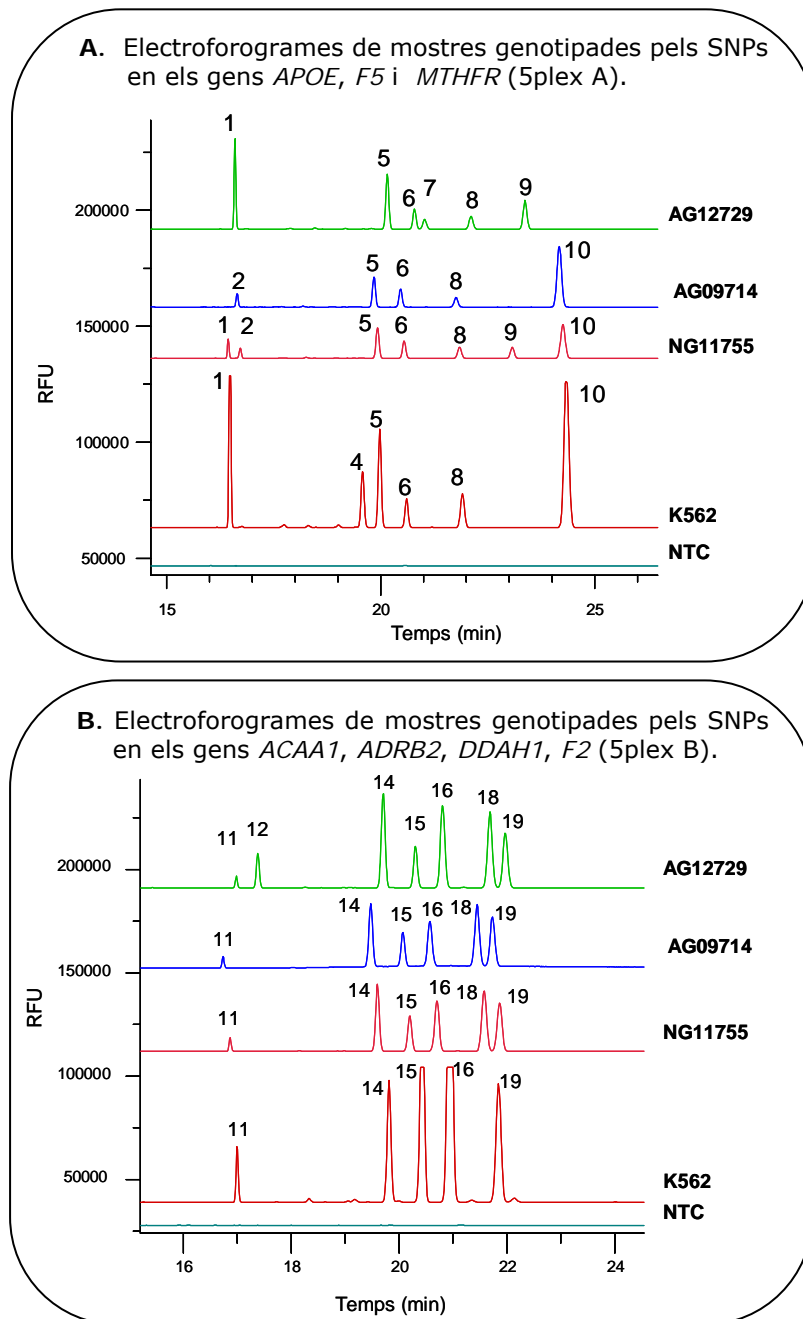
**Figura 22.** Imatge dels productes de PCR per la reacció de PCR multiplicadora per 11 mostres (carrils 1-11) i un control negatiu (carril 12). El carril L correspon als marcadors electroforètics. La fletxa indica la presència d'un producte de PCR inespecífic. Separació electroforètica realitzada en un instrument Agilent 2100 BioAnalyzer.

Totes les sondes al.l.el específiques i Invader van analitzar-se individualment fent ús de DNA sintètic de cadena senzilla i de fragments de PCR de mostres de DNA prèviament caracteritzat. D'aquesta manera es va confirmar l'especificitat de les sondes. En la Figura 23, es mostren els electroferogrames per l'SNP APOE Cys112Arg, a mode d'exemple. Per tal de reduir la complexitat del sistema, els 10 SNPs es van dividir en dues reaccions: 1. Reacció 5plex A, va incloure els SNPs en els gens *APOE*, *F5* i *MTHFR*; 2. Reacció 5plex B, va incloure els SNPs en els gens *ACAA1*, *ADRB2*, *DDAH1* i *FII*. En primer lloc, es va comprovar que els resultats de les reaccions on s'analitzava un únic SNP concordaven amb els obtinguts en les reaccions múltiples per les 12 mostres DNA genòmic control. A continuació, es van analitzar 48 mostres triades a l'atzar del Grup Control.



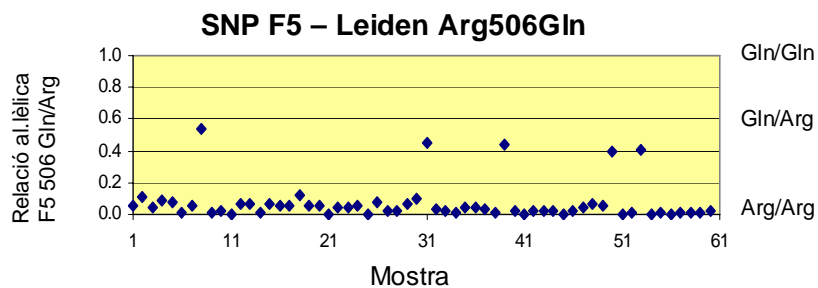
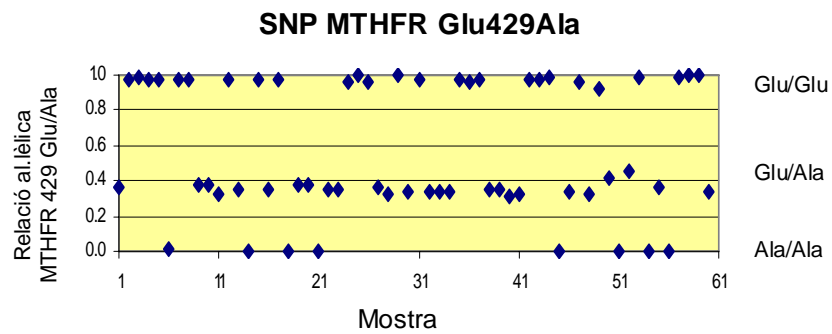
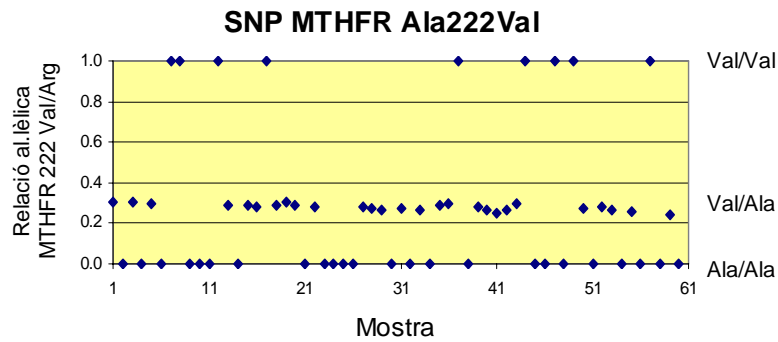
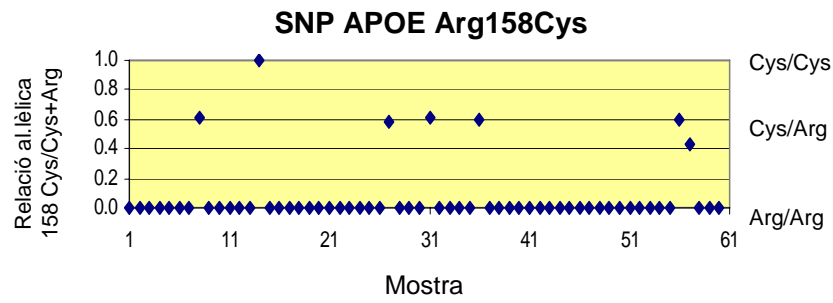
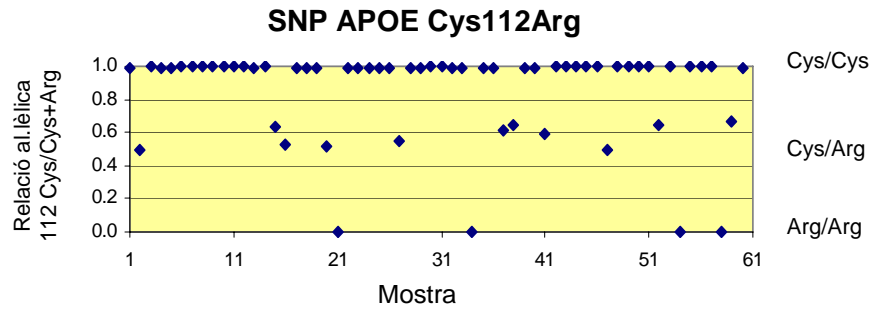
**Figura 23.** Especificitat de l'assaig *eTag Multiplex Invader* per l'SNP APOE 112. Electroforogrames pels DNAs controls de cadena senzilla (APOE 112 T, APOE 112 C i APOE 112 T/C), mostres NG11755 (genotip ApoE 112 T/C), AG12729 (genotip ApoE 112 T/T), AG09714 (genotip ApoE 112 C/C) i control negatiu (NTC). Les sondes corresponents als al·lels APOE 112 T i APOE 112 C es van marcar amb els *eTag reporter* 1 i 2, respectivament. ME: marcador electroforètic. Separació electroforètica realitzada amb MegaBACE1000 (AmershamPharmacia).

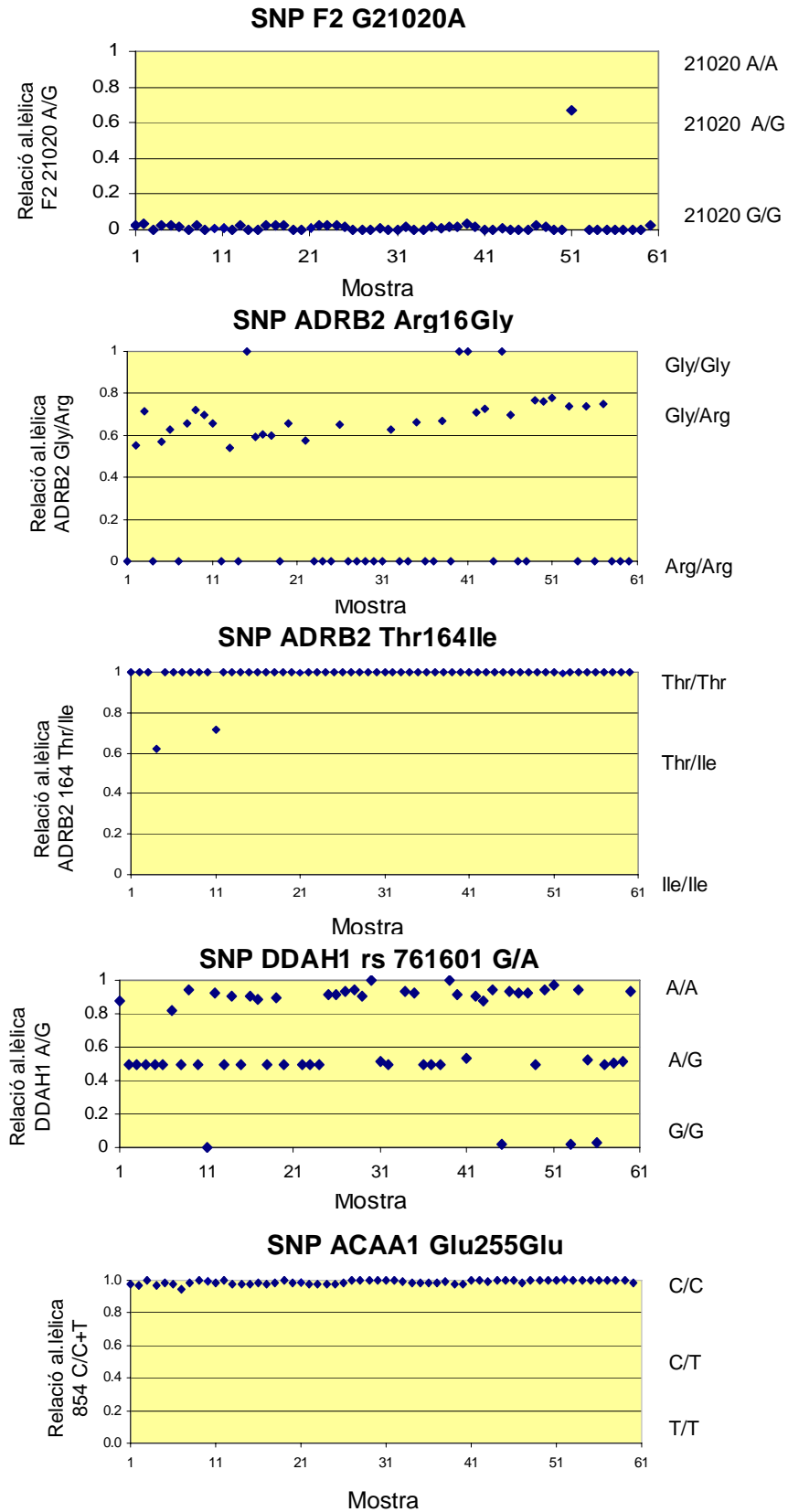
Els electroforogrames corresponents a les reaccions 5plex A i B es mostren en la Figura 24. Els resultats es resumeixen en les gràfiques de la Figura 25, en les quals per cadascuna de les 60 mostres i SNP es representa el valor de la relació al·lèlica. Es va observar que els valors s'agrupaven en 3 *clusters* clarament diferenciats i corresponents als 3 genotips possibles per 6 dels SNPs analitzats. Pels SNPs F5 Leiden, F2 G20210A i ADRB2 Thr164Ile només es van observar dos *clusters*, corresponents als genotips homozigots per l'al·lel salvatge i heterozigots. En el cas de la variant del gen *ACAA1*, totes les mostres van resultar ser homozigotes per l'al·lel C. Les mostres amb una relació al·lèlica  $>0.8$  o  $<0.2$  es van assignar amb genotips homozigots; les mostres amb una relació al·lèlica  $\geq 0.2$  o  $\leq 0.8$  es van assignar amb genotips heterozigots. Aquests valors es van escollir arbitràriament a partir de les representades en els gràfics de la Figura 25. En els gràfics de la Figura 24, es pot observar que, per molts dels SNPs, la intensitat del senyal entre al·lels no era equivalent. És per això que el valor pels genotips heterozigots va oscil·lar entre 0.2 i 0.8. Aquest fet suggeria la presència de diferències en la  $T_m$  de les sondes al·lel específiques, en l'eficiència de reconeixement i tall dels *eTag reporter* per l'enzim Cleavase o bé en la injecció dels *eTag reporter* per EC. Malgrat aquest fet, la diferenciació dels diferents *clusters* corresponents als 3 genotips possibles fou clara.



**Figura 24.** Electroforogrames de les reaccions *eTag Multiplex Invader* 5plex A i 5 plex B. S'inclouen tres mostres DNA control i un control negatiu (NTC). Separació electroforètica realitzada amb MegaBACE1000 (AmershamPharmacia). La correspondència entre *eTag reporters* i al·lels és la següent:

- |                    |                        |
|--------------------|------------------------|
| 1 - APOE 112 Cys   | 11 - ADRB2 Gly16       |
| 2 - APOE 112 Arg   | 12 - ADRB2 Arg16       |
| 3 - FV Leiden Gln  | 13 - ACAA1 854T        |
| 4 - MTHFR 429 Ala  | 14 - ADRB2 Thr164      |
| 5 - FV Leiden Arg  | 15 - ACAA1 854C        |
| 6 - APOE 158 Cys   | 16 - FII 20210G        |
| 7 - MTHFR 429 Glu  | 17 - FI 20210A         |
| 8 - APOE 158 Arg   | 18 - DDAH1 rs7618601 A |
| 9 - MTHFR 222 Val  | 19 - DDAH1 rs7618601 G |
| 10 - MTHFR 222 Ala | 20 - ADRB2 Ile164      |





**Figura 25.** Sumari dels resultats obtinguts en el primer estudi d'avaluació de la tècnica *eTag Multiplex Invader* per l'anàlisi d'SNPs.

El soroll de fons degut al reconeixement i tall no específic dels *eTag reporters* per l'enzim Cleavase X va ser mínim per tots els SNPs, tal i com es pot observar en els electroforogrames corresponents al controls negatius en les Figures 23 i 24. És per això que la raó entre senyal específic:soroll de fons va ser superior a 1:10 i el genotipatge acceptable per tots els SNPs.

En un ~3% de les mostres (3 mostres per cada placa de 96 pouets), no es va detectar cap senyal en els electroforogrames, ni *eTag reporters* ni marcadors electroforètics. Això suggeria la presència de problemes durant la separació electroforètica i no pas de la reacció *eTag Multiplex Invader*. De fet, aquestes mostres es van tornar a injectar i es van obtenir resultats acceptables.

Paral·lelament, les 60 mostres es van genotipar pel mètode PCR-RFLP. La concordança entre els genotips obtinguts entre ambdós mètodes va ser del 100%.

### **Segona avaluació del mètode *eTag Multiplex Invader* per l'anàlisi d'SNPs**

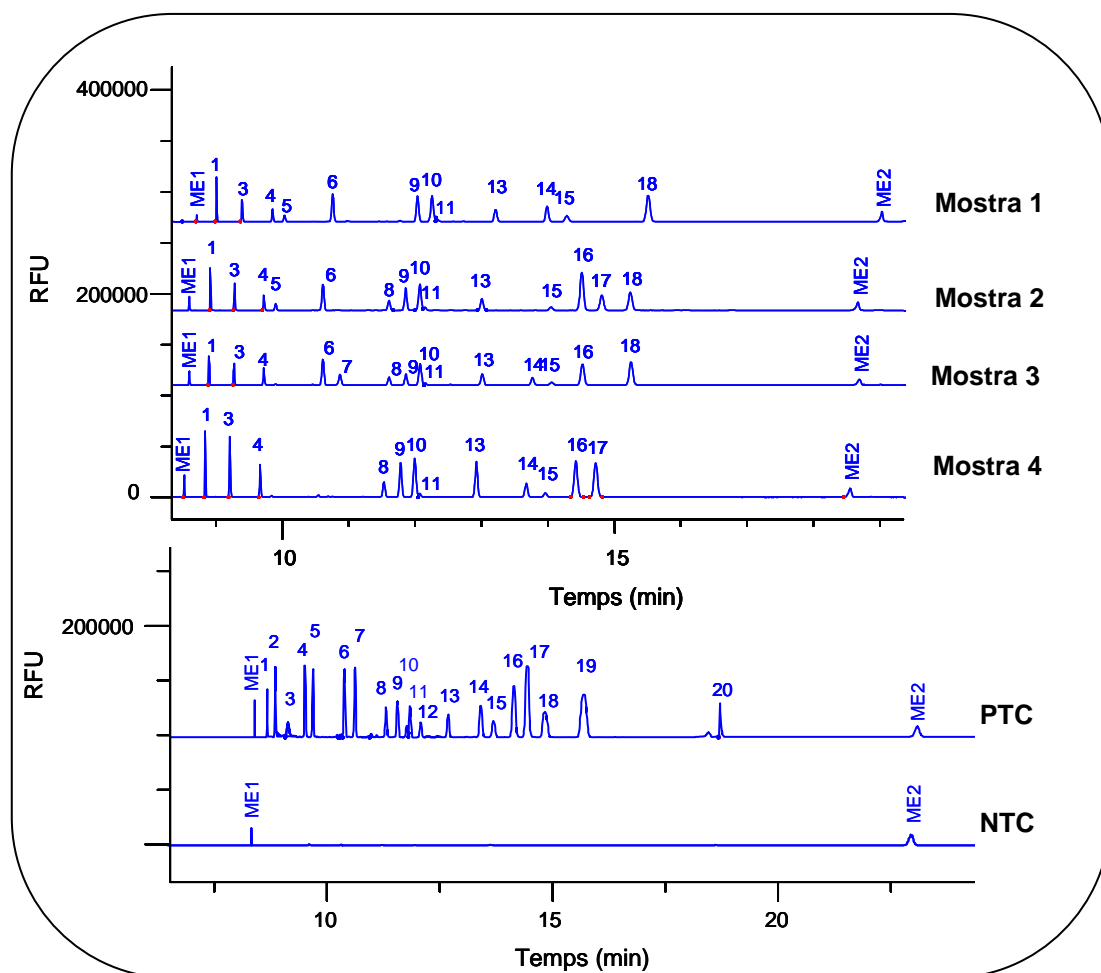
Per tal d'augmentar el nivell de complexitat de l'assaig, els 10 SNPs es van combinar en una sola reacció *eTag Multiplex Invader* (10plex). En primer lloc, es va observar que el mateix protocol podia emprar-se en passar de l'anàlisi de 5 a 10 SNPs. Seguidament, es va comprovar que els genotips de la reacció 10plex per les 12 mostres de DNA genòmic control concordessin amb els resultats de les reaccions 5plex A i 5plex B. A continuació, les 480 mostres van ser genotipades. Per tal d'avaluar potencials problemes en l'amplificació per PCR i la consistència de la reacció *eTag Multiplex Invader*, totes les mostres es van analitzar per duplicat. En la Figura 26 es mostren els electroforogrames per la reacció 10plex. Els resultats es presenten en les gràfiques de la Figura 27 i Taula XX.

A l'igual que en la primera avaluació de la tècnica, es va observar que la intensitat dels pics electroforètics pels dos al·lels d'un determinat SNP no era equivalent. Així, per exemple, per l'SNP APOE Cys112Arg, el valor dels genotips heterozigots fou de ~0.7, indicant que l'*eTag reporter* de la sonda APOE 112Cys era tallat i detectat més eficientment que el de la sonda APOE 112Arg. No obstant, els valors de les relacions al·lèliques per tots els SNP es van agrupar en 3 *clusters* diferenciats corresponents als 3 genotips possibles. En aquest cas, les mostres amb una relació al·lèlica >0.85 o <0.15 es van assignar amb genotips homozigots; mostres amb una relació al·lèlica  $\geq 0.15$  o  $\leq 0.85$  es van assignar amb genotips heterozigots. No es van observar discrepàncies genotípiques entre les mostres duplicades. Les mostres havien estat prèviament genotipades al laboratori del Centre de Recerca Biomèdica pel mètode PCR-RFLP. En comparar els resultats entre ambdós mètodes, es va obtenir una concordança del 100%.

De nou, no es van identificar individus homozigots mutants pels SNPs F5 Leiden, F2 G20210A o ADRB2 Thr164Ile. A més, totes les mostres van resultar ser homozigotes pel per l'al·lel C de l'SNP ACAA1 Glu255Glu.

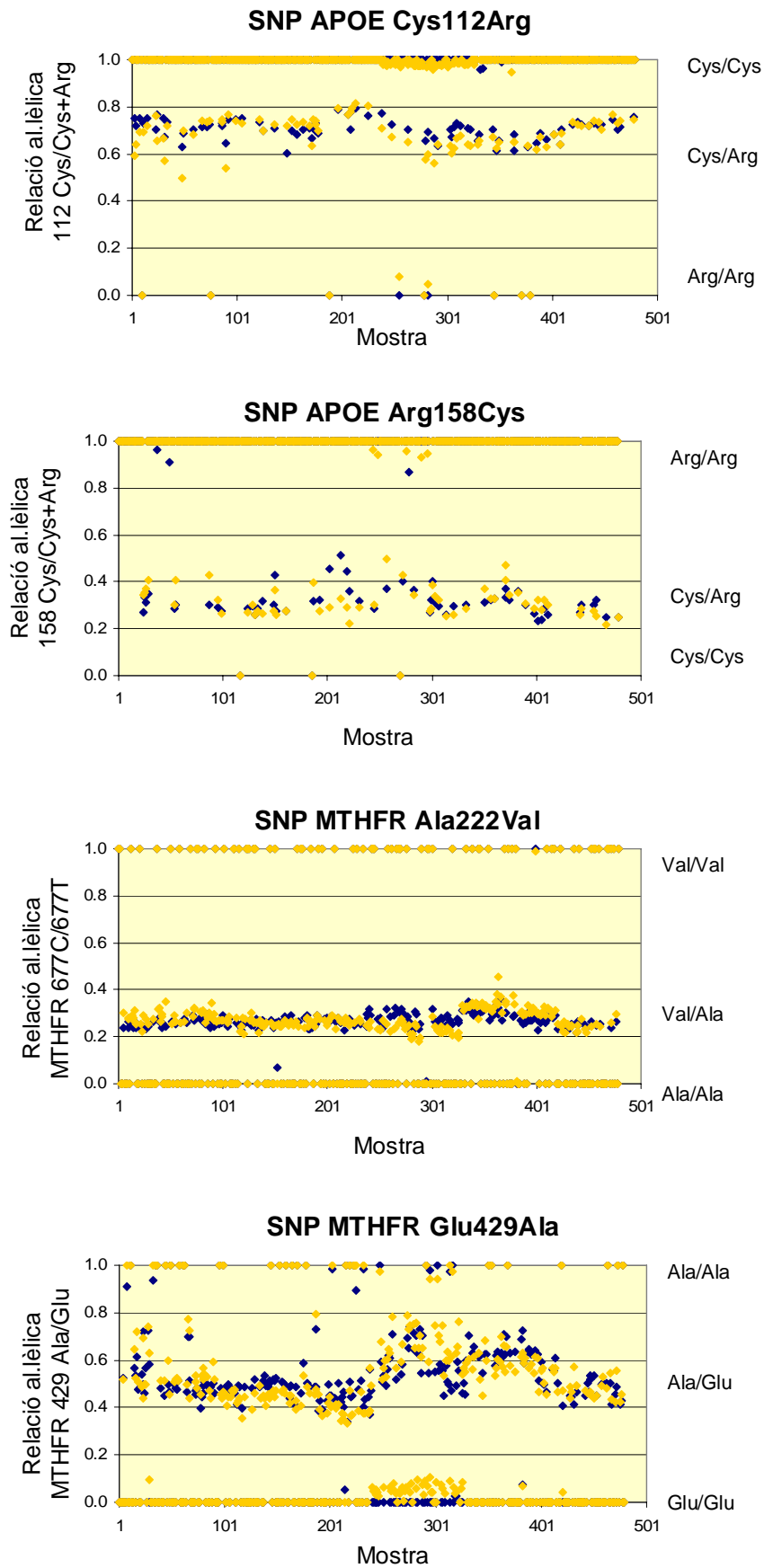


La PCR multiplicadora i reacció de genotipatge es van repetir per 4 de les mostres (0.7%), ja que el senyal detectat per tots els SNPs no era acceptable (relació senyal:soroll inferior a 1:10). A banda, es va observar un 3% de mostres fallides per problemes de separació electroforètica. Aquestes es van tornar a injectar i es van obtenir resultats acceptables.

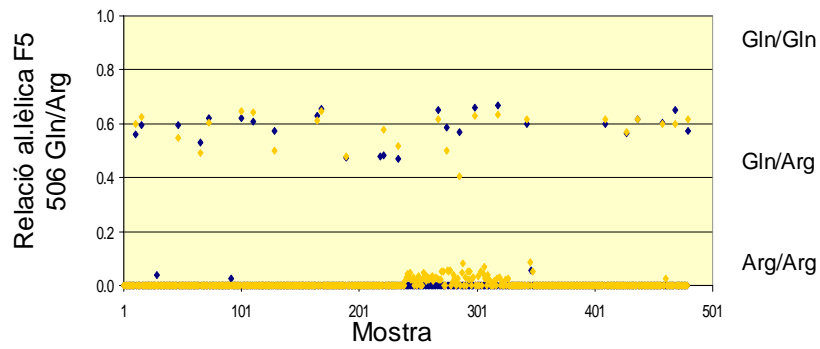


**Figura 26.** Electroforogrames de la reacció *eTag Multiplex Invader* 10plex. S'inclouen les dades de quatre mostres enumerades de l'1 al 4, un control positiu (PTC) i un control negatiu (NTC). Separació electroforètica realitzada amb MegaBACE1000 (AmershamPharmacia). ME1 i ME2: marcadors electroforètics 1 i 2. La correspondència entre *eTag reporters* i al.l els és la següent:

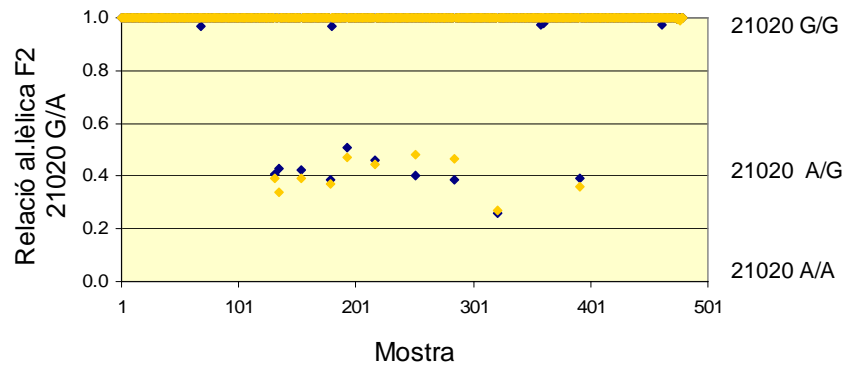
- |                  |                     |                    |                    |
|------------------|---------------------|--------------------|--------------------|
| 1 - ADRB2 Thr164 | 6 - DDAH1 C         | 11 - APOE 158 Arg  | 16 - MTHFR 222 Ala |
| 2 - ACAA1 854T   | 7 - F5 Leiden 1691A | 12 - APOE 158 Cys  | 17 - ADRB2 Gly16   |
| 3 - ACAA1 854C   | 8 - MTHFR 429 Ala   | 13 - DDAH1 T       | 18 - ADRB2 Arg16   |
| 4 - APOE 112 Cys | 9 - F5 Leiden 1691G | 14 - MTHFR 222 Val | 19 - ADRB2 Ile164  |
| 5 - APOE 112 Arg | 10 - F2 20210G      | 15 - MTHFR 429 Glu | 20 - F2 20210A     |



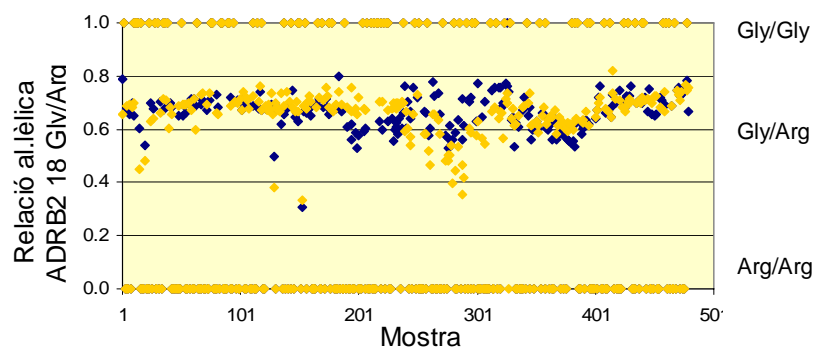
### SNP F5 – Leiden – Arg506Gln



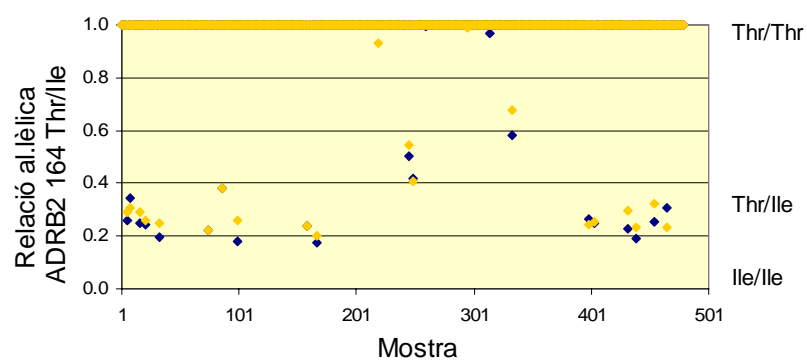
### SNP F2 G21020A

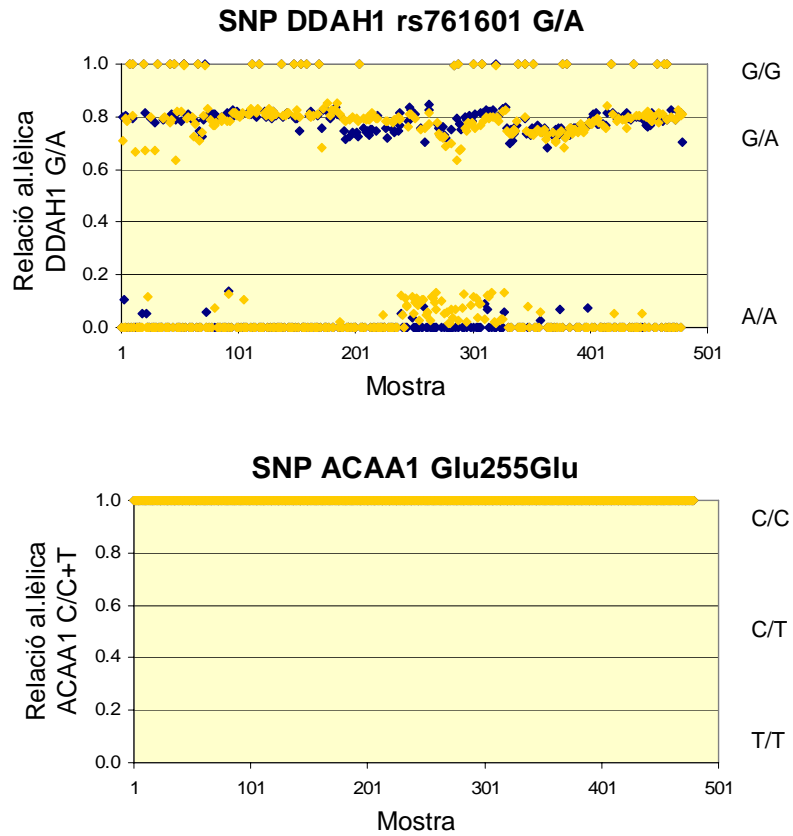


### SNP ADRB2 Arg16Gly



### SNP ADRB2 Thr164Ile





**Figura 27.** Sumari dels resultats obtinguts en el segon estudi d'avaluació de la tècnica *eTag Multiplex Invader* per l'anàlisi de SNPs.

**Taula XX.** Freqüències genotípiques i al·lèliques pels 10 SNPs analitzats mitjançant el mètode *eTag Multiplex Invader*.

<b>SNP</b>	<b>Genotips N (%)</b>			<b>Al·lèls</b>	
<b>APOE Cys112Arg</b>	Cys/Cys 391(81.5)	Cys/Arg 80 (16.7)	Arg/Arg 9 (1.9)	Cys 0.898	Arg 0.102
<b>APOE Arg158Cys</b>	Arg/Arg 421 (87.7)	Arg/Cys 56 (11.7)	Arg/Cys 3 (0.6)	Arg 0.935	Cys 0.065
<b>MTHFR Ala222Val</b>	Ala/Ala 183 (38.1)	Ala/Val 225 (46.9)	Val/Val 72 (15)	Ala 0.616	Val 0.384
<b>MTHFR Glu429Ala</b>	Glu/Glu 235 (49)	Glu/Ala 207 (43.1)	Ala/Ala 38 (7.9)	Glu 0.705	Ala 0.295
<b>F5 – Leiden Arg506Glu</b>	Arg/Arg 454 (94.6)	Arg/Glu 26 (5.4)	Glu/Glu 0	Arg 0.973	Glu 0.027
<b>F2 G21020A</b>	G/G 470 (97.9)	G/A 10 (2.1)	A/A 0	G 0.99	A 0.01
<b>ADRB2 Arg16Gly</b>	Arg/Arg 183 (38.1)	Arg/Gly 218 (45.4)	Gly/Gly 79 (16.5)	Arg 0.608	Gly 0.392
<b>ADRB2 Thr164Ile</b>	Thr/Thr 461 (96)	Thr/Ile 19 (4)	Ile/Ile 0	Thr 0.98	Ile 0.02
<b>DDAH1 rs761601</b>	A/A 247 (51.5)	A/G 200 (41.7)	G/G 33 (6.9)	A 0.723	G 0.277
<b>ACAA1 Glu255Glu</b>	C/C 481 (100)	C/T 0	T/T 0	C 100	T 0

## DISCUSSIÓ

En aquest estudi es presenta una nova plataforma pel genotipatge de múltiples SNPs per combinació del mètode Invader i el marcatge de les sondes al·lel específiques amb els *eTag reporters*. Els resultats demostren que aquest nou mètode, anomenat *eTag Multiplex Invader*, és viable per l'anàlisi d'SNPs a partir de DNA amplificat per PCR. En concret, es van desenvolupar assaigs satisfactòriament per l'anàlisi simultània de 10 SNPs. En comparació amb la tècnica Invader, comporta dos passos addicionals: l'amplificació del DNA genòmic per PCR i la separació i detecció dels *eTag reporter* mitjançant la tècnica LIF-CE. No obstant, totes dues tècniques són d'ús general i rutinari en els laboratoris de genètica. Per tant, la posada punt del mètode *eTag Multiplex Invader* no requereix d'una instrumentació o infraestructura específica.

La incapacitat de detectar senyals específics a partir de mostres de DNA genòmic indicava que l'amplificació del senyal (o generació d'*eTag reporters*) era insuficient perquè fos detectada per LIF-CE. D'altra banda, la utilització de més de 0.5 o 1 µg de DNA genòmic resultava inviable per aplicacions pràctiques. És per això que el DNA genòmic es va amplificar per PCR multiplicadora. La disposició d'un major nombre de còpies de les seqüències d'interès, les quals eren més accessibles a les sondes i oligonucleòtids Invader, va resultar en la generació i detecció de senyals específics. D'aquesta manera, la PCR multiplicadora combinada amb l'anàlisi simultània dels 10 SNPs, va resultar amb l'ús de  $\leq 1$  ng DNA genòmic per SNP. Aquesta reducció de la quantitat de mostra per SNP és molt significativa en comparació als  $> 100$  ng per SNP necessaris pel mètode Invader (Lyamichev i Neri, 2003). Val a dir, que el mètode PCR-Invader presenta també aquesta avantatge. No obstant, els productes de la PCR s'han de diluir i transferir a altres pouets on cada SNP s'analitza individualment (Mein *et al*, 2000; Nevilie *et al*, 2002).

És cert que un dels grans problemes en els laboratoris on es realitzen PCRs diàriament és la contaminació dels mateixos amb fragments de PCR, els quals poden ser amplificats de forma inespecífica i donar lloc a falsos resultats. Ara bé, aquest risc queda minimitzat si es realitza una bona pràctica de laboratori i se separen les àrees de processament de les mostres pre i post PCR, com fou el cas en aquest treball. A més, existeixen d'altres estratègies per evitar aquest tipus de contaminació. Com per exemple, la substitució de dTTP per dUTP durant l'amplificació i tractament previ a la reacció de PCR amb uracil-DNA glicosilasa (Longo *et al*, 1990). De fet, la majoria dels mètodes de genotipatge d'SNPs requereixen de l'amplificació del DNA genòmic per PCR (Syvänen, 2001; Kwok i Chen, 2003). Dit això, vam creure que el fet d'amplificar el DNA genòmic per PCR prèviament a la reacció *eTag Multiplex Invader*, no suposava un gran desavantatge o impediment per tirar endavant el treball.

El protocol del mètode *eTag Multiplex Invader* es va simplificar el màxim possible. Tot l'assaig es va realitzar en una única placa de 96 pouets: PCR, reacció *eTag Multiplex Invader*, separació i detecció per la tècnica LIF-CE (Figura 21). D'aquesta

manera es van evitar possibles errors en la transferència i manipulació de les mostres. La reacció de genotipatge va tenir lloc en solució, no s'utilitzaren micropartícules o altres tipus de superfícies per capturar les múltiples sondes o seqüències diana. Això va permetre una ràpida difusió i mobilitat dels reactius i una reducció del temps d'incubació de la reacció de genotipatge (30 minuts). Amb el protocol presentat en aquest treball i partint de la mostra de DNA genòmic, els genotips de 10 SNPs per 96 mostres podien obtenir-se en 4 hores, aproximadament. Aquest increment en el processament de mostres és considerable en comparació amb el mètode original Invader o bé PCR-Invader. En aquests, només 1 SNP per mostra i pouet és analitzat i la durada del procés oscil·la entre 2 i 5 hores.

Les mateixes condicions de reacció de PCR i *eTag Multiplex Invader* van ser emprades tant pel genotipatge d'1 com dels 10 SNPs combinats. Aquesta uniformitat en els protocols és molt desitjable quan es volen analitzar múltiples SNPs simultàniament. D'aquesta manera, la fase d'optimització per cada SNP individual queda minimitzada a l'avaluació inicial de l'especificitat de les sondes al·lel específiques. En aquest treball, la utilització del programa Invader Creator pel disseny i predicció de  $T_m$  similars per múltiples sondes va resultar satisfactòria. Com era d'esperar, el marcatge de les sondes amb un *eTag reporter* en l'extrem 5' i una molècula de biotina en un dels nucleòtids timina no va afectar la  $T_m$  i funcionalitat de les mateixes. Ara bé, cal tenir en compte, que el disseny de les sondes no és trivial malgrat la utilització d'un programa de predicció. I l'especificitat de les mateixes s'ha d'avaluar individualment. Així, per exemple, en l'estudi de Mein i col.laboradors, un 19% de les sondes (7 de 37 SNPs analitzats) es van haver de tornar a dissenyar.

L'assignació dels genotips va realitzar-se en funció del valor de la relació al·lèlica. Per cada SNP i mostra es va representar gràficament el valor d'aquesta relació (Figures 25 i 27). A partir de les gràfiques, es van escollir arbitràriament els criteris d'assignació de genotips, ja que clarament es diferenciaven 3 *clusters* corresponents als 3 possibles genotips. Ara bé, es va observar que el rendiment de les sondes al·lel específiques era diferent per cada SNP. És per això que, els valors de les relacions al·lèliques pels genotips heterozigots oscil·laren entre 0.2 i 0.8 i 0.15 i 0.85 en la primera i segona avaluació, respectivament. Aquest fet indicava la presència de petites diferències en la  $T_m$  de les sondes al·lel específiques, en l'eficiència de reconeixement i tall dels *eTag reporter* per l'enzim Cleavase o bé en la injecció dels *eTag reporter* per EC. A més, es van observar diferències entre ambdós estudis, en particular pels SNPs en el gen *DDAH1* i *F5*. Això suggeria que la combinació dels 10 SNPs tenia un efecte en el rendiment de la reacció, però no en l'especificitat de la mateixa. Ara bé, per cada nou panell d'SNPs i *eTag reporter* a analitzar, s'haurà de fer un estudi preliminar per determinar el criteri d'assignació de genotips.

La injecció electrocinètica és molt sensible a la presència de sals o d'altres compostos amb càrrega negativa. La presència d'impureses en les mostres de DNA o petits canvis en la concentració de sals per evaporació de la reacció, podrien explicar la

dispersió observada dels valors de la relació al·lèlica. En base a la relació d'absorbàncies a 260 i 280 nm, totes les mostres eren lliures de proteïnes, però no es pot descartar la presència d'altres tipus de contaminants no proteïcs o d'altres variables que afectessin l'assaig.

En un laboratori clínic o amb un elevat processament de mostres, un percentatge de mostres fallides del 3-4%, com el del present treball, suposaria un greu problema pel cost en temps i diners de les mostres a repetir. Aquest percentatge fou similar a l'obtingut pel mètode PCR-Invader (Mein *et al*, 2000; Olivier *et al*, 2002). Un 0.7% de mostres fallides va ser degut a problemes de pipeteig del DNA genòmic i d'amplificació per PCR. El 3% restant va ser causa d'una fallida en la separació electroforètica (injecció nul·la de la mostra o capil·lar malmès). Atès que la injecció electrocinètica permet de conservar la mostra sobrant o no injectada, les mostres fallides es poden tornar a analitzar sense necessitat de repetir la reacció de PCR i genotipatge. Malgrat això, és clar que cal desenvolupar noves condicions d'electroforesi i protocol de manteniment per l'aparell MegaBACE1000 (APBiotech), per tal de garantir una bona separació electroforètica dels *eTag reporter* i minimitzar el percentatge de mostres fallides.

En aquest estudi s'han analitzat un màxim de 10 SNPs, fent ús d'un total de 20 *eTag reporters* diferents. L'especificitat de la tècnica i el relativament ràpid desenvolupament del protocol suggereix que l'addició de més SNPs en la reacció és possible. Aquest increment de SNPs implica la utilització d'un major nombre d'*eTag reporters* els quals han de ser separats i identificats correctament. Actualment es disposa d'una llibreria de més d'un centenar d'*eTag reporters* amb mobilitats electroforètiques úniques. Ara bé, tenint en compte les condicions d'electroforesi emprades i els requisits actuals del programa eTag Informer per la identificació dels pics corresponents als *eTag reporter*, només es poden combinar un màxim de 25-30 *eTag reporters* per reacció. És a dir, un total de 12-15 SNPs per mostra i puet poden ser analitzats. Així, doncs, fent ús de l'aparell MegaBACE1000, es podrien obtenir uns 23.000 genotips per dia. Per tal d'incrementar el processament de genotips, s'estan avaluant d'altres aparells amb més capacitat analítica com el MegaBACE4000 (APBiotech) amb 384 capil·lars i 3730xl DNA Analyzer (Applied Biosystems) amb 96 capil·lars. És clar, que la capacitat multiplicadora actual del mètode *eTag Multiplex Invader* és llunyana a la de plataformes de genotipatge com la dels *microarrays*. No obstant, creiem que és una aproximació vàlida pel processament d'un nombre intermig-alt de mostres i és, a més, de fàcil implementació en tot laboratori que disposi de termocicladors i seqüenciadors automàtics.



## DISCUSSIÓ GENERAL

El tema central d'aquesta tesi ha estat l'anàlisi de les variants en el gen receptor domini discoidina 1 (*DDR1*) en l'esquizofrènia i de les eines necessàries per l'estudi d'aquesta variabilitat. En una primera part del treball, es va realitzar una anàlisi mutacional en *DDR1* en una mostra de malalts d'esquizofrènia i es va explorar la possibilitat que algunes d'aquestes s'associessin a la malaltia. La segona part va tenir un enfocament més metodològic i va consistir en desenvolupar i optimitzar un nou mètode de genotipatge, *eTag Multiplex Invader*, per l'anàlisi simultània d'SNPs.

L'aproximació a l'estudi de *DDR1* en l'esquizofrènia es va abordar mitjançant un estudi d'associació, amb la hipòtesi que les variants estudiades en *DDR1* estarien directament implicades en la malaltia o bé en desequilibri de lligament amb la variant de susceptibilitat. El disseny experimental concret va ser el d'un estudi d'associació de casos i controls. Aquest disseny ens era més convenient pel reclutament dels participants, tant de casos com de controls, que el d'estudi de nuclis familiars. Com ja s'ha comentat anteriorment, hi ha diversos factors limitants a tenir en compte en aquest tipus d'estudi, com són: l'estratificació poblacional, mida de la mostra insuficient o una descripció no acurada de la malaltia o tret d'estudi. Tots poden emmascarar els resultats derivats i en molts casos, tal com s'ha demostrat en la literatura, han estat la causa de falses associacions positives.

Som conscients de la mancança d'una eina de diagnòstic acurada per l'esquizofrènia. En el nostre estudi, el fet de que només s'inclouguessin pacients de la unitat de hospitalització de llarga durada garanteix, en certa manera, que els diagnòstics són consolidats. Es a dir, es tracta de pacients amb una evolució llarga de la malaltia que han estat estudiats durant anys per diferents especialistes. En el moment d'incloure'ls a l'estudi, un psiquiatra va verificar els diagnòstics. A més, el fet de tenir en compte d'altres variables clíniques (com per exemple, l'edat d'inici, el tipus d'inici i d'evolució de la malaltia), en l'anàlisi de les dades reforça els resultats en les relacions genotip-fenotip, tal com el nostre grup s'ha trobat en treballs anteriors (Martorell *et al*, 2001). D'altra banda, del grup d'individus controls, es van excloure aquells participants amb antecedents personals per patologia psiquiàtrica o amb una puntuació positiva pel qüestionari Goldberg de rastreig de patologia psiquiàtrica en població general.

Es va procurar de controlar la presència d'estratificació poblacional mitjançant dues mesures. Per una banda, per control demogràfic com criteri d'inclusió/exclusió i per aparellament de casos i controls amb un mateix origen geogràfic i ètnic. Per una altra, per control de l'estratificació genètica valorat segons el mètode de Pritchard i col.laboradors (Pritchard *et al*, 2000). Amb aquesta última evidenciem que la mostra estudiada era genèticament homogènia.

En el primer estudi, la utilització de la tècnica de la seqüenciació en *pools* o barreges de 10 mostres de DNA va permetre identificar un total de 17 variants: 16 SNPs (2

no-sinònims, 8 sinònims, 6 intrònics) i una deleció (IVS4-18delCT), de les quals 7 no han estat descrites anteriorment (Figura 10). Atès les limitacions del nostre grup, abans d'estudiar en detall totes les variants identificades, es va decidir de valorar l'associació d'algunes d'aquestes en una mostra de casos i controls (Estudi 2). L'Estudi 1 va ser, doncs, merament descriptiu. A més, la determinació de les freqüències al·lèliques de les diferents variants identificades no va ser possible ja que els resultats de la seqüenciació no eren quantificables. Una observació a destacar fou que dels 40 SNPs dipositats a la dbSNP i que es localitzen en les regions de *DDR1* analitzades, només se'n van detectar 10. Aquest fet suggeriria que o bé la resta de SNPs són poc freqüents (freqüència menor al 10% per l'al·lel minoritari), i no foren detectats donada la sensibilitat de la tècnica emprada, o bé no estaven presents en la mostra de malalts d'esquizofrènia estudiada. D'altra banda, la majoria d'aquests SNPs han estat identificats *in silico* i no han estat validats per genotipatge en múltiples mostres i, per tant, està per determinar si es tracta de veritables SNPs.

En el segon estudi, es va valorar si polimorfismes en el *DDR1* estaven associats a la incidència d'esquizofrènia en una mostra de població espanyola. Es van analitzar un total de 5 SNPs, 4 ubicats en *DDR1* i un en *GTF2H4* (Figura 16). Atès la proximitat física dels diferents marcadors analitzats, s'esperava trobar un fort LD entre totes les parelles de *loci* estudiades. Aquest no va ser el cas per la variant N502S. A més, el seu patró de LD fou diferent entre ambdós grups d'estudi. En malalts, l'associació d'aquesta variant amb la resta dels SNPs analitzats es trobava debilitada. Mentre que en controls, es mantenia amb un fort LD amb els SNP1, SNP2 i SNP4 i moderat amb l'SNP5. Aquesta fet explicava, en part, l'aparició d'un major nombre d'haplotips rars (freqüència menor a 0.5%) en el grup de malalts i les diferències estadísticament significatives observades en la distribució de les freqüències haplotípiques entre ambdós grups d'estudi. Tant en malalts com en controls, es va observar que més d'un 93% de la distribució haplotípica es concentrava en tan sols 3 combinacions al·lèliques (haplotips 1, 8, 16), de les quals l'haplotip majoritari estava compost per l'al·lel més comú de cada locus. La resta dels haplotips es trobaren amb freqüències menors a 1.5%.

La debilitació del LD entre parelles de *loci* molt properes en el cromosoma es pot explicar per: i. un fenomen de recombinació; ii. una mutació recurrent o iii. la conversió gènica. D'acord amb les dades publicades, *DDR1* es localitzaria en un bloc haplotípic o de LD contingut entre dos punts calents majors de recombinació (Cullen *et al*, 2002), la qual cosa explicaria l'observació de 3 haplotips majoritaris en la regió estudiada. En aquest treball desconeixem el motiu pel qual el LD s'ha perdut per la variant N502S en el grup de malalts. Ens crida l'atenció el fet d'haver observat aquesta variant només en heterozigosi. No es va detectar cap individu homozigot en casos ni en controls. Es podria especular que o bé es tracta d'una mutació nova o bé es tracta d'una variant que en homozigosi és inviable, la qual cosa explicaria perquè tot i estar en equilibri de Hardy-Weinberg, no la detectem ni en casos ni en controls.

El fet que els resultats de l'anàlisi individual dels SNPs no mostressin cap contribució al·lèlica o genotípica amb la malaltia o amb cap de les variables clíniques considerades i si que trobessim diferències estadísticament significatives en comparar les freqüències haplotípiques, posava de manifest quelcom ben establert que és que l'avaluació de la combinació d'al·lels o haplotips és més informativa que l'anàlisi independent dels marcadors. Malgrat no s'ha pogut trobar una correlació directa entre cap dels haplotips estimats i la malaltia, els resultats d'aquest treball semblen apuntar a que d'altres variants en el gen *DDR1* o en regions flanquejants podrien tenir un paper en l'etiologia de l'esquizofrènia.

A la llum dels resultats de l'Estudi 2, l'anàlisi bioinformàtica i funcional de la resta de variants identificades en l'Estudi 1 pot ser molt informativa i aportar noves dades. Sis d'aquestes variants es localitzaren en regions intròniques i es desconeix si tenen cap efecte en els mecanismes de *splicing* o en la maduració del transcrit. Així, per exemple, tot i que la deleció IVS4-18delCT no va ser estudiada en detall, els tres individus heterozigots identificats per aquesta deleció presentaren el mateix diploip, constituït pels haplotips 1 i 8. Serà d'interès analitzar aquestes variants i ampliar l'estudi d'haplotips en futurs estudis. D'altra banda, l'Estudi 1 va excloure l'anàlisi mutacional de les regions reguladores de l'expressió gènica, com regions promotores i les regions 5' i 3' UTR. Serà també interessant explorar la variabilitat d'aquestes regions i estudiar el seu efecte en l'expressió de *DDR1*.

En el tercer estudi es va explorar la possibilitat de combinar els fluorocroms *eTag reporter* amb la tècnica Invader per l'anàlisi de múltiples SNPs. Aquest és el primer treball on s'ha demostrat que el mètode anomenat *eTag Multiplex Invader SNP Assay* és una aproximació vàlida de genotipatge. Els resultats presentats van demostrar que aquest mètode és viable per l'anàlisi de SNPs a partir de DNA amplificat per PCR. En concret, es van desenvolupar un assaig per l'anàlisi simultània de 10 SNPs. En comparació amb la tècnica Invader, comporta dos passos addicionals: l'amplificació del DNA genòmic per PCR i la separació i detecció dels *eTag reporter* mitjançant la tècnica LIF-CE. No obstant, totes dues tècniques són d'ús general i rutinari en els laboratoris de genètica. Per tant, la posada punt del mètode *eTag Multiplex Invader* no requereix d'una instrumentació o infraestructura específica.

Malgrat demostrar que els *eTag reporter* es podien emprar per l'anàlisi simultània de SNPs per combinació amb la tècnica Invader, ACLARA BioSciences Inc. va decidir no continuar amb el projecte per una qüestió de mercat. Les tècniques de genotipatge d'SNPs a l'abast dels investigadors són múltiples, molt variades i algunes d'elles estan molt ben arrelades en la comunitat científica. Per una empresa petita com ACLARA BioSciences Inc., introduir un nou mètode en un mercat tan competitiu, com el de la detecció d'SNPs, era una operació d'alt risc que no es va voler assumir. Actualment, la prioritat de l'empresa rau en la utilització dels *eTag reporter* per la detecció simultània de proteïnes (de secreció o superfície cel·lular), de proteïnes

modificades (fosforilació) o d'interaccions entre elles (homo o heterodimerització de receptors, formació de complexos entre receptor i proteïnes acobladores).

En síntesi, en aquest treball s'han estudiat un grup de variants en el gen *DDR1* i s'ha observat que la distribució dels haplotips resultants és estadísticament diferent en pacients amb esquizofrènia que amb controls. En concret, en el grup de pacients s'observa un debilitament del LD en l'haplotip format per les variants analitzades. No podem concloure, amb els resultats recollits en aquest treball, si realment hi ha una variant en l'entorn *DDR1* de susceptibilitat per l'esquizofrènia o bé es tracta d'un artefacte. Estudis futurs ens ajudaran a esbrinar-ho.

## CONCLUSIONS

1. La tècnica de la seqüenciació en *pools* o mescles de 10 mostres de DNA és adequada per la identificació de polimorfismes d'un únic nucleòtid, malgrat la limitació observada per la detecció de mutacions rares o poc freqüents.
2. Es troben diferències estadísticament significatives en la distribució dels haplotips en la regió del gen receptor domini discoidina 1 (*DDR1*) entre els malalts d'esquizofrènia i individus controls estudiats. S'observa una major diversitat haplotípica, deguda a la presència d'haplotips rars o poc freqüents en el grup de malalts, suggerint la presència d'altres variants a *DDR1* o zones flanquejants les quals podrien tenir un efecte en l'etiologia de l'esquizofrènia. Podria ser que aquesta situació només s'acomplís per un grup dels malalts.
3. S'ha descartat la presència d'estratificació genètica en la mostra estudiada mitjançant l'aplicació del mètode descrit per Pritchard i col.laboradors.
4. La combinació dels *eTag reporter* amb la tècnica Invader resulta en un nou mètode de genotipatge acurat, de ràpid desenvolupament i propietats multiplicadores el qual no requereix d'una infraestructura o instrumentació específica.

## PERSPECTIVES FUTURES

L'estudi de la possible funció del gen *DDR1* en l'etiologia de l'esquizofrènia és una línia d'investigació en marxa en el nostre grup. Malgrat som conscients que aquesta hipòtesi de treball és arriscada, els resultats de l'estudi genètic que es presenten en aquest treball, conjuntament amb els dels estudis d'expressió que demostren la presència de *DDR1* en oligodendròcits (resultats en fase de publicació), són encoratjadors i ens conviden a continuar investigant.

En un futur pròxim i com a continuació dels resultats que s'han presentat en aquesta tesi, ens plantegem els següents objectius:

- Estudiar la deleció IVS4-18delCT en una mostra de malalts d'esquizofrènia i individus controls.
- Analitzar la presència de mutacions en les regions 5' i 3' UTR de *DDR1* en una mostra de malalts d'esquizofrènia.
- Determinar els haplotips moleculars de la regió del gen *DDR1* en nuclis familiars.
- Caracteritzar la variant N502S i determinar l'efecte del canvi 502S sobre la funcionalitat de la proteïna.

- Abdulhussein R, McFadden C, Fuentes-Prior P, Vogel WF. (2004). Exploring the collagen-binding site of the DDR1 tyrosine kinase receptor. *J Biol Chem* (published online ahead of print May 10)
- Akey JM, Sosnoski D, Parra E, Dios S, Hiester K, Su B, Bonilla C, Jin L, Shriver MD. (2001). Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *Biotechniques* 30, 358-62.
- Alexander S, Sydow LM, Wessels D, Soll DR. (1992). Discoidin proteins of Dictyostelium are necessary for normal cytoskeletal organization and cellular morphology during aggregation. *Differentiation* 51, 149-61.
- Ali SA, Pappas IS, Parnavelas JG. (1998). Collagen type IV promotes the differentiation of neuronal progenitors and inhibits astroglial differentiation in cortical cell cultures. *Brain Res Dev Brain Res* 110, 31-8.
- Alves F, Vogel W, Mossie K, Millauer B, Hofler H, Ullrich A. (1995). Distinct structural characteristics of discoidin I subfamily receptor tyrosine kinases and complementary expression in human cancer. *Oncogene* 10, 609-18.
- Alves F, Saupe S, Ledwon M, Schaub F, Hiddemann W, Vogel WF. (2001). Identification of two novel, kinase-deficient variants of discoidin domain receptor 1: differential expression in human colon cancer cell lines. *FASEB J* 15, 1321-3.
- Andreasen NC, Carpenter WT. (1993). Diagnosis and classification of schizophrenia. *Schizophr Bull* 19, 199-214.
- Antonarakis SE, Blouin JL, Pulver AE, Wolyniec P, Lasseter VK, Nestadt G, Kasch L, Babb R, Kazazian HH, Dombroski B, et al. (1995). Schizophrenia susceptibility and chromosome 6p24-22. *Nat Genet* 11, 235-6.
- Arolt V, Lencer R, Nolte A, Muller-Myhsok B, Purmann S, Schurmann M, Leutelt J, Pinnow M, Schwinger E. (1996). Eye tracking dysfunction is a putative phenotypic susceptibility marker of schizophrenia and maps to a locus on chromosome 6p in families with multiple occurrence of the disease. *Am J Med Genet* 67, 564-79.
- Arolt V, Lencer R, Purmann S, Schurmann M, Muller-Myhsok B, Krecker K, Schwinger E. (1999). Testing for linkage of eye tracking dysfunction and schizophrenia to markers on chromosomes 6, 8, 9, 20, and 22 in families multiply affected with schizophrenia. *Am J Med Genet* 88: 603-6.
- Bailer U, Leisch F, Meszaros K, Lenzinger E, Willinger U, Strobl R, Gebhardt C, Gerhard E, Fuchs K, Sieghart W, et al. (2000). Genome scan for susceptibility loci for schizophrenia. *Neuropsychobiology* 42, 175-82.
- Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, Nelson MP, Nagatomi J, Thomson G. (1997). Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am J Hum Genet* 61, 734-47.
- Barker KT, Martindale JE, Mitchell PJ, Kamalati T, Page MJ, Phippard DJ, Dale TC, Gusterson BA, Crompton MR. (1995). Expression patterns of the novel receptor-like tyrosine kinase, DDR, in human breast tumours. *Oncogene* 10, 569-75.
- Bartzokis G. (2002). Schizophrenia: breakdown in the well-regulated lifelong process of brain development and maturation. *Neuropsychopharmacology* 27, 672-83.
- Bassett AS, Chow EW. (1999). 22q11 deletion syndrome: a genetic subtype of schizophrenia. *Biol Psychiatry* 46, 882-91.

- Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. (1998). The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci* 7, 1626-31.
- Baxter GT, Radeke MJ, Kuo RC, Makrides V, Hinkle B, Hoang R, Medina-Selby A, Coit D, Valenzuela P, Feinstein SC. (1997). Signal transduction mediated by the truncated trkB receptor isoforms, trkB.T1 and trkB.T2. *J Neurosci* 17, 2683-90.
- Benes FM. (1989). Myelination of cortical-hippocampal relays during late adolescence. *Schizophr Bull* 15, 585-93.
- Benes FM, Turtle M, Khan Y, Farol P. (1994). Myelination of a key relay zone in the hippocampal formation occurs in the human brain during childhood, adolescence, and adulthood. *Arch Gen Psychiatry* 51, 477-484.
- Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH. (1994). Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* 369, 64-67.
- Bhatt RS, Tomoda T, Fang Y, Hatten ME. (2000). Discoidin domain receptor 1 functions in axon extension of cerebellar granule neurons. *Gene Dev.* 14, 2216-2228.
- Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ. (2001). Schizophrenia and affective disorders-cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet* 69, 428-33.
- Blazej RG, Paegel BM, Mathies RA. (2003). Polymorphism ratio sequencing: a new approach for single nucleotide polymorphism discovery and genotyping. *Genome Res* 13, 287-93.
- Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, Ullrich G, McGrath J, Kasch L, *et al.* (1998). Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 20, 70-3.
- Blume-Jensen P, Hunter T. (2001). Oncogenic kinase signaling. *Nature* 411, 355-365.
- Bogerts B, Falkai P, Greve B. (1991). Evidence of reduced temporolimbic structure volumes in schizophrenia. *Arch Gen Psychiatry* 48, 956-8.
- Boin F, Zanardini R, Pioli R, Altamura CA, Maes M, Gennarelli M. (2001). Association between -G308A tumor necrosis factor alpha gene polymorphism and schizophrenia. *Mol Psychiatry* 6, 79-82.
- Borg JP, Marchetto S, Le Bivic A, Ollendorff V, Jaulin-Bastard F, Saito H, Fournier E, Adelaide J, Margolis B, Birnbaum D. (2000). ERBIN: a basolateral PDZ protein that interacts with the mammalian ERBB2/HER2 receptor. *Nat Cell Biol* 2, 407-14.
- Braun A, Little DP, Koster H. (1997). Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem* 43, 1151-1158.
- Brunak S, Engelbrecht J, Knudsen S. (1991). Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* 220, 49-65.
- Busjahn A, Li GH, Faulhaber HD, Rosenthal M, Becker A, Jeschke E, Schuster H, Timmermann B, Hoehe MR, Luft FC. (2000). beta-2 adrenergic receptor gene



- variations, blood pressure, and heart size in normal twins. *Hypertension* 35, 555-560.
- Cannon TD, Kaprio J, Lonqvist J, Huttunen M, Koskenvuo M. (1998). The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study. *Arch Gen Psychiatry* 55, 67-74.
- Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC, *et al.* (1999). Heritability estimates for psychotic disorders: the Maudsley twin psychosis Series. *Arch Gen Psychiatry* 56, 162-8.
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22, 231-238.
- Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA. (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet* 33, 518-21.
- Carpenter WT, Kirkpatrick B. (1988). The heterogeneity of the long-term course of schizophrenia. *Schizophr Bull* 14, 645-52.
- Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci USA* 94, 1041-1046.
- Chana G, Landau S, Beasley C, Everall IP, Cotter D. (2003). Two-dimensional assessment of cytoarchitecture in the anterior cingulate cortex in major depressive disorder, bipolar disorder, and schizophrenia: evidence for decreased neuronal somal size and increased neuronal density. *Biol Psychiatry* 53, 1086-98.
- Chiu KC, Chuang LM, Yoon C. (2001). The A54T polymorphism at the intestinal fatty acid binding protein 2 is associated with insulin resistance in glucose tolerant Caucasians. *BMC Genet* 2, 7.
- Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, *et al.* (2002). Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl Acad Sci USA* 99, 13675-13680.
- Clark AG. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* 7, 111-122.
- Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, *et al.* (1998). Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am J Hum Genet* 63, 595-612.
- Combet C, Blanchet C, Geourjon C, Deléage G. (2000). NPS@: Network Protein Sequence Analysis. *TIBS* 25, 147-150.
- Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, Wallace RB. (1983). Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc Natl Acad Sci USA* 80, 278-282.
- Cotter DR, Pariante CM, Everall IP. Glial cell abnormalities in major psychiatric disorders: the evidence and implications. (2001). *Brain Res Bull* 55, 585-595.

- Craig T, Hwang MY, Bromet EJ. (2002). Obsessive-compulsive and panic symptoms in patients with first-admission psychosis. *Am J Psychiatry* 159, 592-598.
- Crow TJ. (1990). The continuum of psychosis and its genetic origins. The sixty-fifth Maudsley lecture. *Br J Psychiatry* 156, 788-797.
- Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M. (2002). High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am J Hum Genet* 71, 759-776.
- Curat CA, Eck M, Dervillez X, Vogel W. (2001). Mapping of epitopes in discoidin domain receptor 1 critical for collagen binding. *J Biol Chem* 276, 45952-45958.
- Curat CA, Vogel WF. (2002). Discoidin domain receptor 1 controls growth and adhesion of mesangial cells. *J Am Soc Nephrol* 13:2648-2656.
- De Stefano V, Dekou V, Nicaud V, Chasse JF, London J, Stansbie D, Humphries SE, Gudnason V. (1998). Linkage disequilibrium at the cystathionine beta synthase (CBS) locus and the association between genetic variation at the CBS locus and plasma levels of homocysteine. The Ears II Group. European Atherosclerosis Research Study. *Ann Hum Genet* 62, 481-490.
- Dejmek J, Dib K, Jonsson M, Andersson T. (2003). Wnt-5a and G-protein signaling are required for collagen-induced DDR1 receptor activation and normal mammary cell adhesion. *Int J Cancer* 103, 344-351.
- DeLisi LE, Hoff AL, Schwartz JE, Shields GW, Halthore SN, Gupta SM, Henn FA, Anand AK. (1991). Brain morphology in first-episode schizophrenic-like psychotic patients: a quantitative magnetic resonance imaging study. *Biol Psychiatry* 29, 159-75.
- DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, Wellman N, Loftus J, Nanthakumar B, Razi K, *et al.* (2002). A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am J Psychiatry* 159, 803-12.
- Devine JM, Williams JG. (1982). Characterization of sequence elements at the 5' end of a discoidin I gene isolated from *Dictyostelium discoideum*. *Nucleic Acids Res* 10, 1231-1241.
- Devlin B, Roeder K. (1999). Genomic control for association studies. *Biometrics* 55, 997-1004.
- Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B, Novasad D, Chu A,TCuenco K, Tiobek S, *et al.* (2002). Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. *Mol Psychiatry* 7, 689-94.
- Di Marco E, Cutuli N, Guerra L, Cancedda R, De Luca M. (1993). Molecular cloning of trkE, a novel trk-related putative tyrosine kinase receptor isolated from normal human keratinocytes and widely expressed by normal human tissues. *J Biol Chem* 268:24290-24295.
- Dionne IJ, Garant MJ, Nolan AA, Pollin TI, Lewis DG, Shuldiner AR, Poehlman ET. (2002). Association between obesity and a polymorphism in the beta(1)-adrenoceptor gene (Gly389Arg ADRB1) in Caucasian women. *Int J Obes Relat Metab Disord* 26, 633-639.

- Douglas JA, Boehnke M, Gillanders E, Trent JM, Gruber SB. (2001). Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat Genet* 28:361-364.
- Elvevag B, Goldberg TE. (2000). Cognitive impairment in schizophrenia is the core of the disorder. *Crit Rev Neurobiol* 14, 1-21.
- Engelke CE, Meini W, Boeing H, Glatt H. (2000). Association between functional genetic polymorphisms of human sulfotransferases 1A1 and 1A2. *Pharmacogenetics* 10, 163-169.
- Ensslin MA, Shur BD. (2003). Identification of mouse sperm SED1, a bimotif EGF repeat and discoidin-domain protein involved in sperm-egg binding. *Cell* 114, 405-417.
- Ewing B, Hillier L, Wendl MC, Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8, 175-185.
- Excoffier L, Slatkin M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 12, 921-927.
- Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A. (2000). Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 10, 853-860.
- Faraci E, Eck M, Gerstmayer B, Bosio A, Vogel WF. (2003). An extracellular matrix-specific microarray allowed the identification of target genes downstream of discoidin domain receptors. *Matrix Biol* 22, 373-381.
- Farmer AE, McGuffin P, Gottesman II. (1987). Twin concordance for DSM-III schizophrenia. Scrutinizing the validity of the definition. *Arch Gen Psychiatry* 44, 634-641.
- Fields RD, Stevens-Graham B. (2002). New insights into neuron-glia communication. *Science* 298, 556-62.
- Fisher RM, Humphries SE, Talmud PJ. (1997). Common variation in the lipoprotein lipase gene: effects on plasma lipids and risk of atherosclerosis. *Atherosclerosis* 135, 145-159.
- Foehr ED, Tatavos A, Tanabe E, Raffioni S, Goetz S, Dimarco E, De Luca M, Bradshaw RA. (2000). Discoidin domain receptor 1 (DDR1) signaling in PC12 cells: activation of juxtamembrane domains in PDGFR/DDR/TrkA chimeric receptors. *FASEB J* 14, 973-981.
- Fors L, Lieder KW, Vavra SH, Kwiatkowski RW. (2000). Large-scale SNP scoring from unamplified genomic DNA. *Pharmacogenomics* 1, 219-229.
- Freedman R, Adler LE, Leonard S. (1999). Alternative phenotypes for the complex genetics of schizophrenia. *Biol Psychiatry* 45, 551-8.
- Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, et al. (2004). Assessing the impact of population stratification on genetic association studies. *Nat Genet* 36, 388-393.
- Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, et al. (1995). A candidate genetic

risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 10, 111-113.

Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, *et al.* (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 98, 13784-13789.

Garner CC, Nash J, Haganir RL. (2000). PDZ domains in synapse assembly and signalling. *Trends Cell Biol* 10, 274-280.

Gershon ES, DeLisi LE, Hamovit J, Nurnberger JI, Maxwell ME, Schreiber J, Dauphinais D, Dingman CW, Guroff JJ. (1988). A controlled family study of chronic psychoses. Schizophrenia and schizoaffective disorder. *Arch Gen Psychiatry* 45, 328-336.

Glantz LA, Lewis DA. (2000). Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *Arch Gen Psychiatry* 57, 65-73.

Goetschy JF, Ulrich G, Aunis D, Ciesielski-Treska J. (1987). Fibronectin and collagens modulate the proliferation and morphology of astroglial cells in culture. *Int J Dev Neurosci* 5, 63-70.

Goldberg D P, Williams P. (1988). A User's Guide to the GHQ. Windsor: NFER-Nelson.

Goldberg TE, Egan MF, Gscheidle T, Coppola R, Weickert T, Kolachana BS, Goldman D, Weinberger DR. (2003). Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype and schizophrenia. *Arch Gen Psychiatry* 60, 889-896.

Gottesman I, Shields J. (1982). Schizophrenia: the epigenetic puzzle. Cambridge: Cambridge University Press.

Gottesman II, Bertelsen A. (1989). Confirming unexpressed genotypes for schizophrenia. Risks in the offspring of Fischer's Danish identical and fraternal discordant twins. *Arch Gen Psychiatry* 46, 867-872.

Gurling HM, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS, Read T, Murphy P, Blaveri E, McQuillin A, *et al.* (2001). Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. *Am J Hum Genet* 68, 661-673.

Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, Haroutunian V, Fienberg AA. (2001). Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc Natl Acad Sci USA* 98, 4746-4751.

Hall JM, LeDuc CA, Watson AR, Roter AH. (1996). An approach to high-throughput genotyping. *Genome Res* 6, 781-790.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22, 239-247.

- Hanayama R, Tanaka M, Miwa K, Shinohara A, Iwamatsu A, Nagata S. (2002). Identification of a factor that links apoptotic cells to phagocytes. *Nature* 417, 182-187.
- Harrison PJ. (1999). The neuropathology of schizophrenia. A critical review of the data and their interpretation. *Brain* 122, 593-624.
- Harrison PJ. (2004). The hippocampus in schizophrenia: a review of the neuropathological evidence and its pathophysiological implications. *Psychopharmacology* Mar 6 (on line).
- Hawley ME, Kidd KK. (1995). HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. *J Hered* 86, 409-11.
- He Z, Tessier-Lavigne M. (1997). Neuropilin is a receptor for the axonal chemorepellent Semaphorin III. *Cell* 90, 739-751.
- Hegele RA, Cao H. (2001). Single nucleotide polymorphisms of RXRA encoding retinoid X receptor alpha. *J Hum Genet* 46, 423-425.
- Heldin CH. (1995). Dimerization of cell surface receptors in signal transduction. *Cell* 80, 213-223.
- Henke W, Herdel K, Jung K, Schnorr D, Loening SA. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res* 25, 3957-3958.
- Hixson JE, Vernier DT. (1990). Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res* 31, 545-548.
- Ho BC, Andreasen NC, Nopoulos P, Arndt S, Magnotta V, Flaum M. (2003). Progressive structural brain abnormalities and their relationship to clinical outcome: a longitudinal magnetic resonance imaging study early in schizophrenia. *Arch Gen Psychiatry* 60, 585-594.
- Hof PR, Haroutunian V, Copland C, Davis KL, Buxbaum JD. (2002). Molecular and cellular evidence for an oligodendrocyte abnormality in schizophrenia. *Neurochem Res* 27, 1193-1200.
- Hof PR, Haroutunian V, Friedrich VL Jr, Byne W, Buitron C, Perl DP, Davis KL. (2003). Loss and altered spatial distribution of oligodendrocytes in the superior frontal gyrus in schizophrenia. *Biol Psychiatry* 53, 1075-1085.
- Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoller J, John A, Cascorbi I, Gerloff T, Roots I, Eichelbaum M, et al. (2000). Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci USA* 97, 3473-3478.
- Holliday R, Grigg GW. (1993). DNA methylation and mutation. *Mutat Res* 285, 61-67.
- Honer WG, Falkai P, Chen C, Arango V, Mann JJ, Dwork AJ. (1999). Synaptic and plasticity-associated proteins in anterior frontal cortex in severe mental illness. *Neuroscience* 91, 1247-1255.
- Hou G, Vogel W, Bendeck MP. (2001). The discoidin domain-receptor tyrosine kinase DDR1 in arterial wound repair. *J Clin Invest* 107, 727-735.

- Hou G, Vogel WF, Bendeck MP. (2002). Tyrosine kinase activity of discoidin domain receptor 1 is necessary for smooth muscle cell migration and matrix metalloproteinase expression. *Circ Res* 90, 1147-1149.
- Hovatta I, Lichtermann D, Juvonen H, Suvisaari J, Terwilliger JD, Arajarvi R, Kokko-Sahin ML, Ekelund J, Lonnqvist J, Peltonen L. (1998). Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. *Mol Psychiatry* 3, 452-457.
- Hwu HG, Lin MW, Lee PC, Lee SF, Ou-Yang WC, Liu CM. (2000). Evaluation of linkage of markers on chromosome 6p with schizophrenia in Taiwanese families. *Am J Med Genet* 96, 74-78.
- Ji J, Chen X, Leung SY, Chi JT, Chu KM, Yuen ST, Li R, Chan AS, Li J, Dunphy N, So S. (2002). Comprehensive analysis of the gene expression profiles in human gastric cancer cell lines. *Oncogene* 21, 6549-6556.
- Johnson JD, Edman JC, Rutter WJ. (1993). A receptor tyrosine kinase found in breast carcinoma cells has an extracellular discoidin I-like domain. *Proc Natl Acad Sci USA* 90, 10891.
- Jonsson M, Andersson T. (2001). Repression of Wnt-5a impairs DDR1 phosphorylation and modifies adhesion and migration of mammary cells. *J Cell Sci* 114, 2043-2053.
- Joober R, Zarate JM, Rouleau GA, Skamene E, Boksa P. (2002). Provisional mapping of quantitative trait loci modulating the acoustic startle response and prepulse inhibition of acoustic startle. *Neuropsychopharmacology* 27, 765-781.
- Jouanolle AM, Fergelot P, Gandon G, Yaouanq J, Le Gall JY, David V. (1997). A candidate gene for hemochromatosis: frequency of the C282Y and H63D mutations. *Hum Genet* 100, 544-547.
- Kafatos FC, Jones CW, Efstratiadis A. (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res* 7, 1541-1552.
- Kamohara H, Yamashiro S, Galligan C, Yoshimura T. (2001). Discoidin domain receptor 1 isoform-a (DDR1alpha) promotes migration of leukocytes in three-dimensional collagen lattices. *FASEB J* 15, 2724-2726.
- Kane WH, Davie EW. (1986). Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. *Proc Natl Acad Sci USA* 83, 6800-6804.
- Katila H, Hanninen K, Hurme M. (1999). Polymorphisms of the interleukin-1 gene complex in schizophrenia. *Mol Psychiatry* 4, 179-181.
- Kawanishi Y, Harada S, Tachikawa H, Okubo T, Shiraishi H. (2000). Novel polymorphisms of the AP-2 gene (6p24): analysis of association with schizophrenia. *J Hum Genet* 45, 24-30.
- Kealey C, Reynolds A, Mynett-Johnson L, Claffey E, McKeon P. (2001). No evidence to support an association between the oestrogen receptor beta gene and bipolar disorder. *Psychiatr Genet* 11, 223-226.

- Kendler KS, Gruenberg AM, Tsuang MT. (1985). Psychiatric illness in first-degree relatives of schizophrenic and surgical control patients. A family study using DSM-III criteria. *Arch Gen Psychiatry* 42, 770-779.
- Kirov G, Ivanov D, Williams NM, Preece A, Nikolov I, Milev R, Koleva S, Dimitrova A, Toncheva D, O'Donovan MC, et al. (2004). Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol Psychiatry* 55, 971-5.
- Kittles RA, Chen W, Panguluri RK, Ahaghotu C, Jackson A, Adebamowo CA, Griffin R, Williams T, Ukoli F, Adams-Campbell L, et al. (2002). CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet* 110, 553-560.
- Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. (1988). Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 43, 520-526.
- Kobayashi S, Inoue S, Hosoi T, Ouchi Y, Shiraki M, Orimo H. (1996). Association of bone mineral density with polymorphism of the estrogen receptor gene. *J Bone Miner Res* 1, 306-311.
- Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC. (1998). MITOMAP: a human mitochondrial genome database--1998 update. *Nucleic Acids Res* 26, 112-115.
- Kolodkin AL, Levengood DV, Rowe EG, Tai YT, Giger RJ, Ginty DD. (1997). Neuropilin is a semaphorin III receptor. *Cell* 90, 753-762.
- Kuimelis RG, Livak KJ, Mullah B, Andrus A. (1997). Structural analogues of TaqMan probes for real-time quantitative PCR. *Nucleic Acids Symp Ser* 37, 255-256.
- Kwok PY, Carlson C, Yager TD, Ankener W, Nickerson DA. (1994). Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* 23, 138-144.
- Kwok PY, Chen X. (2003). Detection of single nucleotide polymorphisms. *Curr Issues Mol Biol* 5, 43-60.
- Labrador JP, Azcoitia V, Tuckermann J, Lin C, Olaso E, Manes S, Bruckner K, Goergen JL, Lemke G, Yancopoulos G, et al. (2001). The collagen receptor DDR2 regulates proliferation and its elimination leads to dwarfism. *EMBO J* 2, 446-452.
- Lai C, Lemke G. (1994). Structure and expression of the Tyro 10 receptor tyrosine kinase. *Oncogene* 9, 877-883.
- Lander ES, Schork NJ. (1994). Genetic dissection of complex traits. *Science* 265, 2037-2048.
- Lander E, Kruglyak L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11, 241-247.
- Landegren U, Kaiser R, Sanders J, Hood L. (1988). A ligase-mediated gene detection technique. *Science* 241, 1077-1080.
- Lara DR, Gama CS, Belmonte-de-Abreu P, Portela LV, Goncalves CA, Fonseca M, Hauck S, Souza DO. (2001). Increased serum S100B protein in schizophrenia: a study in medication-free patients. *J Psychiatr Res* 35, 11-14.

- Layne MD, Endege WO, Jain MK, Yet SF, Hsieh CM, Chin MT, Perrella MA, Blonar MA, Haber E, Lee ME. (1998). Aortic carboxypeptidase-like protein, a novel protein with discoidin and carboxypeptidase-like domains, is up-regulated during vascular smooth muscle cell differentiation. *J Biol Chem* 273, 15654-15660.
- Laval S, Butler R, Shelling AN, Hanby AM, Poulsom R, Ganesan TS. (1994). Isolation and characterization of an epithelial-specific receptor tyrosine kinase from an ovarian cancer cell line. *Cell Growth Differ* 5, 1173-1183.
- Le Hellard S, Ballereau SJ, Visscher PM, Torrance HS, Pinson J, Morris SW, Thomson ML, Semple CA, Muir WJ, Blackwood DH, *et al.* (2002). SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis. *Nucleic Acids Res* 30, e74.
- Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, Hayward NK, Crowe RR, Andreasen NC, Black DW, *et al.* (1998). Genome scan of schizophrenia. *Am J Psychiatry* 155, 741-750.
- Lewine RR. (1980). Sex differences in age of symptom onset and first hospitalization in schizophrenia. *Am J Orthopsychiatry* 50, 316-322.
- L'hote CG, Thomas PH, Ganesan TS. (2002). Functional analysis of discoidin domain receptor 1: effect of adhesion on DDR1 phosphorylation. *FASEB J* 16, 234-236.
- Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. (1996). Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol* 5, 182-187.
- Lieber MR. (1997). The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays* 19, 233-240.
- Lindholm E, Ekholm B, Balciuniene J, Johansson G, Castensson A, Koisti M, Nylander PO, Pettersson U, Adolfsson R, Jazin E. (1999). Linkage analysis of a large Swedish kindred provides further support for a susceptibility locus for schizophrenia on chromosome 6p23. *Am J Med Genet* 88, 369-377.
- Livak KJ, Marmaro J, Todd JA. (1995). Towards fully automated genome-wide polymorphism screening. *Nat Genet* 9, 341-342.
- Livak KJ. (1999). Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* 14, 143-149.
- Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* 19, 225-232.
- Longo MC, Berninger MS, Hartley JL. (1990). Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* 93, 125-128.
- Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, *et al.* (1999). Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol* 17, 292-296.
- Lyamichev VI, Kaiser MW, Lyamicheva NE, Vologodskii AV, Hall JG, Ma WP, Allawi HT, Neri BP. (2000). Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry* 39, 9523-9532.



- Lyamichev V, Neri B. (2003). Invader assay for SNP genotyping. *Methods Mol Biol* 212, 229-240.
- Martinou JC, Falls DL, Fischbach GD, Merlie JP. (1991). Acetylcholine receptor-inducing activity stimulates expression of the epsilon-subunit gene of the muscle acetylcholine receptor. *Proc Natl Acad Sci U S A* 88, 7669-7673.
- Martorell L, Virgos C, Valero J, Coll G, Figuera L, Joven J, Pocovi M, Labad A, Vilella E. (2001). Schizophrenic women with the APOE epsilon 4 allele have a worse prognosis than those without it. *Mol Psychiatry* 6, 307-310.
- Matsuyama W, Faure M, Yoshimura T. (2003). Activation of discoidin domain receptor 1 facilitates the maturation of human monocyte-derived dendritic cells through the TNF receptor associated factor 6/TGF-beta-activated protein kinase 1 binding protein 1 beta/p38 alpha mitogen-activated protein kinase signaling cascade. *J Immunol* 171, 3520-3532.
- Matsuyama W, Kamohara H, Galligan C, Faure M, Yoshimura T. (2003b). Interaction of discoidin domain receptor 1 isoform b (DDR1b) with collagen activates p38 mitogen-activated protein kinase and promotes differentiation of macrophages. *FASEB J* 17, 1286-1288.
- Matsuyama W, Wang L, Farrar WL, Faure M, Yoshimura T. (2004). Activation of discoidin domain receptor 1 isoform b with collagen up-regulates chemokine production in human macrophages: role of p38 mitogen-activated protein kinase and NF-kappaB. *J Immunol* 172, 2332-2340.
- Maziade M, Bissonnette L, Rouillard E, Martinez M, Turgeon M, Charron L, Pouliot V, Boutin P, Cliche D, Dion C, et al. (1997). 6p24-22 region and major psychoses in the Eastern Quebec population. Le Groupe IREP. *Am J Med Genet* 74, 311-318.
- Maziade M, Roy MA, Rouillard E, Bissonnette L, Fournier JP, Roy A, Garneau Y, Montgrain N, Potvin A, Cliche D, et al. (2001). A search for specific and common susceptibility loci for schizophrenia and bipolar disorder: a linkage study in 13 target chromosomes. *Mol Psychiatry* 6, 684-693.
- McCarley RW, Wible CG, Frumin M, Hirayasu Y, Levitt JJ, Fischer IA, Shenton ME. (1999). MRI anatomy of schizophrenia. *Biol Psychiatry* 45, 1099-1119.
- Mein CA, Barratt BJ, Dunn MG, Siegmund T, Smith AN, Esposito L, Nutland S, Stevens HE, Wilson AJ, Phillips MS, et al. (2000). Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res* 10, 330-343.
- Meira-Lima IV, Pereira AC, Mota GF, Floriano M, Araujo F, Mansur AJ, Krieger JE, Vallada H. (2003). Analysis of a polymorphism in the promoter region of the tumor necrosis factor alpha gene in schizophrenia and bipolar disorder: further support for an association with schizophrenia. *Mol Psychiatry* 8, 718-720.
- Michalatos-Beloin S, Tishkoff SA, Bentley KL, Kidd KK, Ruano G. (1996). Molecular haplotyping of genetic markers 10 kb apart by allele-specific long-range PCR. *Nucleic Acids Res* 24, 4841-4843.
- Miller SA, Dykes DD, Polesky HF. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16, 1215.
- Mimmack ML, Ryan M, Baba H, Navarro-Ruiz J, Iritani S, Faull RL, McKenna PJ, Jones PB, Arai H, Starkey M, et al. (2002). Gene expression analysis in schizophrenia:

- reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proc Natl Acad Sci USA* 99, 4680-4685.
- Mirnics K, Middleton FA, Marquez A, Lewis DA, Levitt P. (2000). Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* 28, 53-67.
- Mohan RR, Mohan RR, Wilson SE. (2001). Discoidin domain receptor (DDR) 1 and 2: collagen-activated tyrosine kinase receptors in the cornea. *Exp Eye Res* 72, 87-92.
- Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, Blackwood D, Liu X, Sjogren B, *et al.* (1995). An international two-stage genome-wide search for schizophrenia susceptibility genes. *Nat Genet* 11, 321-324.
- Moises HWM. (2001). Human Genome data analyzed by an evolutionary method suggests a decrease in protein-synthesis rate as cause of schizophrenia and an increase as antipsychotic mechanism. [<http://xxx.arxiv.cornell.edu/abs/condmat/0110189>] *ArXiv.org e-Print archive*.
- Moises HW, Zoega T, Gottesman II. (2002). The glial growth factors efficiency and synaptic destabilization hypothesis of schizophrenia. *BMC Psychiatry* 2, 8-20.
- Morton NE, Collins A. (1998). Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci USA* 95, 11389-11393.
- Murphy KC, Jones LA, Owen MJ. (1999). High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* 56, 940-945.
- Murphy KC. (2002). Schizophrenia and velo-cardio-facial syndrome. *Lancet* 359, 426-430.
- Myles-Worsley M, Coon H, McDowell J, Brenner C, Hoff M, Lind B, Bennett P, Freedman R, Clementz B, Byerley W. (1999). Linkage of a composite inhibitory phenotype to a chromosome 22q locus in eight Utah families. *Am J Med Genet* 88, 544-550.
- Nakayama EE, Wasi C, Ajisawa A, Iwamoto A, Shioda T. (1996). A new polymorphism in the promoter region of the human interleukin-16 (IL-16) gene. *Genes Immun* 1, 293-294.
- Namekata K, Oyama F, Imagawa M, Ihara Y. Human transferrin (Tf): a single mutation at codon 570 determines Tf C1 or Tf C2 variant. (1997). *Hum Genet* 100, 457-458.
- Nemoto T, Ohashi K, Akashi T, Johnson JD, Hirokawa K. (1997). Overexpression of protein tyrosine kinases in human esophageal cancer. *Pathobiology* 65, 195-203.
- Nevillie M, Selzer R, Aizenstein B, Maguire M, Hogan K, Walton R, Welsh K, Neri B, de Arruda M. (2002). Characterization of cytochrome P450 2D6 alleles using the Invader system. *Biotechniques* Suppl: 34-8, 40-43.
- Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 17, 2503-2516.

- Nikiforov TT, Rendle RB, Goelet P, Rogers YH, Kotewicz ML, Anderson S, Trainor GL, Knapp MR. (1994). Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res* 22, 4167-4175.
- Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 265, 2085-2088.
- Nopoulos P, Torres I, Flaum M, Andreasen NC, Ehrhardt JC, Yuh WT. (1995). Brain morphology in first-episode schizophrenia. *Am J Psychiatry* 152, 1721-1723.
- Nyren P, Pettersson B, Uhlen M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem* 208, 171-175.
- Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y. (2001). A high-throughput SNP typing system for genome-wide association studies. *J Hum Genet* 46, 471-477.
- Okayama H, Curiel DT, Brantly ML, Holmes MD, Crystal RG. (1989). Rapid, nonradioactive detection of mutations in the human genome by allele-specific amplification. *J Lab Clin Med* 114, 105-113.
- Olaso E, Ikeda K, Eng FJ, Xu L, Wang LH, Lin HC, Friedman SL. (2001). DDR2 receptor promotes MMP-2-mediated proliferation and invasion by hepatic stellate cells. *J Clin Invest* 108, 1369-1378.
- Olivier M, Chuang LM, Chang MS, Chen YT, Pei D, Ranade K, de Witte A, Allen J, Tran N, Curb D, et al. (2002). High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res* 30, e53.
- Ongusaha PP, Kim JI, Fang L, Wong TW, Yancopoulos GD, Aaronson SA, Lee SW. (2003). p53 induction and activation of DDR1 kinase counteract p53-mediated apoptosis and influence p53 regulation through a positive feedback loop. *EMBO J* 22, 1289-1301.
- Onstad S, Skre I, Torgersen S, Kringlen E. (1991). Twin concordance for DSM-III-R schizophrenia. *Acta Psychiatr Scand* 83, 395-401.
- Orum H, Jakobsen MH, Koch T, Vuust J, Borre MB. (1999). Detection of the factor V Leiden mutation by direct allele-specific hybridization of PCR amplicons to photoimmobilized locked nucleic acids. *Clin Chem* 45, 1898-1905.
- Owen F, Simpson M. (1994). The neurochemistry of schizophrenia. *Mol Cell Biol Hum Dis Ser* 4, 133-159.
- Palko ME, Coppola V, Tessarollo L. (1999). Evidence for a role of truncated trkC receptor isoforms in mouse development. *J Neurosci* 19, 775-782.
- Palmer AA, Breen LL, Flodman P, Conti LH, Spence MA, Printz MP. (2003). Identification of quantitative trait loci for prepulse inhibition in rats. *Psychopharmacology* 165, 270-279.
- Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvanen AC. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res* 7, 606-614.

- Paunio T, Ekelund J, Varilo T, Parker A, Hovatta I, Turunen JA, Rinard K, Foti A, Terwilliger JD, Juvonen H, *et al.* (2001). Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. *Hum Mol Genet* 10, 3037-3048.
- Paus T, Zijdenbos A, Worsley K, Collins DL, Blumenthal J, Giedd JN, Rapoport JL, Evans AC. (1999). Structural maturation of neural pathways in children and adolescents: in vivo study. *Science* 283, 1908-1911.
- Pearlson GD, Marsh L. (1999). Structural brain imaging in schizophrenia: a selective review. *Biol Psychiatry* 46, 627-49.
- Peles E, Nativ M, Lustig M, Grumet M, Schilling J, Martinez R, Plowman GD, Schlessinger J. (1997). Identification of a novel contactin-associated transmembrane receptor with multiple domains implicated in protein-protein interactions. *EMBO J* 16, 978-988.
- Perez JL, Jing SQ, Wong TW. (1996). Identification of two isoforms of the Cak receptor kinase that are coexpressed in breast tumor cell lines. *Oncogene* 12, 1469-1477.
- Pettegrew JW, Keshavan MS, Panchalingam K, Strychor S, Kaplan DB, Tretta MG, Allen M. (1991). Alterations in brain high-energy phosphate and membrane phospholipid metabolism in first-episode, drug-naive schizophrenics. A pilot study of the dorsal prefrontal cortex by in vivo phosphorus 31 nuclear magnetic resonance spectroscopy. *Arch Gen Psychiatry* 48, 563-568.
- Playford MP, Butler RJ, Wang XC, Katso RM, Cooke IE, Ganesan TS. (1996). The genomic structure of discoidin receptor tyrosine kinase. *Genome Res* 6, 620-627.
- Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, *et al.* (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415, 436-442.
- Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. (1996). A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood* 88, 3698-3703.
- Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ. (2001). Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 11, 152-162.
- Pritchard JK, Rosenberg NA. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Gen* 65, 220-228.
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. (2000). Association mapping in structured populations. *Am J Hum Genet* 67, 170-1781.
- Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin JL, Kimberland M, Babb R, Vourlis S, Chen H, *et al.* (1995). Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am J Med Genet* 60, 252-260.
- Reese MG, Eeckman FH, Kulp D, Haussler D. (1997). Improved splice site detection in Genie. *J Comput Biol* 4, 311-323.

- Reich DE, Gabriel SB, Altshuler D. (2003). Quality and completeness of SNP databases. *Nat Genet* 33, 457-8.
- Risch N, Merikangas K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516-1527.
- Ross PL, Lee K, Belgrader P. (1997). Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal Chem* 69, 4197-4202.
- Rothermundt M, Missler U, Arolt V, Peters M, Leadbeater J, Wiesmann M, Rudolf S, Wandinger KP, Kirchner H. (2001). Increased S100B blood levels in unmedicated and treated schizophrenic patients are correlated with negative symptomatology. *Mol Psychiatry* 6, 445-459.
- Rovin BH, Lu L, Saxena R. A novel polymorphism in the MCP-1 gene regulatory region that influences MCP-1 expression. (1999). *Biochem Biophys Res Commun* 259, 344-348.
- Rozen S, Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132, 365-386.
- Rudin E. (1916). Zur Vererbung und Neuentslehung der dementia praecox. Berlin: Springer.
- Ruvkun G, Hobert O. (1998). The taxonomy of developmental control in *Caenorhabditis elegans*. *Science* 282, 2033-2041.
- Ryan D, Nuccie B, Arvan D. (1999). Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. *Mol Diagn* 4, 135-144.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, *et al.* (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928-933.
- Saiki RK, Walsh PS, Levenson CH, Erlich HA. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc Natl Acad Sci USA* 86, 6230-6234.
- Sakamoto O, Suga M, Suda T, Ando M. (2001). Expression of discoidin domain receptor 1 tyrosine kinase on the human bronchial epithelium. *Eur Respir J* 17, 969-974.
- Sakuma S, Saya H, Tada M, Nakao M, Fujiwara T, Roth JA, Sawamura Y, Shinohe Y, Abe H. (1996). Receptor protein tyrosine kinase DDR is up-regulated by p53 protein. *FEBS Lett* 398, 165-169.
- Sanchez MP, Tapley P, Saini SS, He B, Pulido D, Barbacid M. (1994). Multiple tyrosine protein kinases in rat hippocampal neurons: isolation of Ptk-3, a receptor expressed in proliferative zones of the developing brain. *Proc Natl Acad Sci USA* 91, 1819-1823.
- Sanchez-Cespedes M, Cairns P, Jen J, Sidransky D. (1998). Degenerate oligonucleotide-primed PCR (DOP-PCR): evaluation of its reliability for screening of genetic alterations in neoplasia. *Biotechniques* 25, 1036-1038.

- Sauer CG, Gehrig A, Warneke-Wittstock R, Marquardt A, Ewing CC, Gibson A, Lorenz B, Jurklies B, Weber BH. (1997). Positional cloning of the gene associated with X-linked juvenile retinoschisis. *Nat Genet* 17, 164-170.
- Schneider S, Roessli D, Excoffier L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.
- Sham P, Bader JS, Craig I, O'Donovan M, Owen M. (2002). DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet* 3, 862-871.
- Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8. (1996). Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. *Am J Med Genet* 67, 580-594.
- Schultz SK, Andreasen NC. Schizophrenia. (1999). *Lancet* 353, 1425-1430.
- Schwab SG, Albus M, Hallmayer J, Honig S, Borrmann M, Lichtermann D, Ebstein RP, Ackenheil M, Lerer B, Risch N, *et al.* (1995). Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nat Genet* 11, 325-327.
- Schwab SG, Hallmayer J, Albus M, Lerer B, Eckstein GN, Borrmann M, Segman RH, Hanses C, Freymann J, Yakir A, *et al.* (2000). A genome-wide autosomal screen for schizophrenia susceptibility loci in 71 families with affected siblings: support for loci on chromosome 10p and 6. *Mol Psychiatry* 5, 638-649.
- Schwab SG, Hallmayer J, Freimann J, Lerer B, Albus M, Borrmann-Hassenbach M, Segman RH, Trixler M, Rietschel M, Maier W, *et al.* (2002). Investigation of linkage and association/linkage disequilibrium of HLA A-, DQA1-, DQB1-, and DRB1-alleles in 69 sib-pair- and 89 trio-families with schizophrenia. *Am J Med Genet* 114, 315-320.
- Schwab SG, Knapp M, Mondabon S, Hallmayer J, Borrmann-Hassenbach M, Albus M, Lerer B, Rietschel M, Trixler M, Maier W, *et al.* (2003). Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet* 72, 185-190.
- Seidman LJ, Faraone SV, Goldstein JM, Kremen WS, Horton NJ, Makris N, Toomey R, Kennedy D, Caviness VS, Tsuang MT. (2002). Left hippocampal volume as a vulnerability indicator for schizophrenia: a magnetic resonance imaging morphometric study of nonpsychotic first-degree relatives. *Arch Gen Psychiatry* 59, 839-849.
- Sham P, Bader JS, Craig I, O'Donovan M, Owen M. (2002). DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet* 3, 862-871.
- Shrivastava A, Radziejewski C, Campbell E, Kovac L, McGlynn M, Ryan TE, Davis S, Goldfarb MP, Glass DJ, Lemke G, *et al.* (1997). An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. *Mol Cell* 1, 25-34.
- Singh S. (2000). Nucleic acid detection using degradation of a tagged sequence. USPTO 6,514,700.
- Smith HO, Wilcox KW. (1970). A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J Mol Biol* 51, 379-391.

- Smythies JR. (1983). The transmethylation and one-carbon cycle hypotheses of schizophrenia. *Psychol Med* 13, 711-714.
- Sommer SS, Cassady JD, Sobell JL, Bottema CD. (1989). A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. *Mayo Clin Proc* 64, 1361-1372.
- Staal WG, Hulshoff Pol HE, Schnack HG, Hoogendoorn ML, Jellema K, Kahn RS. (2000). Structural brain abnormalities in patients with schizophrenia and their healthy siblings. *Am J Psychiatry* 157, 416-421.
- Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT, *et al.* (2002). Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 71, 877-892.
- Stephens M, Smith NJ, Donnelly P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68, 978-989.
- Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, *et al.* (1995). A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. *Nat Genet* 11, 287-293.
- Straub RE, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, O'Neill FA, *et al.* (2002). Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes. *Mol Psychiatry* 7, 542-759.
- Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, *et al.* (2002b). Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet* 71, 337-348.
- Syvänen AC, Aalto-Setälä K, Harju L, Kontula K, Soderlund H. (1990). A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* 8, 684-692.
- Syvänen AC. (1999). From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat* 13, 1-10.
- Taillon-Miller P, Piernot EE, Kwok PY. (1999). Efficient approach to unique single-nucleotide polymorphism discovery. *Genome Res* 9, 499-505.
- Talbot K, Eidem WL, Tinsley CL, Benson MA, Thompson EW, Smith RJ, Hahn CG, Siegel SJ, Trojanowski JQ, Gur RE, *et al.* (2004). Dysbindin-1 is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. *J Clin Invest* 113, 1353-1363.
- Talmud PJ, Hawe E, Robertson K, Miller GJ, Miller NE, Humphries SE. (2002). Genetic and environmental determinants of plasma high density lipoprotein cholesterol and apolipoprotein AI concentrations in healthy middle-aged men. *Ann Hum Genet* 66, 111-124.
- Tang JX, Zhou J, Fan JB, Li XW, Shi YY, Gu NF, Feng GY, Xing YL, Shi JG, He L. (2003). Family-based association study of DTNBP1 in 6p22.3 and schizophrenia. *Mol Psychiatry* 8, 717-718.

- Terwilliger JD, Weiss KM. (1998). Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 9, 578-594.
- The international classification of diseases. 9th revision, clinical modification; ICD-9-CM. (1979). Geneva, Switzerland: World Health Organization.
- The MHC sequencing consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 401, 921-923.
- The Retinoschisis Consortium. (1998). Functional implications of the spectrum of mutations found in 234 cases with X-linked juvenile retinoschisis. The Retinoschisis Consortium. *Hum Mol Genet* 7, 1185-1192.
- Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB, Starkey M, Webster MJ, Yolken RH, Bahn S. (2003). Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* 362, 798-805.
- Torrey EF, Peterson MR. (1976). The viral hypothesis of schizophrenia. *Schizophr Bull* 2, 136-146.
- Tsai MT, Su YC, Chen YH, Chen CH. (2001). Lack of evidence to support the association of the human prion gene with schizophrenia. *Mol Psychiatry* 6, 74-78.
- Tsuang MT, Lyons MJ, Faraone SV. (1990). Heterogeneity of schizophrenia. Conceptual models and analytic strategies. *Br J Psychiatry* 156,17-26.
- Turki J, Pak J, Green SA, Martin RJ, Liggett SB. (1995). Genetic polymorphisms of the beta-2-adrenergic receptor in nocturnal and nonnocturnal asthma: evidence that gly16 correlates with the nocturnal phenotype. *J Clin Invest* 95, 1635-1641.
- Tyagi S, Kramer FR. (1996). Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* 14, 303-308.
- Tyagi S, Bratu DP, Kramer FR. (1998). Multicolor molecular beacons for allele discrimination. *Nat Biotechnol* 16, 49-53.
- Uranova N, Orlovskaya D, Vikhreva O, Zimina I, Kolomeets N, Vostrikov V, Rachmanova V. (2001). Electron microscopy of oligodendroglia in severe mental illness. *Brain Res Bull* 55, 597-610.
- Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI. (2004). Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. *Schizophr Res* 67, 269-275.
- Valent A, Meddeb M, Danglot G, Duverger A, Nguyen VC, Bernheim A. (1996). Assignment of the NTRK4 (trkE) gene to chromosome 6p21. *Hum Genet* 98, 12-15.
- Van Den Bogaert A, Schumacher J, Schulze TG, Otte AC, Ohlraun S, Kovalenko S, Becker T, Freudenberg J, Jonsson EG, Mattila-Evenden M, et al. (2003). The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease. *Am J Hum Genet* 73, 1438-1443.
- van den Oord EJ, Sullivan PF, Jiang Y, Walsh D, O'Neill FA, Kendler KS, Riley BP. (2003). Identification of a high-risk haplotype for the dystrobrevin binding protein



- 1 (DTNBP1) gene in the Irish study of high-density schizophrenia families. *Mol Psychiatry* 8, 499-510.
- van Eijk MJ, Broekhof JL, van der Poel HJ, Hogers RC, Schneiders H, Kamerbeek J, Verstege E, van Aart JW, Geerlings H, Buntjer JB, *et al.* (2004). SNPWave: a flexible multiplexed SNP genotyping technology. *Nucleic Acids Res* 32, e47.
- Vawter MP, Crook JM, Hyde TM, Kleinman JE, Weinberger DR, Becker KG, Freed WJ. (2002). Microarray analysis of gene expression in the prefrontal cortex in schizophrenia: a preliminary study. *Schizophr Res* 58, 11-20.
- Virgos C, Martorell L, Simo JM, Valero J, Figuera L, Joven J, Labad A, Vilella E. (1999). Plasma homocysteine and the methylenetetrahydrofolate reductase C677T gene variant: lack of association with schizophrenia. *Neuroreport* 10, 2035-2038.
- Virgos C, Martorell L, Valero J, Figuera L, Civeira F, Joven J, Labad A, Vilella E. (2001). Association study of schizophrenia with polymorphisms at six candidate genes. *Schizophr Res* 49, 65-71.
- Virgos C, Cronin MT. (2002). Universal e-Tag primer compositions and methods. PCT/US02/32867 (*pendent*).
- Vogel W, Gish GD, Alves F, Pawson T. (1997). The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol Cell* 1, 13-23.
- Vogel W. (1999). Discoidin domain receptors: structural relations and functional implications. *FASEB J* 13, S77-S82.
- Vogel W, Brakebusch C, Fassler R, Alves F, Ruggiero F, Pawson T. (2000). Discoidin domain receptor 1 is activated independently of beta(1) integrin. *J Biol Chem* 275, 5779-5784.
- Vogel WF, Aszodi A, Alves F, Pawson T. (2001). Discoidin domain receptor 1 tyrosine kinase has an essential role in mammary gland development. *Mol Cell Biol* 21, 2906-2917.
- Vogel WF. (2001). Collagen-receptor signaling in health and disease. *Eur J Dermatol* 11, 506-514.
- Vogel WF. (2002). Ligand-induced shedding of discoidin domain receptor 1. *FEBS Lett* 514, 175-180.
- Vohl MC, Lepage P, Gaudet D, Brewer CG, Betard C, Perron P, Houde G, Cellier C, Faith JM, Despres JP, *et al.* (2000). Molecular scanning of the human PPAR $\alpha$  gene: association of the L162v mutation with hyperapobetalipoproteinemia. *J Lipid Res* 41, 945-952.
- Wang S, Sun CE, Walczak CA, Ziegler JS, Kipps BR, Goldin LR, Diehl SR. (1995). Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. *Nat Genet* 10, 41-46.
- Wang S, Detera-Wadleigh SD, Coon H, Sun CE, Goldin LR, Duffy DL, Byerley WF, Gershon ES, Diehl SR. (1996). Evidence of linkage disequilibrium between schizophrenia and the SCA1 CAG repeat on chromosome 6p23. *Am J Hum Genet* 59, 731-736.

- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077-1082.
- Weber JL, Wong C. (1993). Mutation of human short tandem repeats. *Hum Mol Genet* 2, 1123-1238.
- Wei J, Hemmings GP. (2000). The NOTCH4 locus is associated with susceptibility to schizophrenia. *Nat Genet* 25, 376-377.
- Weinberger DR. (1987). Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry* 44, 660-669.
- Weinberger DR. (1995). Schizophrenia as a neurodevelopment disorder. In Hirsch SR I Weinberger DR eds: Schizophrenia. Oxford: Blackwell Science Ltd, 293-323.
- Weiner HL, Rothman M, Miller DC, Ziff EB. (1996). Pediatric brain tumors express multiple receptor tyrosine kinases including novel cell adhesion kinases. *Pediatr Neurosurg* 25, 64-71.
- Weiner HL, Huang H, Zagzag D, Boyce H, Lichtenbaum R, Ziff EB. (2000). Consistent and selective expression of the discoidin domain receptor-1 tyrosine kinase in human brain tumors. *Neurosurgery* 47, 1400-1409.
- Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. (1998). A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Mol Genet Metab* 64, 169-172.
- Williams NM, Rees MI, Holmans P, Norton N, Cardno AG, Jones LA, Murphy KC, Sanders RD, McCarthy G, Gray MY, *et al.* (1999). A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. *Hum Mol Genet* 8, 1729-1739.
- Williams NM, Preece A, Morris DW, Spurlock G, Bray NJ, Stephens M, Norton N, Williams H, Clement M, Dwyer S, *et al.* (2004). Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry* 61, 336-344.
- Wolford JK, Blunt D, Ballecer C, Prochazka M. (2000). High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC). *Hum Genet* 107, 483-487.
- Wright IC, Rabe-Hesketh S, Woodruff PW, David AS, Murray RM, Bullmore ET. (2000). Meta-analysis of regional brain volumes in schizophrenia. *Am J Psychiatry* 157, 16-25.
- Wu DY, Ugozzoli L, Pal BK, Wallace RB. (1989). Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc Natl Acad Sci USA* 86, 2757-2760.
- Wu DY, Wallace RB. (1989). The ligation amplification reaction (LAR)--amplification of specific DNA sequences using sequential rounds of template-dependent ligation. *Genomics* 4, 560-569.
- Yen CJ, Beamer BA, Negri C, Silver K, Brown KA, Yarnall DP, Burns DK, Roth J, Shuldiner AR. (1997). Molecular scanning of the human peroxisome proliferator activated receptor gamma (hPPAR gamma) gene in diabetic Caucasians:

identification of a Pro12Ala PPAR gamma 2 missense mutation. *Biochem Biophys Res Commun* 241, 270-274.

Zerlin M, Julius MA, Goldfarb M. (1993). NEP: a novel receptor-like tyrosine kinase expressed in proliferating neuroepithelia. *Oncogene* 8, 2731-2739.

## **Referències i informació de base de dades electròniques**

### *Breast Cancer Information Core (BIC)*

Es pot trobar a la direcció URL: <http://research.nhgri.nih.gov/bic/>  
Emprada per la identificació de SNPs en el gen BRCA1.

### *dbSNP*

Es pot trobar a la direcció URL: <http://www.ncbi.nlm.nih.gov/SNP/index.html>  
Emprada per la identificació de SNPs en els gens DDR1 i GTFH4.

### *Online Mendelian Inheritance in Man (OMIM)*

Es pot trobar a la direcció URL:  
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

### *Vertebrate Genome Annotation (VEGA) database*

Es pot trobar a la direcció URL: <http://vega.sanger.ac.uk/>  
Emprada per la anotació de SNPs en el gen DDR1.

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## **INTRODUCCIÓ**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## **HIPÒTESI i OBJECTIUS**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007



## **PARTICIPANTS, MATERIALS i MÈTODES**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## **ESTUDI 1**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## **ESTUDI 2**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## **ESTUDI 3**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007



## **DISCUSSIÓ GENERAL, CONCLUSIONS i PERSPECTIVES FUTURES**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## **BIBLIOGRAFIA**

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

**Hospital Psiquiàtric Universitari  
INSTITUT PERE MATA**  
Departament de Formació i Investigació

**UNIVERSITAT ROVIRA I VIRGILI  
Facultat de Medicina i Ciències de la Salut**

Departament de Medicina i Cirurgia  
Unitat de Psiquiatria i Psicologia Mèdica

**Elisabet Vilella Cuadrada**, doctora en Ciències biològiques, investigadora del Departament de Formació i Investigació de l'Hospital Psiquiàtric Universitari Institut Pere Mata i dels Laboratoris Clínics de l'Hospital Universitari de Sant Joan de Reus i professora associada de la Unitat de Psiquiatria i Psicologia Mèdica de la Facultat de Medicina i Ciències de la Salut:

INFORMA QUE:

La Tesi Doctoral titulada "**Anàlisi mutacional i estudi d'associació del gen receptor domini discodina 1 (DDR1) en l'esquizofrènia**", presentada per Carmen Virgos Matilla, ha estat realitzada sota la meva direcció, i compleix amb els requeriments necessaris per a optar al títol de Doctora en Biologia.

Per a què així consti i tingui els efectes oportuns, signo la present.

Nom i signatura del Director

Nom i Vist i plau del Tutor

Dra. Elisabet Vilella Cuadrada

Dr. Manel Santafé Martínez

Reus, 28 de maig de 2004

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

## Agraïments

Són moltes les persones que han col.laborat directa o indirectament en la realització d'aquesta tesi. Aquest és el fruit del treball i la col.laboració amb els meus companys del CRB, d'ACLARA BioSciences Inc. i de tot el que hem après durant aquests últims anys.

En particular, vull donar les gràcies a la Dra. Elisabet Vilella, per compartir amb mi molt més del necessari i perquè d'ella he après ciència i vida. Gràcies.

Al Dr. Jorge Joven, per la seva empenta i suport.

Al Dr. Antonio Labad, Dr. Joaquín Valero i tots aquells que des de l'HPUIPM han fet possible la recollida de mostres i dades dels malalts participants.

A la Dra. Lourdes Martorell, per oferir-me sempre el seu ajut i estar sempre disponible a 'l'altre costat de la pantalla'.

Als companys del CRB: a les fantàstiques 'nenes' Lúdia, Vicky, Natàlia, Neus, Bàrbara, Carme i Iolanda, per tants cops de mà i per tantes hores compartides d'amistat. Al Jordi i al Josep Maria, que no perdeu mai el bon humor.

A l'Albert Almejide, Dr. Francesc Calafell i Dr. Javier Costas, pel seu ajut en l'anàlisi estadística.

A tots els malalts i participants que han col.laborat desinteressadament en aquest projecte.

This dissertation owes many thanks to a number of people for their assistance, patience and kindness. Without their support, I could not have done what I was able to do. My most sincere gratitude goes to:

Dr. Sharat Singh, for being a judicious mentor.

Dr. Maureen Cronin, a true believer in me, for your tremendous support and enthusiasm for my genotyping research.

Dr. Tina Tian, for allowing me to finish the genotyping project and step into the 'gene expression' world... thank you!

The 'organic chemists' at ACLARA. Without you, guys, this project would not have been possible.

Dr. Elaine Mansfield, for her advice and direction when times were rough.

Dr. Alex Sassi and Dr. Meilan Liu. For deep discussions about oligo separations, lab-cards and multiplex PCR. Why didn't my multiplex design work?

Als meus pares i germans, l'ajut i suport constant que sempre m'han donat.

Al Pere, per ser amb qui em recolzo en tot moment, per encoratjar-me a encarar la vida pel costat positiu.

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007



---

<b>ÍNDEX TEMÀTIC</b>	<i>i</i>
<b>ABREVIATURES</b>	<i>iii</i>
<b>ANTECEDENTS</b>	<i>v</i>
<b>1. INTRODUCCIÓ</b>	
1.1. L'esquizofrènia	3
1.1.1. Antecedents històrics	3
1.1.2. Definició	3
1.1.3. Epidemiologia	4
1.1.4. Aspectes neurobiològics	5
1.1.5. Teories causals	6
1.1.6. Estudis genètics	8
1.1.7. Cromosoma 6 i esquizofrènia	14
1.2. El receptor domini discoidina 1 (DDR1)	16
1.2.1. Per què estudiar <i>DDR1</i> com a gen candidat per l'esquizofrènia?	16
1.2.2. Família dels receptors domini discoidina	16
1.2.3. Estructura gènica i proteica de DDR1	19
1.2.4. Implicació de DDR1 en el sistema nerviós central	25
1.2.5. Implicació de DDR1 en processos d'adhesió, migració, proliferació cel.lular	25
1.3 Mètodes de genotipatge de polimorfismes d'un únic nucleòtid (SNPs)	27
1.3.1. Polimorfismes d'un únic nucleòtid	27
1.3.2. Tècniques de genotipatge d'SNPs: requeriments	29
1.3.3. Estratègies de discriminació al·lèlica	29
1.3.4. <i>eTag<sup>TM</sup> Multiplex Invader<sup>®</sup> SNP Assay</i> : un nou protocol per l'anàlisi simultània d'SNPs	34
<b>2. HIPÒTESI I OBJECTIUS</b>	41
<b>3. PARTICIPANTS, MATERIALS I MÈTODES</b>	
3.1. Disseny experimental	45
3.2. Descripció de la mostra	46
3.3. Aïllament i quantificació de DNA genòmic	48
3.4. Seqüenciació automàtica de les regions codificant i exó-intró del gen <i>DDR1</i>	51
3.5. Clonatge de fragments de PCR en un vector plasmidi	53

3.6. Marcadors i genotipatge mitjançant el mètode PCR-RFLP ( <i>polymerase chain reaction–restriction fragment length polymorphism</i> )	54
3.7. Marcadors i genotipatge mitjançant el mètode <i>eTag Multiplex Invader</i> per l'anàlisi d'SNPs	57
3.8. Anàlisi de les dades	60
<b>4. RESULTATS</b>	
<b>Estudi 1:</b> Rastreig de mutacions en el gen <i>DDR1</i> en una mostra de malalts d'esquizofrènia: seqüenciació en <i>pools</i> de DNA	65
Introducció	65
Resultats	66
Discussió	73
<b>Estudi 2:</b> Anàlisi de l'associació del gen <i>DDR1</i> amb l'esquizofrènia: estudi de casos i controls	81
Introducció	81
Resultats	84
Discussió	90
<b>Estudi 3:</b> Validació del mètode <i>eTag Multiplex Invader</i> per l'anàlisi d'SNPs	95
Introducció	95
Resultats	98
Discussió	111
<b>5. DISCUSSIÓ GENERAL, CONCLUSIONS I PERSPECTIVES</b>	117
<b>6. BIBLIOGRAFIA</b>	125

---

A	Adenina
aa	Aminoàcid
ARMS	<i>Amplification refractory mutation system</i>
ARPE	Alçada relativa del pic electroforètic
C	Citosina
DNA	Àcid desoxiribonucleic
dNTP	Deoxiribonucleòsid trifosfat
DDR1	Receptor domini discoidina 1
DDR2	Receptor domini discoidina 2
DGGE	<i>Denaturing gradient gel electrophoresis</i>
dHPLC	<i>Denaturing high-performance liquid chromatography</i>
EC	Electroforesi capil.lar
EDTA	Àcid etilen diaminotetraacètic
EST	<i>Expressed sequence tag</i>
G	Guanina
GTF2H4	<i>General transcription factor IIIH, polypeptide 4</i>
h	Hora
HDA	<i>Heteroduplex analysis</i>
Kda	Quilodaltons
Kv	Quilovolts
LIF	<i>Laser-induced fluorescence</i>
LIF-CE	<i>Laser-induced fluorescence - capillary electrophoresis</i>
LOD	Logaritme en base 10 de la probabilitat de lligament ( <i>Lod score</i> )
M, mM, $\mu$ M	Molar, mil.limolar, micromolar
Mb, Kb, pb	Megabase, kilobase, parell de base
MEC	Matriu extracel.lular
min	Minut
mL, $\mu$ L	Mil.lilitre, microlitre
MMP	Metal.loproteasa
mRNA	RNA missatger
MT	Mutat
ng	Nanogram
nm	Nanòmetre
NPL	<i>Lod score no paramètric</i>
OD	<i>Optical density</i>
PCR	<i>Polymerase chain reaction</i>
RFLP	<i>Restriction fragment length polymorphism</i>
RNA	Àcid ribonucleic
RPTK	<i>Receptor protein-tyrosine kinase</i>
seg	Segon
SNC	Sistema nerviós central
SNP	Polimorfisme d'un únic nucleòtid
SSCP	<i>Single strand conformation polymorphism</i>
STRP	<i>Short tandem repeat polymorphism</i>
T	Timina
Tm	Temperatura de fusió
UTR	<i>Untranslated region</i>
WT	<i>Wild type</i>

UNIVERSITAT ROVIRA I VIRGILI  
ANÀLISI MUTACIONAL I ESTUDI D'ASSOCIACIÓ DEL GEN RECEPTOR DOMINI DISCOIDINA 1 (DDR1) EN L'ESQUIZOFRÈNIA.  
Carmen Virgos Matilla  
ISBN: 978-84-690-7776-4 / DL: T.1209-2007

En el moment que ens vam plantejar aquest treball (1998), el col.lagen tot just s'havia identificat com a lligand fisiològic dels receptors domini discoidina (DDRs): receptor domini discoidina 1 (DDR1) i receptor domini discoidina 2 (DDR2). Malgrat la funció dels DDrs era desconeguda, s'havia proposat la seva implicació en interaccions cèl.lula-cèl.lula i en vies de transducció del senyal d'adhesió i migració cel.lular, donada la presència de la regió extracel.lular homòloga a la proteïna discoidina 1 de *Dictyostelium discoideum*.

En humans, el gen *DDR1* es localitza al cromosoma 6p21, regió candidata a contenir un o més gens de susceptibilitat per l'esquizofrènia. Malgrat la funció del DDR1 era i és encara desconeguda, l'estructura gènica i el seu patró d'expressió han estat parcialment caracteritzats en models animal i humà. En múrids, *Ddr1* (el gen homòleg a *DDR1*) presenta un patró d'expressió característica en sistema nerviós central (SNC). Durant els estadis embrionaris, *Ddr1* es detecta en zones proliferatives i, durant els estadis adults, la seva expressió es localitza en subpoblacions glials, substància blanca i zones de neurogènesi secundària. D'altra banda, hi havia precedents que apuntaven al col.lagen com a modulador de la proliferació de les cèl.lules neuroepitelials en rata. Era lògic pensar, doncs, que DDR1 podria tenir un paper rellevant durant el desenvolupament del SNC.

S'han formulat diverses hipòtesis sobre què desencadena l'aparició de l'esquizofrènia. Durant l'última dècada, la hipòtesi del neurodesenvolupament ha estat la més àmpliament recolzada. Aquesta sosté que la malaltia s'origina per alteracions durant la neurogènesi i gliogènesi en les etapes de desenvolupament perinatal. I és al final de l'adolescència quan apareixen els símptomes clínics de la malaltia. Aquesta hipòtesi encaixa bé el supòsit que l'esquizofrènia és una malaltia complexa, resultat de la combinació de factors genètics i factors ambientals. Dins del marc d'aquesta hipòtesi, ens vam plantejar estudiar el *DDR1* com a gen candidat en l'esquizofrènia. Era necessari, doncs, caracteritzar en detall la seva expressió durant el neurodesenvolupament i en el cervell adult. Així com identificar les variants genètiques en el *DDR1* humà.

Ja iniciat aquest projecte, per motius personals, la doctoranda va desplaçar-se als Estats Units a principis del 2000 i va unir-se al grup de la Dra. Tina Tian a l'empresa ACLARA BioSciences Inc. (Mountain View, CA, EUA), dedicant-se al desenvolupament de nous mètodes de genotipatge. En aquell moment es va establir una col.laboració entre el grup de la Dra. Elisabet Vilella i ACLARA BioSciences Inc. que havia de proporcionar les eines i possibilitats metodològiques per a desenvolupar aquest projecte de tesi.

L'estudi de l'expressió de DDR1 durant el neurodesenvolupament i en cervell adult (en ratolins i humans) és treball de tesi de dos altres membres del grup de la Dra. Elisabet Vilella. La present tesi doctoral se centra en l'estudi de variants genètiques

en el gen *DDR1* en una mostra de malalts d'esquizofrènia i individus controls i en el desenvolupament d'un nou mètode d'anàlisi de polimorfismes.

## 1.1. L'ESQUIZOFRÈNIA

### 1.1.1. ANTECEDENTS HISTÒRICS

La paraula esquizofrènia deriva dels termes grecs *skhizo* i *phren* que volen dir escissió i ment, respectivament. Tot i que el terme esquizofrènia va estar encunyat pel psiquiatre suís Eugen Bleuler l'any 1911, la condició a la que es refereix ha acompanyat a la humanitat des de fa molt de temps. Les descripcions més antigues d'aquest trastorn es troben en el papyrus d'Ebers (1550 aC). Les primeres teories sobre la base física de les malalties mentals, la integració del cos i la ment i els processos mentals no conscients van ser formulades per Plató (427-347 aC). Contemporàniament, Hipòcrates (470-? aC) proposà que els trastorns mentals tenen el seu origen en el cervell. Però no és fins a començaments del segle XVIII que es fan les primeres descripcions clíniques detallades de les malalties mentals.

La primera aproximació clínica a aquesta síndrome fou descrita pel psiquiatre alemany Emil Kraepelin (1856-1926) amb el nom de demència precoç. L'apreciació que en els malalts inclosos sota el diagnòstic de Kraepelin, predominaven més les característiques de dissociació i disgregació de la personalitat que les d'autèntica demència, va conduir a Bleuler a designar aquest grup de malalts com afectats d'esquizofrènia.

El criteri de diagnòstic i la lògica d'un diagnòstic diferencial pels trastorns mentals no apareix fins l'any 1948 amb la sisena edició de la Classificació Internacional de Malalties (International Classification of Disease, ICD-6) establert per l'Organització Mundial de la Salut. Paral·lelament, a l'any 1952, l'Associació Americana de Psiquiatria publicà la primera edició del Manual Diagnòstic i Estadístic de Trastorns Mentals (*Diagnostic and Statistical Manual of Mental Disorders*, DSM). Des d'aleshores, ambdós manuals han modificat i incorporat diagnòstics d'acord amb els avenços realitzats en psiquiatria. Aquests dos sistemes són els d'ús comú a tot el món. Les versions més recents ICD-10 (1992) i DSM-IV (1994) tenen definicions similars d'esquizofrènia i el diagnòstic que es realitza amb un i altre coincideix amb un elevat grau d'exactitud.

### 1.1.2. DEFINICIÓ

L'esquizofrènia es caracteritza per una alteració profunda del pensament i la percepció, afectant els atributs humans més fonamentals: llenguatge, pensament, percepció, afectivitat i sentit d'un mateix. La diversitat de símptomes freqüentment inclouen manifestacions psicòtiques com són sentir veus internes o experimentar sensacions no vinculades a cap font (al·lucinacions auditives i visuals) i tenir creences falses i persistents, com que el malalt és perseguit o es creu ser el centre de tot el passa al seu voltant (idees delirants).

No hi ha cap símptoma definitiu pel diagnòstic. Aquest es basa en la presentació d'una sèrie de signes i símptomes peculiars que han estat presents una part

significativa de temps durant el període d'un mes (o durant un període més breu si hi ha hagut tractament amb èxit) i amb alguns signes del trastorn que han persistit com a mínim 6 mesos. Aquests símptomes són: idees delirants, al·lucinacions, llenguatge desorganitzat, comportament greument desorganitzat o catatònic i símptomes negatius (pèrdua de les habilitats socials, falta d'interès, incapacitat d'experimentar plaer, incapacitat d'expressar emocions, pensament i llenguatge empobrits).

Es tracta d'un trastorn que sol ser crònic, recurrent i greu que té un gran impacte en cadascun dels aspectes de la vida de la persona que el pateix, així com també en l'entorn familiar i social. Les persones afectades pel trastorn se senten com si haguessin perdut la identitat, l'autonomia i la capacitat mental. El suïcidi és una causa de mort freqüent en aquests pacients: un 50% intenten suïcidar-se almenys una vegada a la vida i d'un 10 a un 15% moren per aquesta causa.

El tractament, per la majoria de malalts, combina l'administració de fàrmacs antipsicòtics amb la psicoteràpia. No existeix cap tractament curatiu per aquest trastorn. La tendència actual és la d'evitar l'hospitalització de llarga durada, per no allunyar el malalt del seu entorn social. L'hospitalització és indicada si l'agitació psicomotriu és incontrolable, hi ha perill d'autoagressió o heteroagressió i per establir la medicació.

### **1.1.3. EPIDEMIOLOGIA**

L'esquizofrènia té una prevalença al voltant de l'1% en tots els àmbits geogràfics, culturals i socioeconòmics (Gottesman i Shields, 1992). Les primeres manifestacions d'aquesta malaltia apareixen entre el final de l'adolescència i el principi de l'edat adulta. En homes s'ha establert que l'edat d'inici oscil·la entre els 18 i els 25 anys, i en dones entre els 23 i 35 (Lewine, 1980).

Ja a l'any 1916, Ernst Rüdin va observar que l'esquizofrènia s'agregava en famílies (Rüdin, 1916). Els estudis de risc familiar duts a terme en els darrers 20 anys amb pacients amb esquizofrènia observen de forma consistent l'existència d'agregació familiar. Aquests estudis indiquen que els familiars de primer grau dels pacients tenen un risc d'afectació de 5 a 17 vegades major que els individus no relacionats (Kendler *et al*, 1985; Gershon *et al*, 1988; Gottesman i Shields, 1992). Aquesta agregació familiar és una indicació important d'una base genètica però no suficient, ja que altres components no genètics (ambientals, culturals) poden estar implicats en aquesta observació.

Els estudis de bessons han detectat una concordança del 5-15% i 40-50% en bessons dizigòtics i monozigòtics, respectivament (Farmer *et al*, 1987; Onstad *et al*, 1991; Gottesman i Shields, 1992). D'aquests estudis se'n desprèn que l'heretabilitat del trastorn és del 60-89% (Cannon *et al*, 1998; Cardno *et al*, 1999; Schultz i



Andreasen, 1999), recolzant la idea que hi ha factors genètics que contribueixen a la transmissió familiar del trastorn.

Els estudis de la descendència de bessons discordants per l'esquizofrènia indiquen que el risc d'afectació és el mateix pels fills de bessons monzigòtics discordants (17%). Mentre que en bessons dizigòtics, el risc és major pels fills del bessó malalt (17% versus 2%). Aquests resultats suggereixen que genotips que confereixen susceptibilitat a desenvolupar esquizofrènia poden no expressar-se (penetrança incompleta) i que la freqüència d'individus que no hereten la predisposició genètica però que desenvolupen la malaltia (fenocòpies) és baixa (Gottesman i Bertelsen, 1989). Aquestes dues característiques són importants a tenir en compte en fer les anàlisis dels estudis genètics ja que poden emascarar els resultats.

Els estudis de risc familiar, de bessons i d'adopció no són capaços de determinar l'existència d'un patró Mendelià de transmissió de la malaltia. El patró d'herència és complex i no està clar si l'agregació familiar en l'esquizofrènia és deguda a la contribució d'un gen major, varis gens o molts gens d'efecte menor. El model d'herència de la malaltia és encara desconegut.

L'esquizofrènia és sovint comòrbida amb altres condicions trastorn de pànic (10-20%) i trastorn obsessiu compulsiu (10-20%) (Craig *et al*, 2002). Un 50% dels pacients presenten algun trastorn d'abús de substàncies. I un 80-90% presenten dependència de nicotina.

#### **1.1.4. ASPECTES NEUROBIOLÒGICS**

El processament de la informació i, concretament, la tasca de correlacionar i integrar la informació emmagatzemada amb noves situacions en el context apropiat, sembla estar alterat en els malalts d'esquizofrènia. Les evidències d'un dèficit cognitiu en aquests malalts són múltiples (Elvevag i Goldberg, 2000), però el sistema o circuits neuronals subjacents són encara desconeguts. Els estudis de neuroimatge, neurofisiologia i anatomo-patològics *post-mortem* indiquen que l'esquizofrènia és una malaltia estructural i funcional del cervell que afecta varies regions corticals i subcorticals implicades en els aspectes cognitius, emocionals i de motivació de la conducta humana.

A nivell macroanatòmic, s'ha arribat al consens que els malalts d'esquizofrènia presenten una ampliació dels ventricles i una reducció del volum en àrees del sistema límbic (hipocamp i amígdala) i regions corticals (còrtex prefrontal) (Harrison, 1999 i 2004; McCarley *et al*, 1999; Pearlson i Marsh, 1999; Wright *et al*, 2000; Ho *et al*, 2003). Aquestes alteracions s'observen tant en malalts de llarga evolució com en primers episodis de malalts no medicats (Bogerts *et al*, 1991; DeLisi *et al*, 1991; Nopoulos *et al*, 1995). I, en un menor grau, també s'observen en familiars no afectats pel trastorn (Staal *et al*, 2000; Seidman *et al*, 2002). Alguns autors

consideren aquestes alteracions estructurals com a potencials 'marcadors biològics' de la malaltia.

Els estudis a nivell microanatòmic han detectat anomalies en la citoarquitectura neuronal, disminucions dels cossos neuronals, de la densitat i longitud de les espines dendrítiques i de proteïnes sinàptiques que serien, en part, responsables de les reduccions de les regions límbiques i corticals mencionades anteriorment (Harrison, 1999 i 2004; Honer *et al*, 1999; Glantz i Lewis, 2000; Chana *et al*, 2003). Estudis més recents convergeixen en la implicació de la glia (Cotter *et al*, 2001) i, en particular, de l'oligodendroglia i el procés de mielinització en l'esquizofrènia (Uranova *et al*, 2001 i 2004; Bartzokis, 2002; Hof *et al*, 2002 i 2003). Aquests resultats es veuen recolzats pels estudis d'expressió gènica, els quals es comenten més endavant, on es detecta una baixa expressió (o *downregulation*) de gens expressats per l'oligodendroglia i implicats en el procés de mielinització (Hakak *et al* 2001, Tkachev *et al*, 2003). En aquesta mateixa línia, s'han detectat nivells elevats de la proteïna S100B en sang (marcador de la integritat de les cèl.lules glials i dany cerebral) en malalts d'esquizofrènia no medicats (Lara *et al*, 2001; Rothermund *et al*, 2001). A més, les concentracions de S100B es mantenen elevades després del tractament i s'associen a una simptomatologia negativa (Rothermund *et al*, 2004).

Hi ha molta controvèrsia sobre si aquestes anomalies cerebrals fluctuen o són estàtiques amb l'aparició de la malaltia. S'accepta que el seu origen té lloc durant les etapes primerenques del neurodesenvolupament. Tot i que també es disposen de dades que apunten a l'existència d'un procés subtil de neurodegeneració amb la progressió de la malaltia. Les hipòtesis etiopatogèniques formulades més recentment per l'esquizofrènia integren ambdós processos, de neurodesenvolupament i neurodegeneració.

### **1.1.5. TEORIES CAUSALS**

El procés patofisiològic i els factors etiològics causants de l'esquizofrènia són, doncs, desconeguts. L'heterogeneïtat que presenta l'esquizofrènia, manifestada a través del ventall de símptomes, resposta al tractament i evolució, complica la formulació de models teòrics per la mateixa (Andreasen i Carpenter, 1993). S'ha postulat que aquesta heterogeneïtat seria resultat de múltiples etiologies responsables de diferents mecanismes patofisiològics (Tsuang *et al*, 1990) o bé, d'un únic procés etiopatogènic (Crow, 1990), que donat el context individual donaria lloc a les diferents manifestacions del trastorn (Carpenter i Kirkpatrick, 1988). A més, com ja s'ha comentat prèviament, els factors ambientals contribueixen a la variabilitat de la manifestació i evolució de la malaltia.

Les hipòtesis etiopatogèniques per l'esquizofrènia van des de la concepció més organicista fins les psicològiques i socials. Les teories psicosocials postulen que són determinats successos quotidians estressants els que desborden la capacitat

defensiva de l'individu en el seu rol social i desencadenen la malaltia; un canvi ambiental, l'estrès o una tensió ambiental excessiva, o una capacitat defensiva i suport social escàs poden donar lloc a una crisi i exacerbació. Aquestes teories, que foren influents a la meitat del segle XX, han estat difícils d'avaluar científicament. Les teories actuals consideren que l'esquizofrènia és el resultat de la combinació d'uns factors genètics i uns factors ambientals. El gen o gens implicats en la malaltia conferirien una predisposició que acabaria en manifestació clínica després de l'exposició a determinats factors ambientals. Des d'aquest punt de vista, l'esquizofrènia s'inclou dins del grup de les malalties complexes o multifactorials (Lander i Schork, 1994).

Algunes de les hipòtesis més populars per l'esquizofrènia són les que impliquen un canvi en la neurotransmissió de dopamina, glutamat, acetilcolina, serotonina, GABA i/o noradrenalina (Owen i Simpson, 1994), la hipòtesi del neurodesenvolupament (Weinberger, 1986), la hipòtesi de la transmetilació (Smythies, 1983), la hipòtesi viral (Torrey i Peterson, 1976) i la hipòtesi dels fosfolípids de membrana (Pettegrew *et al*, 1991). Totes elles es recolzen de dades epidemiològiques i derivades d'estudis de resposta a tractament, farmacològics, de neuroimatge i de determinacions d'hormones i neurotransmissors. Les hipòtesis dopaminèrgica i del neurodesenvolupament són potser les més acceptades. Però ni aquestes dues ni la resta són incompatibles entre elles.

Les hipòtesis més recentment formulades impliquen una alteració en la glia i proporcionen una explicació heurística a les observacions neurobiològiques i genètiques realitzades en l'esquizofrènia. A continuació se'n fa una breu descripció.

#### *Hipòtesi de la deficiència de factors de creixement glials i desestabilització sinàptica*

Aquesta postula que una deficiència funcional dels factors de creixement de la glia i de factors de creixement sintetitzats per la glia comportaria una desestabilització de la sinapsi, que seria la causa principal de l'aparició de l'esquizofrènia (Moises 2001, Moises *et al*, 2002). La hipòtesi suggereix que les cèl.lules glials serien el punt de trobada de les interaccions gen-ambient implicades en l'esquizofrènia. La vulnerabilitat a desestabilitzar la sinapsi i, conseqüentment, desenvolupar l'esquizofrènia, es deuria per tenir uns nivells basals baixos de factors de creixement glials i d'estabilitat sinàptica determinats genèticament i epigenètica.

#### *Hipòtesi de la disfunció de l'oligodendroglia i mielinització*

Aquesta postula que una alteració en el procés de mielinització i/o el seu manteniment afectaria profundament la connexió neuronal que resultaria en l'aparició de l'esquizofrènia (Bartzokis, 2002; Hof *et al*, 2002). La mielinització és un procés dinàmic que s'estén fins l'edat adulta (Benes *et al*, 1994). En particular, la mielinització entre les estructures límbiques i el còrtex prefrontal dorsolateral sembla tenir lloc durant la segona i tercera dècada de vida, període de vulnerabilitat per

l'aparició de l'esquizofrènia (Benes, 1989; Paus *et al*, 1999). Totes aquelles interaccions gen-ambient que alteressin el funcionament dels oligodendròcits i el procés de mielinització durant el neurodesenvolupament i maduració del cervell desencadenarien el trastorn. La dispersa distribució de l'oligodendroglia explicaria, en part, les diferents troballes neurobiològiques i també el ventall de símptomes que presenten aquests malalats.

### 1.1.6. ESTUDIS GENÈTICS

Com se'n deriva de les dades epidemiològiques revisades anteriorment, el component genètic de l'esquizofrènia és considerable. Ara bé, no presenta un patró d'herència Mendeliana clàssica i, per això, és considerada una malaltia complexa. Les característiques que explicarien la dificultat en caracteritzar els gens en les malalties complexes són les següents:

- *Herència oligogènica o poligènica*. Presència de mutacions en varis o múltiples gens confereix susceptibilitat.
- *Penetrança incompleta*. Presència de la variant de susceptibilitat, però es requereix la interacció amb determinats factors ambientals (interaccions gen-ambient) o altres variants gèniques (interaccions gen-gen) per l'aparició dels símptomes.
- *Heterogeneïtat genètica*. Mutacions en qualsevol dels gens implicats poden resultar en el mateix fenotip.
- *Factors ambientals* comuns o compartits pels diferents membres familiars que faciliten l'aparició d'agregació familiar.
- *Fenocòpies*. Individus que no hereten la predisposició genètica desenvolupen la malaltia.
- *Elevada freqüència dels al·lels de susceptibilitat* de causar la malaltia en la població general.

Existeixen moltes evidències que qualsevol d'aquestes característiques pot estar operant en l'esquizofrènia. En són un clar exemple la manca de *LOD scores* positius en els estudis de lligament o la incapacitat de replicar-ne els resultats, junt amb la incapacitat de replicar els resultats positius de la gran majoria dels estudis d'associació.

S'han identificat varis mecanismes d'herència complexa o no mendeliana, i inclouen l'herència mitocondrial, l'empremta (*imprinting*), la disomia uniparental i l'anticipació. Fins el moment, no hi ha prou evidència de l'acció d'aquests mecanismes en l'esquizofrènia.

Els estudis d'investigació de *loci* de susceptibilitat per l'esquizofrènia es poden agrupar en sis categories (Taula I). Aquests no són mútuament excloents i la utilització d'un o altre mètode dependrà del tipus i la grandària de la mostra, el tret fenotípic estudiat i mecanisme o model genètic proposat. A continuació es fa una

breu revisió de l'estat actual de la investigació de gens de susceptibilitat per l'esquizofrènia.

**Taula I.** Resum de les aproximacions emprades per estudiar les regions cromosòmiques i/o gens implicats en l'esquizofrènia.

Mètode	Estudis d'anàlisi de lligament		Estudis d'associació		Estudis citogenètics i anomalies cromosòmiques	Estudis d'expressió gènica per microxips
	Model dependent (paramètric)	Model independent (no paramètric)	Basats en famílies	Basats en població		
<b>Nom alternatiu</b>	Anàlisi del <i>LOD score</i>	<i>Allele sharing method</i>	Test de desequilibri de la transmissió (TDT); Test de risc relatiu d'haplotip (HRR)	Estudi de casos i controls		Anàlisi per <i>microarray</i>
<b>Objectiu</b>	Localitzar locus implicat en la malaltia		Determinar si un al·lel o haplotip concret està associat o en desequilibri de lligament amb la malaltia		Localitzar alteracions cromosòmiques associades a la malaltia	Detectar si l'expressió determinada d'un gen s'associa a la malaltia
<b>Mostra</b>	Genealogies grans amb múltiples individus afectats	Parelles de germans afectats o pare o mare i fill afectats	Individu afectat i pares heterozigots pels marcadors estudiats	Individus afectats (casos) i individus no afectats (controls)	Casos i controls; genealogies	Casos i controls
<b>Utilització</b>	Anàlisi de tot el genoma (estudis globals del genoma) o d'un locus específic		Anàlisi d'un locus o gen candidat		Anàlisi d'una regió cromosòmica o gen candidat	Anàlisi de milers de gens simultàniament
<b>Avantatge</b>	Capaç de detectar l'efecte d'un locus en gran distàncies genètiques		Utilització dels genotips parentals com a control intern	Capaç de detectar gens d'efecte moderat o lleu	Capaç d'acotar la regió amb el gen o gens implicats	Capaç d'identificar nous gens candidats en absència d'hipòtesi preconcebuda
<b>Desavantatge</b>	Dificultat de trobar grans genealogies	Poc sensible al definir la localització del locus implicat	Dificultat de trobar ambdós pares	Estratificació de la mostra	Poc sensible	Falsos positius per manca de controls aparellats

### Estudis de lligament

Els estudis globals del genoma en famílies amb esquizofrènia han trobat valors suggestius de lligament amb mètodes paramètrics (depenen d'un model genètic d'herència) i no paramètrics (no depenen d'un model genètic d'herència) en més de la meitat dels 23 cromosomes (Taula II). Molts d'aquests resultats positius no han estat replicats quan s'estudien cohorts més grans o les mateixes regions són estudiades per grups independents. Les explicacions a aquesta disparitat de resultats inclouen un poder insuficient de les mostres per a detectar lligament, heterogeneïtat genètica, manca d'una definició acurada del fenotip o bé existència de múltiples al·lells de susceptibilitat amb un efecte moderat. Si la predisposició a la malaltia és resultat de la combinació de diferents gens d'efecte moderat, és d'esperar que els resultats positius dels estudis de lligament siguin costosos de replicar.

**Taula II.** Principals estudis de lligament genètic en l'esquizofrènia.

Estudi <sup>a</sup>	Regió cromosòmica	Mostra	Diagnòstic
Blouin <i>et al</i> , 1998 (452 marcadors)	8p21-22 13q32 14q13 22q11	54 famílies (origen divers)	DSM-III-R
Levinson <i>et al</i> , 1998 (310 marcadors)	2q 4q 9 10q 11q	43 famílies (origen divers)	DSM-III-R
Hovatta <i>et al</i> , 1999 (351 marcadors)	1q32 4q34 9q21 Xp11	20 famílies de Finlàndia	DSM-IV
Williams <i>et al</i> , 1999 (229 marcadors)	4p 18q Xcen	196 parelles de germans afectes	DSM-IV
Gurling <i>et al</i> , 2001 (365 marcadors)	1q33.2 4q13-31 8p22.1-22 11q21 11q23.3-24	13 famílies	RDC <sup>b</sup> DSM-III
Paunio <i>et al</i> , 2001 (315 marcadors)	5q 2q	238 famílies de Finlàndia	DSM-IV
DeLisi <i>et al</i> , 2002 (396 marcadors)	2cen 10p15-p13 22q12	382 parelles de germans afectes	
Devlin <i>et al</i> , 2002 (546 marcadors)	3q 5q 9p	5 famílies de Palau	DSM-IV
Stefansson <i>et al</i> , 2002 (950 marcadors)	8p12-21	33 famílies d'Islàndia	RDC
Straub <i>et al</i> , 2002 (166 marcadors)	2p14-13 4q24-32 5q21-31 6p24-21 8p22-21 10p15-p11	270 famílies d'Irlanda	DSM-III-R

<sup>a</sup> S'han inclòs aquells estudis amb valors de *LOD score* i *NPL* superiors a 1. Aquests són els estadístics que expressen la probabilitat de què un marcador cosegregui o estigui lligat amb el tret o malaltia d'estudi. Els valors que oscil·len entre 1 i 3 són considerats 'valors suggestius de lligament' i la seva rèplica en mostres independents és considerada una prova de confiança (Lander i Kruglyak, 1995). Per la confirmació de lligament d'una determinada regió cromosòmica o marcador es consideren valors superiors a 3. En el cas de les malalties complexes i en concret, l'esquizofrènia, la majoria dels estudis de lligament obtenen valors entre 1-3; <sup>b</sup>RDC, Criteris de Diagnòstic en Investigació (*Research Diagnostic Criteria*).

### Estudis d'associació

S'ha postulat que els estudis d'associació són més efectius per detectar gens de susceptibilitat d'efecte moderat o lleu respecte els estudis de lligament (Risch i Merikangas, 1996). Ara bé, per l'estudi global del genoma per associació es requereix d'analitzar milers de marcadors, la qual cosa no és viable actualment. I, d'altra banda, tampoc és clar quin és el nombre de marcadors necessaris. És per això que

els investigadors s'han centrat en l'estudi de gens o regions candidates. Un gen o una regió es defineix com a candidat si es troba en un segment cromosòmic que ha estat prèviament implicat mitjançant el mapatge posicional i/o la seva funció és rellevant dins del marc d'una de les hipòtesis etiopatogèniques de la malaltia. Donat el nombre de regions cromosòmiques implicat en l'esquizofrènia i, que aquestes estan poc acotades, el número de gens candidats és extremadament elevat. Com ja s'ha mencionat, són moltes les hipòtesis etiopatogèniques formulades per l'esquizofrènia i la tria de gens candidats és igualment difícil en aquesta situació. La Taula III recull una llista parcial amb gens candidats per l'esquizofrènia.

**Taula III.** Principals gens estudiats com a candidats a participar en la etiologia de l'esquizofrènia.

Hipòtesi de treball	Gen	Proteïna
Hipòtesis que impliquen canvi en la neurotransmissió	<i>DR</i>	Receptors dopamina (D1, D2, D3, D4)
	<i>HTR</i>	Receptors serotonina (HTR1A, 1B, 2C)
	<i>SLC6A4</i>	Serotonin transporter
	<i>GRIN2B</i>	Glutamate receptor
	<i>COMT</i>	Cathecol-O-methyltransferase
	<i>MAO</i>	Monoamina oxidases A i B
Hipòtesi del neurodesenvolupament	<i>CNTF</i>	Ciliary neurotrophic factor
	<i>BDNF</i>	Brain derived neurotrophic factor
	<i>NT-3</i>	Neurotrophin 3
	<i>TRKs</i>	Receptors de neurotrofines (TRKA, TRKB, TRKC)
	<i>RELN</i>	Reelin
Hipòtesi de la disfunció de l'oligodendroglia i mielinització	<i>GSN</i>	Gelsolin
	<i>MAG</i>	Myelin associated glycoprotein
	<i>TF</i>	Transferrin
	<i>MAL</i>	T-cell differentiation protein
Hipòtesi de la deficiència de factors de creixement glials i desestabilització sinàptica	<i>NRG</i>	Neuregulin (NGR1, NGR2, NGR3, NGR4)
	<i>ErbB</i>	Receptors de NRG (ErbB1, ErbB2, ErbB3, ErbB4)
	<i>GDNF</i>	Glial cell line derived neurotrophic factor
	<i>IGF1</i>	Insuline-growth factor I
	<i>EGF</i>	Epidermal growth factor
Altres	<i>RGS4</i>	G-protein signaling 4
	<i>NOTCH4</i>	Notch homolog 4 (Drosophila)
	<i>IL</i>	Interleuquina (IL2, IL12)
	<i>APO</i>	Apolipoproteïnes (Apo E, Apo D, Apo L)
	<i>MTHFR</i>	Methylen tetrahydrofolate reductase
	<i>SCA1</i>	Ataxia spinocerebelar I

La majoria dels estudis d'associació de gens candidats s'han centrat en gens implicats en la neurotransmissió dopaminèrgica i serotoninèrgica i en la hipòtesi del neurodesenvolupament. La majoria d'aquests estudis han analitzat un únic polimorfisme per locus o gen candidat. Les interpretacions que deriven d'un resultat positiu en aquest tipus d'estudi poden ser dues: 1. l'al·lel estudiat comporta un canvi biològic i és la variant causant o de susceptibilitat (associació directa); 2. l'al·lel estudiat es troba tan proper a l'al·lel causant que ambdós s'hereten conjuntament, és a dir, hi ha associació mantinguda per un desequilibri de lligament entre *loci* (associació indirecta).

A l'igual que pels estudis de lligament, els resultats positius dels estudis d'associació tampoc s'han pogut replicar de forma consistent. Aquestes discrepàncies rauen sovint en una selecció inadequada de la mostra de casos i controls. Això pot donar lloc a:

- i. Falsos positius (error de tipus I) deguts a la presència d'estratificació en la mostra o un nivell de significació inadequada.
- ii. Falsos negatius (error de tipus II) deguts a una mida insuficient de la mostra, presència d'heterogeneïtat genètica o diferències en la definició del fenotip.

Malgrat aquesta disparitat de resultats, estudis recents han suggerit els gens *dysbindin (DTNBP1)* i *neuregulin 1 (NRG1)* com a *loci* de susceptibilitat per l'esquizofrènia (Owen *et al*, 2004). Ambdós gens es localitzen en regions cromosòmiques les quals havien estat prèviament implicades mitjançant estudis de lligament. *DTNBP1* es localitza en la regió 6p22.3 i *NRG1* en la regió 8p22-p11. Malgrat la identificació de múltiples polimorfismes i haplotips específics de risc, els mecanismes patogènics són encara desconeguts.

### **Estudis de citogenètica i anomalies cromosòmiques**

S'han descrit un gran nombre d'anomalies cromosòmiques en l'esquizofrènia; però tres s'han associat a aquest trastorn: les microdeleccions cromosòmiques de 22q11.2 o síndrome de la deleció de 22q (22qDS), la inversió pericèntrica del cromosoma 9 i la translocació (1;11)(q41;q14.3). Tots tres resultats es veuen recolzats per estudis de lligament.

Només la síndrome 22qDS, coneguda com a síndrome velo-cardio-facial (VCF) o de Digeorge, presenta un fenotip específic associat. És un trastorn molt variable que presenta una prevalença d'esquizofrènia del 24% (Murphy *et al*, 1999; Murphy, 2002). Tenint en compte que la prevalença de la síndrome VCF és de 0.03%, s'ha estimat que un 0.5–2% dels malalts d'esquizofrènia tindrien una microdeleció a 22q11 i que representaria un subtipus genètic de la malaltia d'inici en l'edat adulta (Bassett i Chow, 1999).

### **Estudis d'expressió gènica**

Els estudis recents d'expressió gènica que utilitzen la tecnologia dels *microarrays* mesuren simultàniament els nivells d'expressió de milers de gens. Aquesta aproximació permet la identificació de gens amb una expressió diferencial entre teixits de malalts i individus controls. Aquests tipus d'estudi són una aproximació complementària als estudis d'anàlisi de lligament i associació, ja que analitzen milers de gens i poden suggerir nous mecanismes moleculars i gens candidats implicats en la malaltia sense la formulació preconcebuda d'hipòtesis etiopatogèniques.

Els estudis realitzats fins el moment s'han centrat principalment en l'estudi de l'expressió gènica diferencial del còrtex prefrontal dorsolateral de malalts



d'esquizofrènia (Mirnics *et al*, 2000; Hakak *et al*, 2001; Mimmack *et al*, 2002; Vawter *et al*, 2002). Donades les diferències metodològiques en aquests treballs (com per exemple, en els criteris de diagnòstic, preparació de la mostra i extracció del mRNA, tipus de plataforma o *microarray* i anàlisi de les dades), les conclusions que se'n deriven són diferents. Mirnics i col.laboradors detecten diferències en gens implicats en la funció secretora presinàptica, per la qual cosa proposen que l'esquizofrènia és una malaltia de la sinapsi. Hakak i col.laboradors observen una *downregulation* de gens implicats en el procés de mielinització i expressats de forma majoritària en oligodendròcits. Aquests resultats han estat recolzats per estudis més recents (Tkachev *et al*, 2003). Aquests autors suggereixen que part de les alteracions cerebrals observades en els malalts d'esquizofrènia són degudes a una alteració en la funció dels oligodendròcits i els processos de mielintzació. Mimmack i col.laboradors observen una sobre-expressió (*upregulation*) de tres membres de la família de les apoliproteïnes L (apo L) que localitzen en el cromosoma 22q12, regió propera a la de la síndrome VCF. De moment, es desconeix quina és la funció de les apo L en cervell.

### Estudis d'endofenotips

Els endofenotips són trets o característiques mesurables que reflecteixen un genotip subjacent i que estarien més relacionats amb el genotip que amb la categoria diagnòstica del trastorn (Gottesman i Gould, 2003). Aquests trets poden implicar mecanismes neurofisiològics, bioquímics, endocrinològics, neuroanatòmics, o neuropsicològics i han de complir amb els criteris d'heretabilitat, estabilitat, especificitat i replicabilitat. Els endofenotips s'identifiquen amb més freqüència en els malalts i també sovint en familiars no afectes, indicant l'existència de penetrança incompleta. Els endofenotips identificats i lligats genèticament a l'esquizofrènia són els següents:

- i. Dèficit en els moviments de persecució ocular (*smooth pursuit eye tracking*) mapat al cromosoma 6p, regió implicada prèviament en l'esquizofrènia (Arolt *et al*, 1999).
- ii. Inhibició del potencial evocat P50 lligat als cromosomes 15 i 22 (Freedman *et al*, 1999; Myles-Worsley *et al*, 1999).
- iii. Disminució de l'amplitud potencial evocat P300 lligat al cromosoma 1q42 (Blackwood *et al*, 2001).
- iv. Dèficit en la memòria de treball associada al genotip Val158Met del gen *COMT* (Goldberg *et al*, 2003).

### Models animals

L'esquizofrènia és una malaltia característica dels humans. Aquesta no s'observa o no es pot mesurar en animals. No obstant això, la utilització de models animals, en particular el múrid, pot ser molt informativa per identificar gens implicats en trets fenotípics considerats de rellevància per l'esquizofrènia. Com per exemple, l'estudi amb múrids de la inhibició prepuls (*prepulse inhibition*, PPI) de la resposta a

sobresalt (*startle response*, SR) (Joober *et al*, 2002; Palmer *et al*, 2003). La PPI és un comportament incondicionat que té lloc de forma natural en un ampli nombre d'espècies animals en les quals, la presentació d'un lleuger estímul sensorial (estímul prepols) unes mil·lèsimes de segons abans de la presentació d'un estímul més intens que indueix la SR, disminueix la magnitud d'aquesta resposta. S'hipotitza que la presentació del prepols activa un procés inhibitor que minva la SR. Malalts d'esquizofrènia així com rates tractades amb fàrmacs d'efecte psicòtic presenten un dèficit en la PPI, que és revertit per l'administració de fàrmacs antipsicòtics.

### 1.1.7. CROMOSOMA 6 I ESQUIZOFRÈNIA

Una de les regions cromosòmiques més estudiada en l'esquizofrènia és la regió 6p (locus SCZD3 [MIM 60051]). Els primers estudis de lligament amb resultats positius i replicats per grups independents són els de Wang *et al* (1995) i Straub *et al* (1995), involucrant la regió 6p24-p21. Evidències addicionals de lligament en aquesta regió deriven de diferents estudis (Moises *et al*, 1995; Schwab *et al*, 1995; Antonarakis *et al*, 1995; Bailer *et al*, 2000; Schwab *et al*, 2002; Straub *et al*, 2002b; Schwab *et al*, 2003). Tot i que d'altres no han pogut replicar els resultats (Pulver *et al*, 1995; Hovatta *et al*, 1998; Maziade *et al*, 2001; DeLisi *et al*, 2002).

Els marcadors amb valors suggestius de lligament a 6p per l'esquizofrènia es distribueixen per una regió àmplia de més de 25 Mb, entre els marcadors D6S309 a D6S282 (Taula IV). No obstant, sembla que aquests s'agrupen en 4 subregions: 6p24 a prop de D6S309-D6S296, 6p24 a prop de D6S940/D6S470, 6p23-22 a prop de D6260 i 6p21.3 on es troba el locus del sistema HLA (de l'anglès *Human Leukocyte Antigen*).

S'han dut a terme varis estudis d'associació amb gens candidats localitzats a 6p i algun d'ells amb resultats positius (Taula IV). Recentment, un estudi d'associació basat en famílies ha detectat desequilibri de lligament entre el gen *DTNBP1* i un diagnòstic estricte d'esquizofrènia (Straub *et al*, 2002). Aquests resultats han estat replicats en diverses poblacions mitjançant estudis basats en famílies (Schwab *et al*, 2003; Tang *et al*, 2003; van den Oord, *et al*, 2003; Kirov *et al*, 2004) i estudis de casos i controls (Van den Bogaert, *et al*, 2003; Williams *et al*, 2004). Es tracta de la primera replicació d'un estudi de desequilibri de lligament en esquizofrènia. El patró de desequilibri de lligament és consistent amb la presència de més d'un al·lel de susceptibilitat depenent de la població estudiada. De moment, no s'han identificat variants funcionals en les regions codificants de *DTNBP1*, però se sospita que les variants afectarien la seva expressió i processament proteic. En aquesta línia, s'han observat diferències subtils en l'expressió de *DTNBP1* en la regió de l'hipocamp en malalts d'esquizofrènia, però no s'han associat a cap dels polimorfismes identificats (Talbot *et al*, 2004).

**Taula IV.** Resultats més significatius dels estudis genètics (de lligament i associació) i gens candidats en la regió 6p24-21 per l'esquizofrènia.

Regió cromosòmica	Referència	Marcador genètic	Posició en el mapa <sup>a</sup>	Gen candidat <sup>b</sup>	Referència
6p24	Lindholm <i>et al</i> , 1999 Bailer <i>et al</i> , 2000	D6S309	14.07	<i>NRN1</i> <sup>c</sup>	Moises <i>et al</i> , 2002
	Antonarakis <i>et al</i> , 1995 Straub <i>et al</i> , 1995 Maziade <i>et al</i> , 1997 Hwu <i>et al</i> , 2000	D6S296	14.07	<i>TFAP2A</i> <sup>d</sup>	Kawanishi <i>et al</i> , 2000
	Maziade <i>et al</i> , 1997	D6S277	14.61		
	Straub <i>et al</i> , 2002a	D6S940	16.84		
	Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8, 1996	D6S470	18.22		
6p23	Lindholm <i>et al</i> , 1999	D6S1578	26.71	<i>SCA1</i> <sup>e</sup>	Wang <i>et al</i> , 1996
	Straub <i>et al</i> , 2002a	D6S259	27.78	<i>DTNBP1</i> <sup>f</sup>	Straub <i>et al</i> , 2002b Schwab <i>et al</i> , 2003 Williams <i>et al</i> , 2004
	Wang <i>et al</i> , 1995 Schwab <i>et al</i> , 2000	D6S260	29.93		
6p22	Schwab <i>et al</i> , 1995 Moises <i>et al</i> , 1995 Schwab <i>et al</i> , 2000	D6S274	32.6		
	Straub <i>et al</i> , 1995 Straub <i>et al</i> , 2002a	D6S285	34.23		
6p21	Moises <i>et al</i> , 1995	D6S291	49.5		
	Arolt <i>et al</i> , 1996	D6S271	66.37	<i>NOTCH4</i>	Wei i Hemmings, 2000
	Arolt <i>et al</i> , 1996	D6S282	66.37	<i>TNF</i> <sup>g</sup>	Boin <i>et al</i> , 2001 Meira-Lima <i>et al</i> , 2003

<sup>a</sup> Posició en el mapa dels marcadors en cM respecte pter (final del braç curt del cromosoma). Dades del mapa Mfd98 del Marshfield Medical Research Center (<http://research.marshfieldclinic.org/genetics/>); <sup>b</sup> Gens candidats de la regió 6p que han estat estudiats en l'esquizofrènia. A excepció de *NRN1* que ha estat recentement proposat (Moises *et al*, 2002); <sup>c</sup> *NRN1*, neuritin; <sup>d</sup> *TFAP2A*, transcription factor ap2-alpha; <sup>e</sup> *SCA1*, spinocerebellar ataxia 1; <sup>f</sup> *DTNBP1*, dystrobrevin-binding protein 1; <sup>g</sup> *TNF*, tumor necrosis factor alpha.

## 1.2. EL RECEPTOR DOMINI DISCOIDINA 1 (DDR1)

### 1.2.1. PER QUÈ ESTUDIAR *DDR1* COM A GEN CANDIDAT PER L'ESQUIZOFRÈNIA?

A finals de l'any 1997, en la recerca de nous gens candidats per l'esquizofrènia, diferents treballs ens van fer suggerir l'estudi del gen *DDR1* en el context de la hipòtesi del neurodesenvolupament per l'esquizofrènia. Aquestes troballes es poden resumir en:

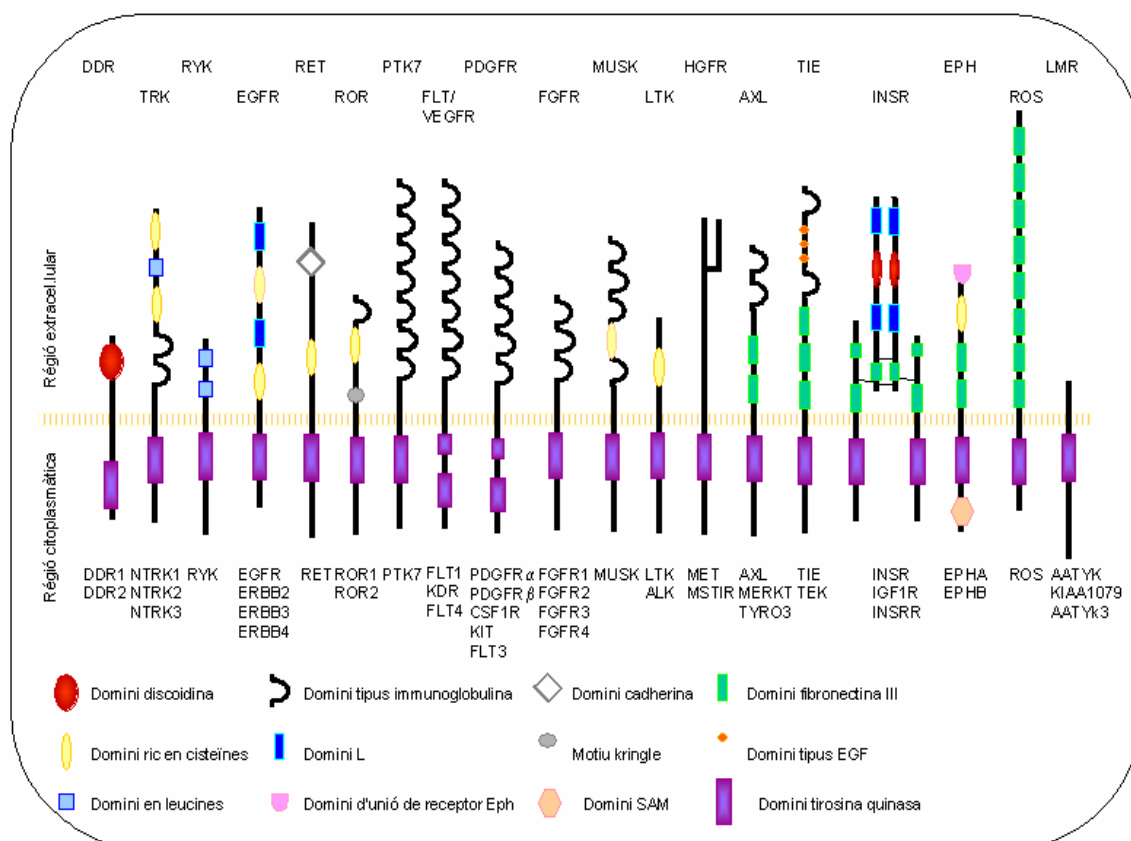
- i. Localització de *DDR1* a 6p24-21 (Valent *et al*, 1996), regió candidata a contenir un o més gens de susceptibilitat per l'esquizofrènia, com s'ha comentat prèviament.
- ii. Caracterització de *DDR1* com a receptor tirosina quinasa i amb homologia elevada amb el domini tirosina quinasa dels receptors de neurotrofines (Di Marco *et al*, 1993).
- iii. Expressió de *DDR1* en sistema nerviós central (SNC) i sobretot durant el neurodesenvolupament en ratolí i rata (Zerlin *et al*, 1993; Sanchez *et al*, 1994).
- iv. Modulació de la proliferació i diferenciació de cèl·lules neuroepiteliales i astroglija per col·lagen, lligand fisiològic de *DDR1* (Goetschy *et al*, 1987; Shrivastava *et al*, 1997; Vogel *et al*, 1997; Ali *et al*, 1998).

### 1.2.2. FAMÍLIA DELS RECEPTORS DOMINI DISCOIDINA

Els receptors domini discoidina 1 i 2 (*DDR1* i *DDR2*) constitueixen una família de receptors tirosina quinasa (RPTK, de l'anglès *receptor protein-tyrosine kinase*) per les seves homologies estructurals i funcionals. Els RPTK són proteïnes transmembrana implicades en el control i regulació dels processos cel·lulars de proliferació, diferenciació, transformació, senescència i apoptosi durant el desenvolupament i en la vida adulta. Les aproximadament 20 famílies de RPTKs identificades es caracteritzen per tenir un domini extracel·lular d'unió al lligand, un únic domini transmembrana i un domini intracel·lular tirosina quinasa (Figura 1). La transmissió del senyal dels RPTKs requereix de l'oligomerització del receptor, com a resultat de la unió del lligand, la qual promou l'autofosforilació de residus tirosina de la regió citoplasmàtica del receptor (Heldin, 1995). La fosforilació específica dels residus tirosina estableix la conformació activa del receptor i permet l'ancoratge de molècules de transmissió del senyal com proteïnes amb dominis SH2 i PTB, les quals reclutaran proteïnes addicionals activant les cascades corresponents de transmissió del senyal cap a l'interior cel·lular (Blume-Jensen i Hunter, 2001).

Els DDRs es distingeixen de la resta de RPTKs per tenir en la regió extracel·lular un domini discoidina homòleg a la proteïna discoidina I de l'ameba *Dictyostelium discoideum*. Discoidina I és una lectina (d'unió a galactosa i N-acetil-galactosamina) amb un paper rellevant en el manteniment de la morfologia cel·lular, organització citoesquelètica i agregació cel·lular de *Dictyostelium* quan hi ha manca de nutrients al medi (Devine i Williams, 1982; Alexander *et al*, 1992). En humans, regions homòlogues al domini discoidina s'han identificat en proteïnes implicades en

adhesió cel.lular o interaccions cèl.lula-cèl.lula (Baumgartner *et al*, 1998) (Taula V). Aquestes proteïnes inclouen receptors de membrana (neuropilin 1 i 2, caspr/neurexin IV) i proteïnes de secreció (factors de coagulació V i VIII, AEBP1, RS1, MFGE8). A més, en ratolins, resultats recents suggereixen que els dominis discoidina de la proteïna SED1 estan implicats en la comunicació cèl.lula-matriu extracel.lular durant la unió de l'espermatozou a l'oòcit (Ensslin i Shur, 2003).



**Figura 1.** Esquema de les famílies estructurals de receptors tirosina quinasa en humans. El receptor prototípic de cada família s'indica a sobre el receptor i, a sota, els receptors representatius identificats. Els diferents dominis estructurals s'especifiquen a la dreta. Les abreviatures dels receptors prototípics són: DDR: receptor domini discoidina; TRK: receptor tirosina quinasa; RYK: receptor relacionat amb tirosina quinases; EGFR: receptor de creixement epidèrmic; ROR: receptor òrfan; PTK7: proteïna tirosina quinasa 7; FLT: receptor relacionat a fms (o receptor de factor de creixement endotelial); PDGFR: receptor de factor de creixement derivat de plaquetes; FGFR: receptor de creixement de fibroblasts; MUSK: receptor tirosina quinasa específica de múscul; LTK: tirosina quinasa de leucòcits; HGFR: receptor de creixement d'hepatòcits; TIE: receptor tirosina quinasa de cèl.lules endotel·lials; INSR: receptor d'insulina; EPH: receptor d'ephrin; ROS: RPTK expressat en determinades cèl.lules epitel·lials; LMR: lemur (modificada de Blume-Jensen i Hunter, 2001).

**Taula V.** Proteïnes amb domini discoidina en humans.

Gen	Proteïna	Núm. dominis <sup>a</sup>	Lligand	Funció	Referència
<b>Receptors de membrana</b>					
<i>CASPR4</i>	Proteïna de reconeixement cel.lular 4	1	Formació d'un complex trimèric amb contactina i neurofascina	Adhesió axó-glia en la regió paranodal	Peles <i>et al</i> , 1997
<i>DDR1</i>	Receptor domini discoidina 1	1	Col.lagen	Proliferació, migració, diferenciació cel.lular	Jonson <i>et al</i> , 1993
<i>DDR2</i>	Receptor domini discoidina 2	1	Col.lagen	Interacció cèl.lula-MEC <sup>b</sup>	Lai i Lemke, 1994
<i>NRP1</i>	Neuropilin-1	2	Semaforines classe 3 i VEGF <sub>165</sub> <sup>c</sup>	Adhesió cel.lular	He i Tessier-Lavigne, 1993
<i>NRP2</i>	Neuropilin-2	2	Semaforines classe 3 i VEGF <sub>165</sub>	Adhesió cel.lular	Kolodkin <i>et al</i> , 1997
<b>Proteïnes de secreció</b>					
<i>AEBP1</i>	Proteïna 1 d'unió a AE (proteïna tipus carboxipeptidasa aòrtica)	1	Matriu extracel.lular (desconegut)	Desenvolupament embrionari i reparació de teixits	Layne <i>et al</i> , 1998
<i>F5</i>	Factor de coagulació V	2	Factor Xa	Cofactor del Factor Xa; unió a fosfolípids	Kane i Davie, 1986
<i>F8</i>	Factor de coagulació VIII	2	Factor IXa	Cofactor del Factor IXa; unió a fosfolípids	Kane i Davie, 1986
<i>MFGE8</i>	Proteïna grassa de la llet – factor de creixement epidèrmic 8	2	Aminofosfolípids (cèl.lules apoptòtiques)	Factor d'unió entre cèl.lules apoptòtiques i fagòcits	Hanayama <i>et al</i> , 2002
<i>RS1</i>	RS1	2	(desconegut)	Proteïna d'adhesió cel.lular	Sauer <i>et al</i> , 1997

<sup>a</sup> Número de dominis discoidina; <sup>b</sup> MEC: matriu extracel.lular; <sup>c</sup> VEGF<sub>165</sub>: isoforma 165 del factor de creixement de l'endoteli vascular.

Fins el present moment, el col.lagen ha estat identificat com a lligand fisiològic dels DDRs (Shrivastava *et al*, 1997; Vogel *et al*, 1997), tot i que no es descarta la presència d'altres lligands o cofactors. El DDR1 s'activa per la unió al col.lagen de tipus I a VI i també VIII. Mentre que DDR2 només és activat per col.lagens fibril·lars, en particular el tipus I i III.

No obstant, les funcions biològiques dels DDRs són encara desconegudes. La supressió de l'expressió de *DDR1* en ratolins (ratolins *DDR1 null*) és compatible amb el desenvolupament embrionari (Vogel *et al*, 2001). Els animals tenen una mida i pes menor fins la pubertat. En les femelles, aquesta disminució es manté en l'edat adulta

i presenten defectes en la implantació de blastòcits i en el desenvolupament de la glàndula mamària. També presenten problemes motors en les extremitats posteriors i una malformació de l'oida externa (Vogel, comunicació personal). L'eliminació del gen *DDR2* en ratolins també dona lloc a animals viables, però amb clars defectes en el creixement postnatal dels ossos degut a una disminució en la proliferació dels condrocits (Labrador *et al*, 2001).

DDR1 és particularment abundant en cèl.lules epitelials. En humans, la seva expressió s'ha detectat en cervell, ronyó, pulmó, tracte gastrointestinal, glàndula mamària, illots de Langerhans, la tiroide i en fibroblasts de la còrnia i dermis (Alves *et al*, 1995; Mohan *et al*, 2001; Sakamoto *et al*, 2001). Mentre que l'expressió de *DDR2* es detecta en múscul esquelètic, cardíac i teixit connectiu (Vogel, 1999; Olaso *et al*, 2001).

Ambdós receptors presenten una expressió elevada en determinats tumors malignes. En concret, s'ha detectat una sobreexpressió de *DDR1* en càncer de mama, ovari, d'esòfag, pulmó, gliomes i tumors cerebrals infantils (Laval *et al*, 1994; Alves *et al*, 1995; Barker *et al*, 1995; Perez *et al*, 1996; Weiner *et al*, 1996; Nemoto *et al*, 1997; Weiner *et al*, 2000; Garber *et al*, 2001; Ji *et al*, 2002; Pomeroy *et al*, 2002).

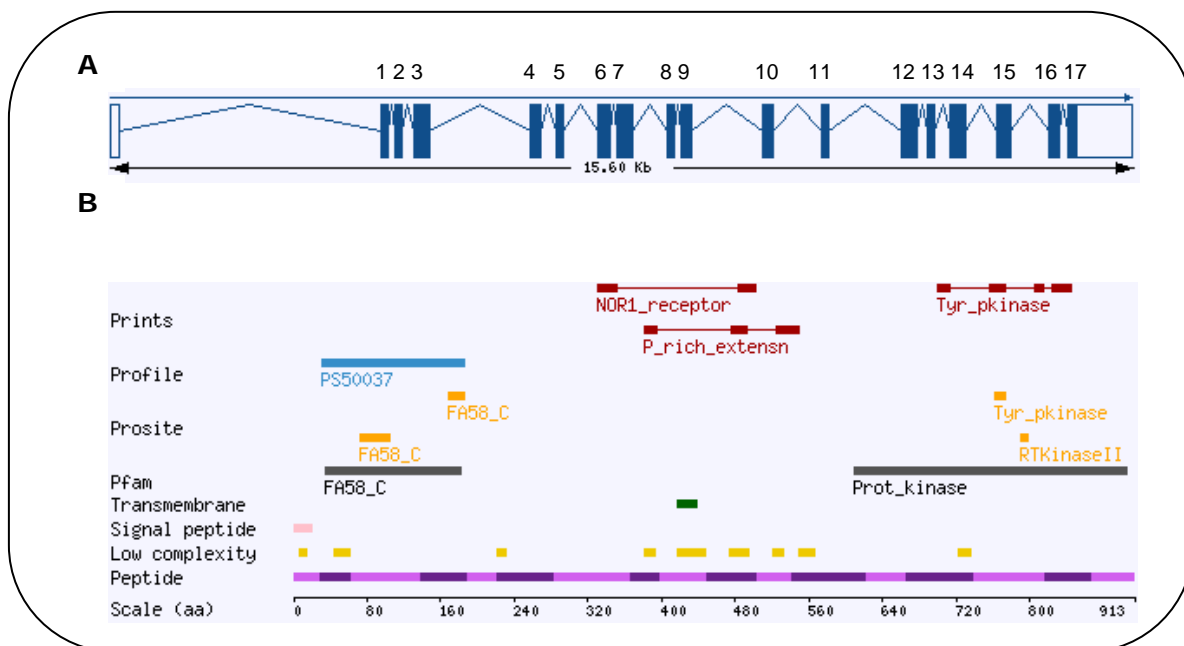
Mitjançant estudis d'hibridació *in situ* de seccions de tumors d'ovari i pulmó en humans, s'ha observat una expressió diferencial dels *DDRs* (Alves *et al*, 1995). *DDR1* es localitza en les cèl.lules tumorals invasives, mentre que *DDR2* en les cèl.lules circumdants al tumor, fet que indica una expressió coordinada entre *DDR1* i *DDR2* en la progressió del tumor. L'activació d'aquests receptors per col.lagen estimula la síntesi de metal.loproteases de matriu extracel.lular (MMPs) (Vogel *et al*, 1997; Hou *et al*, 2001; Hou *et al*, 2002). Aquestes són endopeptidases que degraden els components de la matriu extracel.lular (MEC). Les MMPs estan implicades en la remodelació tissular, processos de cicatrització, angiogènesi i invasió tumoral. Per això, s'ha suggerit que els *DDRs* actuarien com a sensors al col.lagen en la superfície cel.lular i estarien implicats en processos de diferenciació, proliferació i migració cel.lular (Vogel *et al*, 1997).

### 1.2.3. ESTRUCTURA GÈNICA I PROTEICA DE *DDR1*

*DDR1* es va clonar per primera vegada a partir d'una genoteca de cDNA de placenta humana utilitzant una sonda amb una seqüència consens del domini tirosina quinasa dels RPTK i va ser anomenat *DDR* (Johnson *et al*, 1993). Simultàniament, va ser identificat per altres grups i anomenat amb els noms *NEP* (Zerlin *et al*, 1993), *trkE* (Di Marco *et al*, 1993), *Ptk-3* (Sanchez *et al*, 1994), *Cak* (Perez *et al*, 1994), *RTK6* (Laval *et al*, 1994) i *MCK-10* (Alves *et al*, 1995). Gens homòlegs a *DDR1* s'han identificat en ratolí (Zerlin *et al*, 1993), rata (Sanchez *et al*, 1994), i *Caenorhabditis elegans* (Ruvkun i Hobert, 1998).





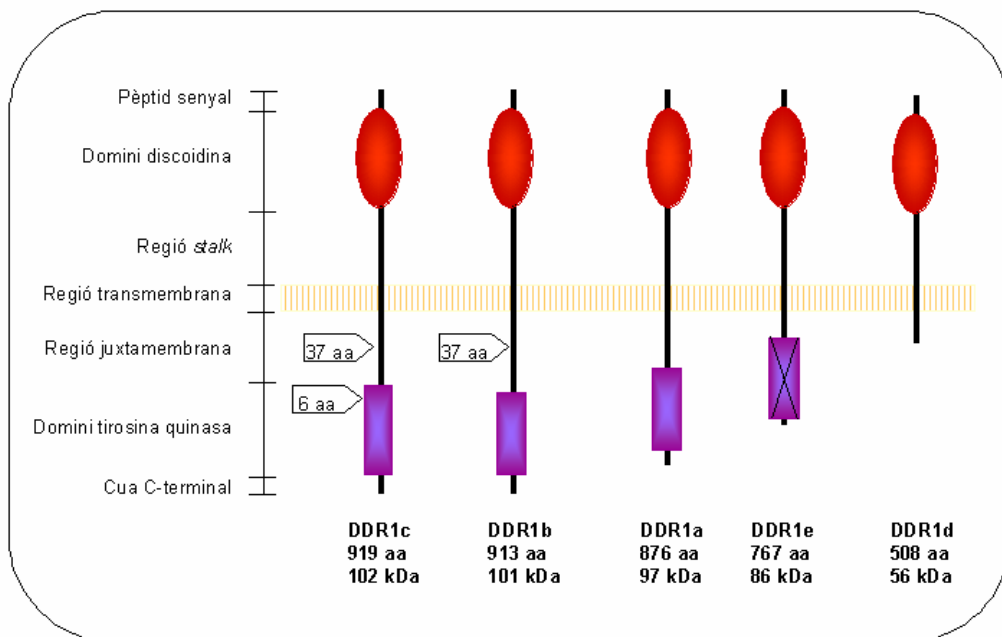


**Figura 3. Estructura gènica i característiques proteiques del DDR1.** A. Estructura exònica del gen *DDR1*. Les barres blaves corresponen als exons. Les barres blanques a les regions 5' i 3' UTR. B. Característiques proteiques d'acord amb la base de dades Vertebrate Genome Annotation (VEGA) ([http://vega.sanger.ac.uk/Homo\\_sapiens/geneview?gene=OTTHUMG00000016356&db=core](http://vega.sanger.ac.uk/Homo_sapiens/geneview?gene=OTTHUMG00000016356&db=core)).

Les isoformes, ordenades de més llarga a més curta, són les següents:

- i. *isoforma c*. Codifica per una proteïna de 919 aminoàcids, essent la isoforma més llarga i l'única que conté tota la regió codificant.
- ii. *isoforma b*. Codifica per una proteïna de 913 aminoàcids. L'extrem 5' de l'exó 14 conté un lloc acceptor de *splicing* alternatiu críptic, el qual dóna lloc a l'eliminació de les primeres 18 pb de l'exó 14 (codifiquen per 6 aminoàcids SFSLFS). Es desconeix quina és la conseqüència funcional de l'eliminació d'aquests 6 aminoàcids. Atès que l'expressió d'aquesta isoforma és la més abundant durant l'embriogènesi i la vida adulta, en aquest treball s'ha considerat la seva seqüència com la de referència per la posició de nucleòtids i aminoàcids.
- iii. *isoforma a*. Codifica per una proteïna de 875 aminoàcids. Eliminació per *splicing* alternatiu de les 18 pb entre els exons 13 i 14 i de l'exó 11 (37 aminoàcids). Com es descriu més endavant, l'exó 11 conté un motiu d'unió a dominis PTB i SH2.
- iv. *isoforma e*. Codifica per una proteïna de 767 aminoàcids. La regió exó-intró 5' de l'exó 10 té 2 llocs acceptors de *splicing* alternatiu. El més distal a 5' dóna lloc a l'eliminació d'una part de l'exó 10, 11 i 12. En aquest cas la seqüència jxtamembrana està alterada en 28 aminoàcids i hi ha una pèrdua del lloc d'unió a ATP (seqüència d'aminoàcids GEGAFG). Sense aquest lloc d'unió a ATP, DDR1 molt probablement perd l'activitat intrínseca tirosina quinasa.

v. *isoforma d*. Codifica per una proteïna de 508 aminoàcids. Els exons 11 i 12 són eliminats per *splicing* alternatiu. La unió de l'exó 10 amb el 13 provoca un canvi en el marc de lectura i una terminació prematura de la traducció. Dóna lloc a una proteïna ancorada a la membrana, però sense el domini tirosina quinasa i la cua C-terminal.



**Figura 4.** Esquema de les isoformes de DDR1. Isoformes ordenades de major a menor pes molecular. La isoforma b és la que es pren com referència.

L'expressió diferencial d'aquestes isoformes, tant a nivell gènic com proteic, i el seu significat biològic no ha estat estudiat en detall. S'ha observat, per exemple, que l'expressió de DDR1b augmenta amb el desenvolupament postnatal en comparació amb la isoforma DDR1a en rata (Sakuma *et al*, 1996). La isoforma DDR1a se sobre-exprés en determinats tumors mamaris (Perez *et al*, 1994) i macròfags infiltrant-se en tumors (Kamohara *et al*, 2001). Les isoformes DDR1d i DDR1e només s'han observat en línies cel·lulars de càncer de còlon (Alves *et al*, 2001).

Recentment s'ha observat que l'expressió i activació de DDR1 són induïdes per p53 en resposta a estrès genotòxic o dany del DNA (Ongusaha *et al*, 2003). Els mateixos autors han demostrat que la seqüència consens o element de resposta a p53 es localitza a l'intró 3 de *DDR1*, i no en la regió promotora tal i com havien proposat d'altres autors (Sakuma *et al*, 1996). No obstant, els mecanismes de regulació de l'expressió diferencial de les diferents isoformes i l'estabilitat dels mRNA estan encara per caracteritzar.

A continuació es fa una breu descripció dels dominis funcionals i estructurals i seqüències de rellevància per la funcionalitat del DDR1.

### **Regió extracel.lular: domini discoidina i regió *stalk***

El domini discoidina està constituït pels 160 aminoàcids N-terminals. La regió que va des del final del domini discoidina fins la regió transmembrana és anomenada regió *stalk* (Curat *et al*, 2001). Aquesta té 215 aminoàcids i no presenta homologia amb d'altres proteïnes. Ambdós, domini discoidina i regió *stalk*, són essencials per la unió a col.lagen, la dimerització i activació del receptor (Curat *et al*, 2001).

Els residus d'unió a col.lagen es troben en el domini discoidina i corresponen a aminoàcids conservats entre els dominis discoidina de les proteïnes incloses a la Taula VI (Abdulhussein *et al*, 2004). Curiosament, mutacions en aquests residus en la proteïna RS1 provoquen retinosis juvenil lligada al cromosoma X. Un 80% de les mutacions identificades en aquesta malaltia es localitzen en el domini discoidina del gen *RS1*, suggerint que aquest domini és crucial per la funcionalitat de la proteïna (The Retinoschisis Consortium, 1998).

La regió *stalk* conté dues seqüències rellevants per la funcionalitat del receptor:

- i. Una seqüència de glicosilació. Aquesta és necessària per la funcionalitat del receptor (Curat *et al*, 2001).
- ii. Una seqüència diana per una proteasa de superfície cel.lular. El DDR1, en resposta a la interacció amb el col.lagen, és processat per una proteasa de superfície de membrana en un fragment soluble, el qual pot difondre per la MEC, i un fragment ancorat a la membrana i que roman fosforilat durant dies (Vogel, 2002). El mecanisme i la implicació biològica del processament del DDR1 són encara desconeguts. Se sospita que, a l'igual que en d'altres receptors d'adhesió (VCAM-1, L-selectin, syndecan), receptors acoblats a proteïnes G (receptor de vasopressina V2) o precursors de lligands ancorats a membrana (TNF- $\alpha$ , Fas ligand), aquest processament proteolític és un mecanisme de regulació de l'activitat de DDR1.

### **Regió juxtamembrana**

La regió juxtamembrana està codificada pels exons 10, 11 i 12 i té 176 aminoàcids. Aquesta regió és la més afectada pels mecanismes de *splicing* alternatiu i conté diverses seqüències consens de fosforilació de tirosines que indica que és una regió important per la transducció del senyal.

L'exó 11, present en les isoformes DDR1b i DDR1c, conté el motiu LLXNPXY que és fosforilat en el residu tirosina per l'activació del domini tirosina quinasa en resposta a la unió a col.lagen. Aquest motiu correspon al motiu consens  $\psi$ XNPXpY (essent  $\psi$  un aminoàcid hidrofòbic) d'unió al domini PTB de la proteïna adaptora Shc. Recentment, s'ha demostrat que l'activació de la isoforma DDR1b té un paper important en la

diferenciació de monòcits a macròfags i aquesta depèn de la unió de Shc a DDR1, activant la via de la p38 MAPK (Matsuyama *et al*, 2003).

Hi ha d'altres tirosines conservades en l'exó 12 (tirosina 543 (YSGD) i tirosina 547 (YMEP)). Aquestes són potencials llocs d'unió a proteïnes amb un domini SH2 o PTB i, per tant, serien importants en la via de senyalització de DDR1. Fins el moment no s'ha demostrat la seva fosforilació o interacció amb d'altres proteïnes adaptadores. No obstant, s'ha observat que la proteïna FRS2, amb un domini PTB, interacciona amb la regió juxtamembrana de la isoforma DDR1a (Foehr *et al*, 2000).

L'homologia de la regió juxtamembrana entre DDR1 i DDR2 és baixa, però hi ha una seqüència de 25 aminoàcids (aminoàcids 566-590 del DDR1) localitzats en l'exó 12, que es manté pràcticament igual en ambdós receptors, suggerint que pot ser una regió crítica per la transducció del senyal d'ambdós DDRs (Alves *et al*, 2001).

### **Domini tirosina quinasa i cua C-terminal**

El domini tirosina quinasa presenta l'homologia més elevada (del 55-58%) amb els RPTK de neurotrofines (NTRK1, 2 i 3 i NGFR). No obstant, l'autofosforilació del receptor per unió a col.lagen té una cinètica completament diferent comparada amb la dels RPTKs. La fosforilació s'observa al cap d'unes hores en cèl.lules adherents i d'uns 30 minuts en cèl.lules en suspensió (Vogel *et al*, 1997; L'Hote *et al*, 2002) i es manté durant dies (Vogel, 2002). Si el DDR1 actua com a sensor al col.lagen en la superfície cel.lular, aquesta cinètica donaria el temps suficient per permetre l'establiment de les connexions cèl.lula-MEC necessàries pel posicionament cel.lular.

La funció de les isoformes sense activitat catalítica (DDR1d i DDR1e) són desconegudes. S'ha postulat que tindrien una funció reguladora del mateix receptor DDR1 per analogia a les isoformes truncades (sense domini tirosina quinasa) dels receptors NTRK2 i NTRK3 (Baxter *et al*, 1997; Palko *et al*, 1999; Alves *et al*, 2001).

En l'extrem C-terminal de les isoformes DDR1a, DDR1b, DDR1c hi ha un motiu LNTV. Aquest motiu pot actuar com a lloc d'unió a les proteïnes amb un o varis dominis PDZ. Els dominis PDZ són punts d'interacció entre proteïnes i faciliten l'ancoratge i localització de complexos proteics (Garner *et al*, 2000). Per exemple, s'ha demostrat que els dominis PDZ són essencials per l'agrupació i localització de receptors eph i HER2 (Borg *et al*, 2000). Curiosament, el canvi en el marc de lectura de la isoforma DDR1d dóna lloc a l'aparició del motiu GAPV que també és lloc d'unió a PDZ (Alves *et al*, 2001).

La presència de mutacions en qualsevol dels dominis mencionats, així com en les seqüències de *splicing* alternatiu, podrien afectar la funcionalitat del receptor. De moment, però, no s'han descrit mutacions en el gen *DDR1* associades a cap malaltia o tret fenotípic.

#### **1.2.4. IMPLICACIÓ DE DDR1 EN EL SISTEMA NERVIÓS CENTRAL**

En sistema nerviós central (SNC), l'expressió de DDR1 ha estat estudiada en els models múrids de ratolí (Zerlin *et al*, 1993; Bhatt *et al*, 2000) i rata (Sanchez *et al*, 1994) mitjançant les tècniques d'hibridació *in situ* i RT-PCR. En aquests treballs s'observa una expressió de DDR1 en zones proliferatives durant el desenvolupament embrionari i en zones de neurogènesi secundària en estadis adults. Sanchez i col.laboradors també detecten expressió de DDR1 en substància blanca.

El nostre grup ha estudiat l'expressió de DDR1 durant el neurodesenvolupament i en l'adult mitjançant la tècnica de la hibridació *in situ* en talls de cervell de ratolí (estudi realitzat per la doctoranda Neus Franco). El patró d'expressió que observem és comparable amb el dels estudis anteriors. No obstant, en l'animal postnatal i en l'adult detectem una expressió en una subpoblació cel.lular compatible amb oligodendròcits. Actualment estem treballant per esbrinar de quina subpoblació es tracta.

A finals del 2000, es va publicar el primer i únic estudi que posa de manifest la importància de DDR1 en el desenvolupament del SNC (Bhatt *et al*, 2000). Aquest estudi demostra que la interacció de DDR1 amb el col.lagen i la corresponent via de senyalització activada és essencial per la formació axonal de les cèl.lules granulars durant el desenvolupament del cerebel. El fet que DDR1 s'expressi també durant l'edat adulta, va fer suggerir a Bhatt i col.laboradors que DDR1 estaria implicat en processos de manteniment d'estructures neuronals, regulació de la formació de sinapsis i en els processos de *sprouting*. Precisament, en el SNC de l'adult, aquestes tasques són realitzades per l'astroglia i oligodendroglia (Fields i Stevens-Graham, 2002).

D'altra banda, hi ha varis treballs puntuals i independents que demostren la modulació de la proliferació i diferenciació de cèl.lules neuroepitelials i astroglia per col.lagen (Goetschy *et al*, 1987; Ali *et al*, 1998). Però cap d'aquests estudis explora la possible participació del DDR1.

#### **1.2.5. IMPLICACIÓ DE DDR1 EN PROCESSOS D'ADHESIÓ, MIGRACIÓ, DIFERENCIACIÓ I PROLIFERACIÓ CEL.LULAR**

El conjunt de macromolècules que constitueixen la MEC no només té un paper important en el manteniment de la integritat dels teixits i els òrgans, sinó també en l'adhesió, la proliferació, la migració i la diferenciació de les cèl.lules que envolten. És més, les interaccions cèl.lula-MEC tenen implicacions molt importants durant el desenvolupament i morfogènesi.

El component majoritari de la MEC és el col.lagen. S'han descrit 3 tipus de receptors de col.lagen: les integrines, els DDRs i la glicoproteïna VI (Vogel, 2001). Els tres

activen diverses vies de transmissió del senyal implicades en processos d'adhesió, proliferació, migració i diferenciació cel·lulars. La glicoproteïna VI s'expressa exclusivament en plaquetes, mentre que integrines i DDRs es troben expressats en diferents tipus cel·lulars i tenen la capacitat de monitoratge i modulació de la integritat de la MEC mitjançant l'activació de mecanismes de degradació o síntesi de col·lagen. En concret, l'expressió de DDR1 i de col·lagen de tipus I, III i VIII té lloc en paral·lel en cèl·lules musculars de rata (Hou *et al*, 2001). Per altra banda, en cèl·lules de ratolí *DDR1 null* s'observa una inhibició de l'expressió de metal·loproteases (MMP-2 i MMP-9) (Hou *et al*, 2001). Ambdues observacions són clars indicadors del paper de DDR1 en la migració i adhesió cel·lular.

Fins el present moment, s'ha demostrat que la unió de col·lagen a DDR1 promou la diferenciació de cèl·lules musculars esquelètiques (Vogel *et al*, 2000), l'adhesió, proliferació i migració de cèl·lules musculars llises en el procés de reparació de la paret arterial (Hou *et al*, 2001), l'adhesió cel·lular i inhibició de la proliferació cel·lular de cèl·lules mesangials i epitel·lials (Curat i Vogel, 2002), la diferenciació de monòcits a macròfags (Matsuyama *et al*, 2003b) i alliberació de citoquines per macròfags (Matsuyama *et al*, 2004). I, com ja s'ha mencionat, també promou l'elongació dels axons en cèl·lules granulars de cerebel (Bhatt *et al*, 2000).

No obstant, en diverses línies cel·lulars, no es detecta una activació de DDR1 per col·lagen, suggerint que aquesta activació requereix de la presència de cofactors o lligands addicionals (Foehr *et al*, 2000; L'hote *et al*, 2002). Recentment, s'ha demostrat que la proteïna Wnt-5a és co-factor necessari per l'activació del DDR1 per col·lagen en l'adhesió i migració de cèl·lules epitel·lials de mama HB2 (Jonsson i Andersson, 2002; Dejmek *et al*, 2003), implicant el DDR1 en la via de senyalització de les proteïnes Wnt. Aquestes són glicoproteïnes de secreció autocrina o paracrina que modulen gran varietat de processos cel·lulars (diferenciació, polarització, migració i proliferació) tant durant el desenvolupament com en l'adult (Miller, 2001).

## 1.3. MÈTODES DE GENOTIPATGE DE POLIMORFISMES D'UN ÚNIC NUCLEÒTID (SNPs)

### 1.3.1. POLIMORFISMES D'UN ÚNIC NUCLEÒTID

Els SNPs són substitucions d'un sol nucleòtid en la seqüència de DNA que resulten en la generació de variants de seqüència (o al·lels), les quals han de coexistir en la població i la freqüència de l'al·lel menys representat ha de ser superior a l'1%. En principi, els SNPs podrien ser bi, tri o tetra-al·lèlics. No obstant, en humans, tan sols s'han detectat SNPs bi-al·lèlics en DNA genòmic. Els SNP tri i tetra-al·lèlics tan sols s'han identificat en molt baixa freqüència en el genoma mitocondrial (Kogelnik *et al*, 1998). Les freqüències de les diferents substitucions nucleotídiques possibles no són equivalents. Així, unes 2/3 parts dels SNPs impliquen la transició C/T (o G/A, en la cadena complementària), mentre que les transversions s'observen en igual proporció. L'elevat percentatge de SNPs C/T (o G/A) es deu probablement a la reacció de deaminació de la 5-metilcitosina resultant en timina (Holliday i Grigg, 1993).

El 90% dels polimorfismes observats en el genoma humà són SNPs. S'ha estimat que la seva distribució és d'un SNP per cada 1000 pb de DNA (Sachidanandam *et al*, 2001). El genoma haploide humà consta d'unes  $3.3 \times 10^9$  pb, de manera que, un individu qualsevol tindrà més de 3 milions de SNPs en el seu genoma. Fins el present moment, hi ha registrats més de 4 milions de SNPs candidats en la base de dades d'accés públic dbSNP mantinguda pel National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/SNP/>). Molts d'aquests s'han identificat *in silico* o en mostres aïllades i estan essent verificats i les seves freqüències estudiades en diverses poblacions.

A finals dels 90, els SNPs es van proposar com a marcadors genètics per la identificació de *loci* associats a malalties complexes i a la farmacogenètica. L'estudi de la variabilitat en la resposta a fàrmacs determinada genèticament, o farmacogenètica, és anàloga a l'estudi de malalties complexes en quant a la implicació de factors genètics i ambientals. Fins aquell moment, els polimorfismes més estudiats havien estat els microsatèl·lits o STRPs (*Short Tandem Repeat Polymorphisms*). Aquests són repeticions en tàndem d'una unitat bàsica de seqüència de DNA, d'entre dos i sis nucleòtids. Són polimorfismes abundants i presenten un elevat nombre d'al·lels. Hom estima que hi ha un STRP per cada 30-60 kb de DNA. El fet de què els STRPs fossin més informatius que els SNPs (*loci* multi-al·lèlics versus bi-al·lèlics, respectivament), tinguessin una distribució homogènia en el genoma i es poguessin analitzar fent ús de la tècnica de la PCR, va promoure el desenvolupament de sistemes semi-automàtics d'anàlisi de STRP (Livak *et al*, 1995; Hall *et al*, 1996). Els STRPs van esdevenir els marcadors genètics ideals en els estudis de lligament per la identificació de gens responsables de malalties monogèniques.

Ara bé, el viratge a finals dels 90 per l'estudi de malalties complexes i farmacogenètica, va comportar un canvi en els dissenys experimentals. En ambdós casos, com ja s'ha comentat anteriorment, els estudis d'associació poden ser més efectius per detectar gens d'efecte moderat o lleu que els estudis de lligament (Risch i Merikangas, 1996). En aquests estudis, els SNPs són els marcadors genètics més adequats a estudiar pels següents motius:

1. Menor taxa de mutació que els STRPs ( $\sim 10^{-8}$  i  $\sim 10^{-5}$ - $10^{-2}$ , per SNPs i STRPs, respectivament (Weber i Wong, 1993; Chakraborty *et al*, 1997; Li *et al*, 1996)). L'elevada variabilitat o múltiples al·lels observats en els STRPs reflecteix, en part, una elevada taxa de mutació, la qual pot ser un factor confusionari en els estudis d'associació.
2. Elevada densitat en el genoma humà (1 cada 1000 pb) que permet la caracterització d'un mapa d'haplotips per la identificació de gens de susceptibilitat per malalties complexes (Sachidanandam *et al*, 2001).
3. Disponibilitat d'un gran ventall de tècniques d'identificació i rastreig d'SNPs.
4. Implicació directa en la malaltia o tret fenotípic d'aquells SNPs amb conseqüències funcionals (afectant les regions codificants o reguladores del gen).

L'interès en l'anàlisi d'SNPs rau en l'esperança que aquests seran una eina poderosa en el procés d'identificació de gens de susceptibilitat de malalties complexes, en la farmacogenètica i en d'altres àrees de la biomedicina. No obstant, hi ha una gran disparitat entre el ritme amb què es descobreixen nous SNPs i el desenvolupament de noves tècniques que permetin l'anàlisi d'un gran nombre d'aquests. Malgrat l'existència de múltiples mètodes de genotipatge, hi ha una demanda de nous mètodes capaços d'analitzar gran nombre d'SNPs de forma fiable, ràpida i amb un baix cost. En aquest context, la doctoranda ha estat treballant en el desenvolupament nous protocols de genotipatge fent ús de la tecnologia dels *eTag*<sup>TM</sup> *reporters* a ACLARA BioSciences Inc.

Els múltiples mètodes de genotipatge d'SNPs actuals han evolucionat a partir de diferents principis bàsics molts dels quals es coneixien des de fa més de 20 anys. Com per exemple, la hibridació al·lel específica d'oligonucleòtids o la restricció enzimàtica al·lel específica (Smith i Wilcox, 1970; Kafatos *et al*, 1979). No obstant, no ha estat fins el descobriment de la PCR que el genotipatge s'ha pogut realitzar a mitjana i gran escala. La majoria dels mètodes actuals requereixen de l'amplificació prèvia per PCR de les seqüències que contenen els SNPs d'interès. Hi ha mètodes alternatius (que no depenen de PCR), però requereixen de quantitats més grans de DNA, el qual és un factor limitant quan es treballa amb mostres humanes.

A continuació es revisen els principis bàsics per la discriminació al·lèlica i es mencionen les tècniques de genotipatge d'SNPs d'ús actual.



### 1.3.2. TÈCNiques DE GENOTIPATGE D'SNPS: REQUERIMENTS

S'han desenvolupat un gran nombre de mètodes de genotipatge. La tria d'un o altre mètode depèn dels objectius experimentals i recursos disponibles de l'investigador o grup de recerca. Per exemple, un projecte pot requerir el genotipatge d'un nombre limitat d'SNPs en una mostra de mida gran o bé de molts SNPs en un o pocs individus. La flexibilitat en el disseny, l'habilitat de quantificar freqüències al·lèliques en *pools* de mostres de DNA i d'analitzar més d'un SNP per reacció, així com la semi-automatització del procés, són alguns dels aspectes a considerar en l'avaluació d'una tècnica de genotipatge. En resum, les propietats del mètode de genotipatge ideal capaç d'acomplir amb totes les necessitats actuals serien les següents:

- Específic
- Robust, fins i tot en presència de poca quantitat de DNA
- Disseny senzill i flexible
- Assaig homogeni, en un sol tub o pouet
- Mínima optimització de la reacció
- Quantitatiu, permet l'anàlisi en *pools* de mostres
- Multiplicador, permet l'anàlisi de més d'un SNP per reacció
- Assaig de fàcil automatització
- Automatització de l'anàlisi de resultats
- Assaig escalable i adaptable (processament des de baix a elevat nombre de mostres)
- Cost baix per genotip (reactius, instruments, manteniment i mà d'obra inclosos)

### 1.3.3. ESTRATÈGIES DE DISCRIMINACIÓ AL·LÈLICA

Les estratègies bàsiques emprades en la discriminació al·lèlica d'SNPs es poden agrupar en quatre categories. Aquestes són:

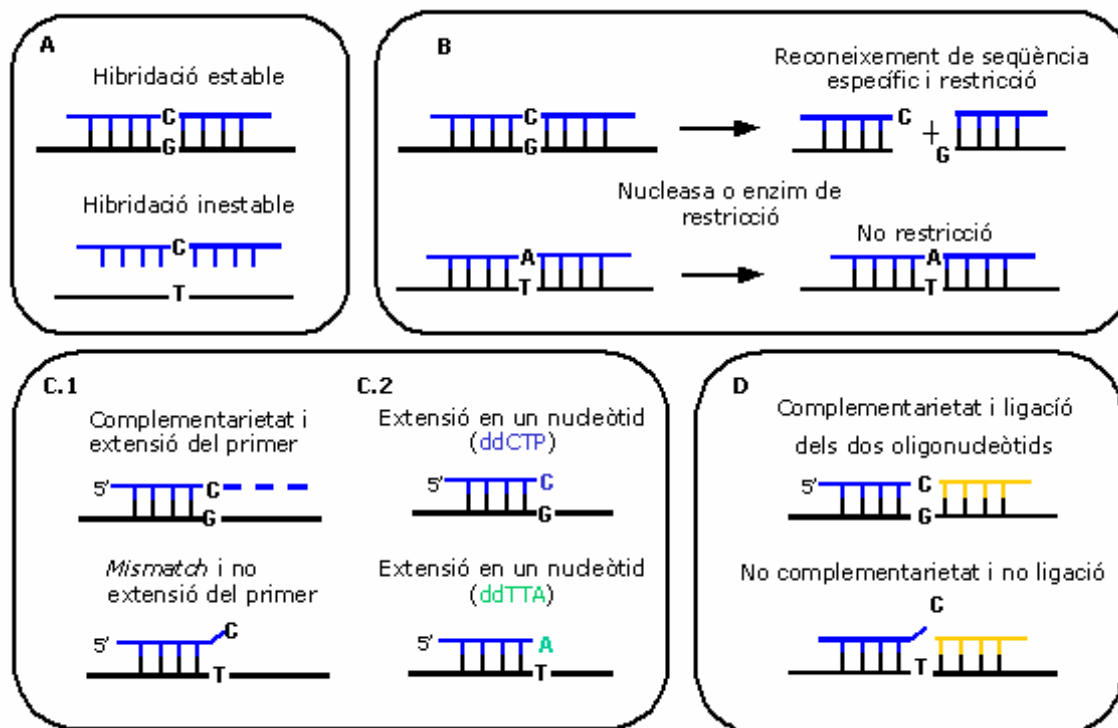
1. Hibridació al·lèl específica
2. Discriminació al·lèl específica per una endonucleasa
3. Extensió al·lèl específica d'un oligonucleòtid o *primer*
4. Ligació al·lèl específica d'oligonucleòtids

Cadascuna d'aquestes estratègies pot acoblar-se a diferents mètodes de detecció els quals detecten canvis en la massa, la conductivitat elèctrica o l'emissió de llum de molècules presents en els productes al·lèl específics. La detecció de fluorescència és potser el mètode de detecció més utilitzat. El marcatge fluorescent pot acoblar-se fàcilment a la majoria de tècniques i disposa de múltiples plataformes de mesura (lectors de plaques, electroforesi de gel i capil·lar i *microarrays* de DNA). L'espectrofotometria de masses i la quimioluminiscència (on l'emissió de llum és resultat d'una reacció bioquímica, com és el cas particular de la tècnica Pyrosequencing<sup>®</sup>) són dos sistemes de detecció que estan guanyant acceptació en els últims tres anys. Tots tres mètodes de detecció són compatibles amb el desenvolupament de protocols que permeten el rastreig semi- i automatitzat d'SNPs a mitjana i gran escala.

El format d'aquestes estratègies és també variable i molts dels mètodes poden realitzar-se en solució o bé en fase sòlida (immobilitzant part dels components de la reacció en membranes de niló, *microarrays*, microesferes o plaques de 96 o 384 pouets).

### 1. Hibridació al.llel específica

Es basa en les diferències en l'estabilitat entre dos cadenes de DNA perfectament complementàries o amb un nucleòtid desparellat. Sota unes condicions determinades d'hibridació (temperatura, tampó de reacció), una sonda o oligonucleòtid de seqüència curta (entre 10 i 30 nucleòtids) formarà un dúplex amb el DNA diana només si són perfectament complementaris (Figura 5.A).



**Figura 5.** Estratègies bàsiques emprades en la discriminació al·lèlica d'SNPs. A. Hibridació al.llel específica; B. Discriminació al.llel específica per una endonucleasa; C.1. Extensió al.llel específica d'un oligonucleòtid o *primer*; C.2. Incorporació específica d'un dideoxinucleòtid; D. Ligació d'oligonucleòtids al.llel específica.

Els primers mètodes d'hibridació al.llel específica van sorgir a partir de les tècniques del *dot blot* (Conner *et al*, 1983) i *reverse dot blot* (Saiki *et al*, 1989). Aquesta última pot considerar-se com la precursora dels *microarrays* pel genotipatge simultani de centenars d'SNPs. Les tècniques basades en la hibridació al.llel específica es resumeixen en la Taula VI. Es diferencien per com té lloc la discriminació al·lèlica, el sistema de detecció i el format de la reacció. Són tècniques que poden automatitzar-se fàcilment i permeten l'anàlisi de múltiples mostres simultàniament. Les dues tècniques més populars són TaqMan® (Applied Biosystems, FosterCity, CA, EUA)

(Livak *et al*, 1995; 1999) i Molecular Beacons (Tyagi i Kramer, 1996; Tyagi *et al*, 1998). En aquestes, la discriminació al·lèlica té lloc durant l'amplificació del DNA per PCR. Es tracta, doncs, d'assaigs homogenis i els resultats poden obtenir-se en menys de 2 hores. No obstant, no permeten l'anàlisi simultània de múltiples SNPs ja que les condicions òptimes d'hibridació són diferents per cada SNP. I, d'altra banda, no hi ha fluorocroms o molècules de marcatge suficients com per analitzar més d'1 o 2 SNPs per reacció. Malgrat la utilització dels *microarrays* (com per exemple, el GeneChip d'Affymetrix, Santa Clara, CA, EUA) pot semblar una solució a aquests dos problemes, varis estudis han demostrat que aquests no poden diferenciar els genotips homozigots dels heterozigots per un percentatge d'SNPs considerable (2–12%) (Wang *et al*, 1998; Cargill *et al*, 1999; Halushka *et al*, 1999).

**Taula VI.** Tècniques de discriminació al·lèlica basades en la hibridació al·lèl específica.

Nom (empresa) <sup>a</sup>	Mètode de detecció	Format	Avantatge	Desavantatge	Referència
Reverse dot blot	F <sup>b</sup> , Q <sup>c</sup> , C <sup>d</sup>	FS <sup>e</sup>	Multiplicatiu	Soroll de fons per hibridació inespecífica	Saiki <i>et al</i> , 1989
TaqMan (Applied Biosystems)	FR <sup>f</sup>	S <sup>g</sup>	Homogeni, simple, ràpid, automatitzable	Elevat cost	Livak <i>et al</i> , 1995
Molecular Beacons	FR	S	Homogeni, simple i ràpid	Elevat cost	Tyagi i Kramer, 1996; Tyagi <i>et al</i> , 1998
Microarray GeneChip (Affymetrix)	F	FS	Multiplicatiu	Elevat cost i percentatge elevat de resultats erronis per hibridació inespecífica	Wang <i>et al</i> , 1998; Cargill <i>et al</i> , 1999
DASH <sup>h</sup> (ThermoHybaid)	F	S, FS	Baix cost	Disseny complex	Prince <i>et al</i> , 2001

<sup>a</sup>Indica l'empresa que comercialitza la tècnica; <sup>b</sup>F, fluorescència; <sup>c</sup>Q, quimioluminiscència; <sup>d</sup>C, colorimetria; <sup>e</sup>FS, fase sòlida; <sup>f</sup>FR, transferència d'energia per ressonància de fluorescència; <sup>g</sup>S, fase en solució; <sup>h</sup>DASH, *dynamic allele-specific hybridization*.

Per tal de millorar la discriminació al·lèlica, s'han utilitzat anàlegs de nucleòtids com els àcids nucleics peptídics (PNA, *peptide nucleic acid*) o bé els àcids nucleics tancats (LNA, *locked nucleic acid*) en la síntesi de les sondes (Ross *et al*, 1997; Orum *et al*, 1999). De la mateixa manera, les sondes TaqMan també han estat modificades afegint en l'extrem 3' una molècula d'unió al solc petit del DNA (MGB, *minor groove-binder molecule*) (Kuimelis *et al*, 1997). Aquestes modificacions incrementen la temperatura de fusió (T<sub>m</sub>) de les sondes i, per tant, milloren la capacitat discriminatòria de les mateixes, però encareixen considerablement el cost de la reacció.

## 2. Discriminació al.lél específica per una endonucleasa

Es basen en la capacitat que tenen algunes endonucleases en reconèixer i tallar seqüències específiques de DNA (Figura 5.B, Taula VII). La tècnica més emprada en els laboratoris de genotipatge ha estat la que combina l'amplificació per PCR de la regió de DNA que conté l'SNP amb la posterior digestió amb un enzim de restricció (PCR-RFLP). No obstant, aquesta està essent substituïda per d'altres ja que no sempre es disposa d'una endonucleasa o enzim de restricció per la seqüència d'interès, és un procés llarg (la restricció enzimàtica requereix d'incubacions de 2 a 10 h) i de difícil automatització.

La tècnica Invader o de digestió invasiva per una 5'endonucleasa es descriu en detall més endavant (vegeu apartat *eTag™ Multiplex Invader® SNP Assay*: un nou protocol per l'anàlisi simultània d'SNPs).

**Taula VII.** Tècniques de discriminació al.lèlica basades en l'acció d'una nucleasa específica de seqüència.

Nom (empresa) <sup>a</sup>	Mètode de detecció	Format	Avantatge	Desavantatge	Referència
PCR-RFLP <sup>b</sup>	F <sup>c</sup> , Q <sup>d</sup> , C <sup>e</sup>	S <sup>f</sup>	Barat; no requereix d'equips sofisticats	Laboriós	Hixson i Vernier, 1990
Invader (ThirdWave Technologies Inc)	FR <sup>g</sup>	S	Homogeni; no requereix d'amplificació del DNA d'interès	Requereix gran quantitat de DNA genòmic	Lyamichev <i>et al</i> , 1999
McSNP <sup>h</sup>	F	S	Homogeni; barat	Laboriós	Akey <i>et al</i> , 2001

<sup>a</sup>Indica l'empresa que comercialitza la tècnica; <sup>b</sup>PCR-RFLP, *polymerase chain reaction-restriction length polymorphism*; <sup>c</sup>F, fluorescència; <sup>d</sup>Q, quimioluminiscència; <sup>e</sup>C, colorimetria; <sup>f</sup>S, fase en solució; <sup>g</sup>FR, transferència d'energia per ressonància de fluorescència; <sup>h</sup>McSNP, *Melting Curve Single Nucleotide Polymorphism*.

## 3. Extensió al.lél específica d'un oligonucleòtid o primer

Es basa en la precisió de la incorporació de nucleòtids complementaris a una seqüència diana de DNA per una DNA polimerasa (Sylvänen *et al*, 1990). Hi ha dos tipus de mètodes els quals es descriuen a continuació (Figura 5.C, Taula VIII).

### 3.1. Amplificació al.lél específica per PCR

Aquesta també es coneix amb el nom de ARMS (*Amplification Refractory Mutation System*) (Newton *et al*, 1989), ASA (*Allele Specific Amplification*) (Okayama *et al*, 1989), ASPCR (*Allele-Specific PCR*) (Wu *et al*, 1989) i PASA (*PCR Amplification of Specific Alleles*) (Sommer *et al*, 1989). Es basa en què l'extensió de l'encebador o oligonucleòtid durant la PCR depèn de la complementarietat del seu extrem 3' a la seqüència diana (Figura 7.C.1). Malgrat la simplicitat i baix cost d'aquest mètode, ja que no cal fer cap tractament previ o posterior a la PCR, no ha estat gaire utilitzat ja

que les DNA polimerases sovint donen lloc a falsos positius. El disseny dels oligonucleòtids al·lel específics no és simple i cal, sovint, optimitzar el protocol de PCR per cadascun dels SNPs a analitzar. Tot i que no és objectiu del present treball, mencionar que la doctoranda ha desenvolupat un protocol pel genotipatge simultani d'SNPs combinant els mètodes TaqMan i ARMS amb el marcatge dels *eTag reporters* (Virgos i Cronin, PCT/US02/32867).

### 3.2. Incorporació al·lel específica de nucleòtids

En aquest disseny, l'oligonucleòtid híbrida amb el seu extrem 3' en un nucleòtid *upstream* respecte la posició de l'SNP i es detecta el tipus de nucleòtid incorporat (Figura 5.C.2). Per tal d'evitar l'elongació de l'oligonucleòtid en més d'una base, s'utilitzen dideoxinucleòtids en la barreja de PCR (Syvänen *et al*, 1990). Cadascun dels quatre dideoxinucleòtids està marcat amb un fluorocrom diferent, de manera que el genotip es dedueix per anàlisi de la llum emesa pel dideoxinucleòtid incorporat (Pastinen *et al*, 1997).

**Taula VIII.** Tècniques de discriminació al·lèlica basades en l'extensió d'un oligonucleòtid per una DNA polimerasa. Incloses només aquelles tècniques amb més capacitat multiplicadora.

Nom (empresa) <sup>a</sup>	Mètode de detecció	Format	Avantatge	Desavantatge	Referència
<b>Amplificació al·lel específica</b>					
ARMS <sup>b</sup>	F <sup>c</sup> , FR <sup>d</sup> , C <sup>e</sup>	S <sup>f</sup>	Baix cost, homogeni	Disseny complexe	Newton <i>et al</i> , 1989
<b>Incorporació al·lel específica de nucleòtids</b>					
SnaPshot (Applied Biosystems)	F	S <sup>g</sup>	Multiplicador	Purificació i quantificació del producte de PCR	www.appliedbiosystems.com
Pyrosequencing (Pyrosequencing)	LI <sup>h</sup>	S	Multiplicador, capaç de seqüenciar fins 50 bases	Elevat cost i disseny complexe	Nyren <i>et al</i> , 1993
SNP-it, SNPstream (Orchid Biosciences, Luminex)	C	FS	Multiplicador, barat	Múltiples reaccions	Nikiforov <i>et al</i> , 1994
MassExtend, MassArray (Sequenom)	MS <sup>i</sup>	FS	Multiplicador	Elevat cost (instrumentació)	Braun <i>et al</i> , 1997
Microarray universal GenFlex (Affymetrix)	F	S	Multiplicador, disseny flexible	Elevat cost (reacció i instrumentació)	Fan <i>et al</i> , 2000

<sup>a</sup>Indica l'empresa que comercialitza la tècnica; <sup>b</sup>ARMS, *Amplification refractory mutation system*; <sup>c</sup>F, Fluorescència; <sup>d</sup>FR, Transferència d'energia per ressonància de fluorescència; <sup>e</sup>C, Colorimetria; <sup>f</sup>S, Fase en solució; <sup>g</sup>FS, Fase sòlida; <sup>h</sup>LI, llum; <sup>i</sup>MS, Espectrofotometria de masses.

Actualment, el mètode d'extensió al·lel específica d'un oligonucleòtid és el mètode més emprat pels grups de genotipatge d'SNPs ja que el disseny és simple i ofereix una gran diversitat de sistemes de detecció (Syvänen, 1999). Es tracta d'un mètode altament específic, quantitatiu, multiplicador i escalable, però que requereix de la

purificació de la reacció de PCR i, molt sovint també, de la reacció d'incorporació al.llel específica de nucleòtids prèviament a la detecció i lectura dels genotips.

#### **4. Ligació d'oligonucleòtids al.llel específica**

Es basa en la unió covalent de dos oligonucleòtids adjacents per una DNA ligasa, si aquests són perfectament complementaris al DNA diana en la regió de ligació (Wu i Wallace, 1989) (Figura 5.D). El disseny consta de dos oligonucleòtids, un dels quals és al.llel específic en el seu extrem 3' o 5'. La detecció del producte resultat de la ligació és indicatiu de la presència de l'al.llel complementari a l'oligonucleòtid al.llel específic. Els mètodes basats en la ligació al.llel específica d'oligonucleòtids es poden classificar en dues categories:

4.1. Assaig de la ligació d'oligonucleòtids (Landegren *et al*, 1988) i que generalment inclou l'amplificació posterior dels oligonucleòtids ligats. Recentment, s'han desenvolupat dos mètodes basats en aquesta aproximació i que permeten l'anàlisi de múltiples SNPs per reacció. Aquests són l'assaig SNPlex™ (<http://www.appliedbiosystems.com>) i SNPWave™ (van Eijk *et al*, 2004; <http://www.keygene.com>).

4.2. Assaig de ligació i amplificació per gir-volta del cercle (Nilsson *et al*, 1994; Lizardi *et al*, 1998), que no depèn de la PCR. El problema que presenten aquests mètodes és el cost de la DNA ligasa i del marcatge dels múltiples oligonucleòtids.

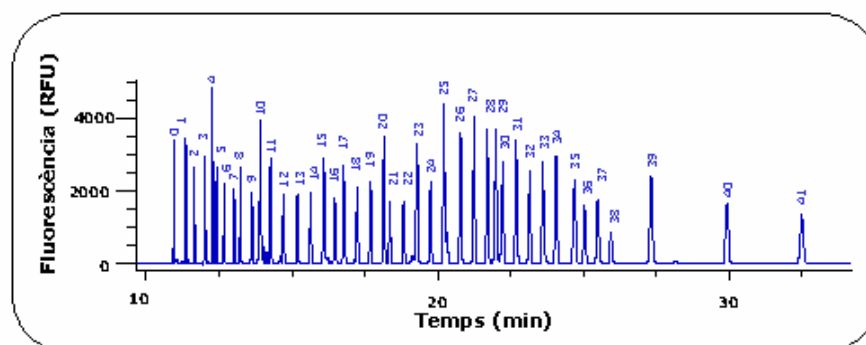
#### **1.3.4. eTAG™ MULTIPLEX INVADER® SNP ASSAY: UN NOU PROTOCOL PER L'ANÀLISI SIMULTÀNIA D'SNPs**

##### **Què són els eTag reporters?**

Els *eTag reporters* són molècules fluorescents de baix pes molecular (150-5.000 da) amb una relació càrrega/massa única que permet la seva separació per electroforesi capil·lar (EC) (Figura 6). Els *eTag reporters* es poden acoblar a sondes, siguin aquestes oligonucleòtids, pèptids o proteïnes. Quan la sonda interacciona amb la corresponent diana, l'*eTag reporter* és alliberat a la solució i separat per EC (Singh, USPTO 6,514,700). Així, doncs, la identificació d'un *eTag reporter* concret és indicatiu de la presència d'un determinat component en la reacció. ACLARA BioSciences Inc. (<http://www.aclara.com>) ha desenvolupat una llibreria de més d'un centenar d'*eTag reporters*. Aquests es poden acoblar a múltiples sondes i permeten la identificació simultània de diferents components en una mateixa reacció ja sigui per l'anàlisi d'àcids nucleics (DNA i RNA) o proteïnes.

La separació i quantificació dels *eTag reporters* té lloc mitjançant la utilització d'aparells d'electroforesi capil·lar (EC) o seqüenciadors automàtics (com per exemple, ABI PRISM 3100 Genetic Analyzer (Applied Biosystems), MegaBACE1000(APBiotech)). L'anàlisi dels resultats pot fer-se de forma semi-

automàtica utilitzant el programa eTag Informer™ desenvolupat a ACLARA BioSciences Inc.



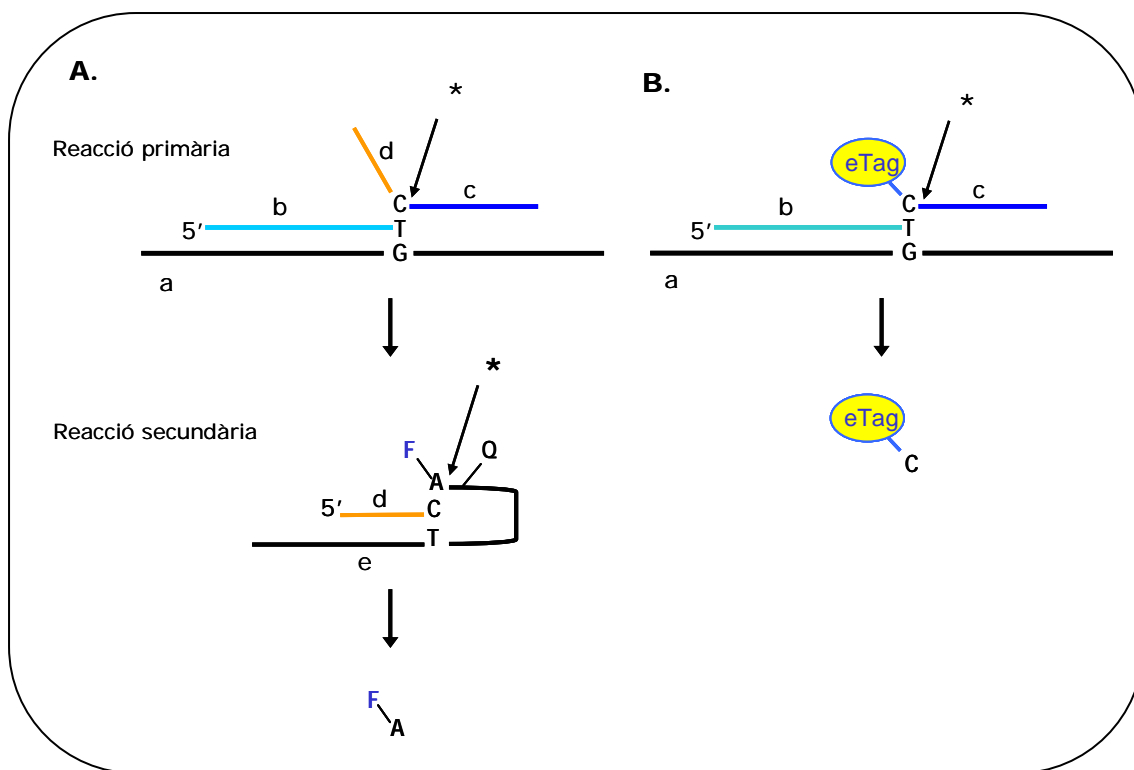
**Figura 6.** Electroforograma amb 42 *eTag reporters*. Separació electroforètica realitzada en un ABI3100 Genetic Analyzer (Applied Biosystems).

### **Combinació dels *eTag reporter* amb la tècnica Invader per l'anàlisi d'SNPs**

En la tècnica Invader (ThirdWave Technologies Inc, Madison, WI, EUA), dos oligonucleòtids (les sondes Invader i al.llel específica), hibriden en tàndem a una cadena de DNA complementària i formen una estructura triple com s'indica en la Figura 7. En aquesta estructura, l'extrem 3' de la sonda Invader se solapa en un nucleòtid amb la sonda al.llel específica. Aquesta última té una cua de nucleòtids en l'extrem 5' (anomenada *flap*), la qual no és complementària a la mostra de DNA a analitzar. Les 5'endonucleases, conegudes amb el nom de Cleavases®, reconeixen aquesta estructura triple i tallen la sonda al.llel específica en el nucleòtid superposat (Lyamichev *et al*, 1999). Aquest tall allibera la *flap* a la solució i es detecta en una segona reacció Invader la qual és comuna per tots els SNPs (Invader Square). La *flap* hibrida amb un *cassette* de transferència d'energia per ressonància de fluorescència (FRET, *fluorescence resonance energy transfer*) i forma una estructura triple que serà reconeguda i tallada per la 5'endonucleasa. La digestió del *cassette* FRET separa el fluorocrom del *quencher* i genera un senyal fluorescent que és detectat en un lector de fluorescència (Figura 7.A).

Les Cleavases són enzims de la família de les 5' *flap* endonucleases (FEN), les quals tenen un paper clau en la replicació, recombinació i reparació del DNA (Lieber, 1997). L'especificitat d'aquesta tècnica està determinada per l'habilitat d'aquests enzims de discriminar entre substrats on les dues sondes se superposen de les que no. La substitució d'un únic nucleòtid en el DNA diana en el punt de superposició entre sondes impedeix la formació de l'estructura triple, la qual és necessària per l'activitat de l'enzim. Aquesta forta dependència entre l'activitat de l'enzim i la formació de l'estructura triple és la que permet que aquesta tècnica sigui molt adequada per la detecció d'SNPs (Ryan *et al*, 1999; Mein *et al*, 2000). De fet, Third Wave

Technologies disposa d'assajos de detecció d'SNPs d'aplicació clínica (<http://www.twt.com>).



**Figura 7.** Discriminació al·lelica per la tècnica Invader. A. Tècnica Invader Square. B. Combinació de la tècnica Invader amb els *eTag reporters*. a, DNA diana; b, sonda Invader; c, sonda al·lel específica; d, *flap*; e, *cassette* FRET (transferència d'energia per ressonància de fluorescència); \* punt de tall per l'enzim 5'nucleasa o Cleavase.

Per la identificació d'un SNP es requereix, doncs, del disseny d'una sonda Invader i dues al·lel específiques, una per cada al·lel. El nucleòtid de l'extrem 3' de la sonda Invader no cal que sigui complementari al DNA diana i no hi ha preferència per cap dels quatre nucleòtids (Lyamichev *et al*, 1999). Les sondes al·lel específiques es marcaran amb una *flap* diferent en el seu extrem 3' la qual hibrida de forma específica a un *cassette* FRET. L'amplificació del senyal i el tall de múltiples *flaps* per molècula de DNA, té lloc per incubació de la reacció a una temperatura propera a la  $T_m$  de les sondes al·lel específiques, les quals estan en excés. Sota aquestes condicions, la sonda tallada es dissocia ràpidament del DNA diana i és reemplaçada per una sonda intacta, iniciant-se un nou cicle. El nombre de sondes tallades per molècula diana i minut és aproximadament  $30 \text{ min}^{-1}$ , que resulta amb una amplificació del senyal de  $10^{e3}$  a  $10^{e4}$  en una reacció de 2 a 4 hores (Lyamichev *et al*, 1999; 2000). La sonda Invader es dissenya que tingui una  $T_m$  més elevada que les sondes al·lel específiques i es mantingui hibridada al DNA diana a la temperatura de la reacció. Així, doncs, sota unes condicions isoterms, múltiples sondes són tallades



per molècula de DNA i dóna lloc a una amplificació lineal de les *flap* respecte la concentració de DNA inicial.

El disseny Invader original només permet la detecció d'un SNP per reacció donat el nombre limitat de fluorocroms disponibles. Les principals avantatges d'aquesta tècnica són tres: *i.* Assaig homogeni; *ii.* Assaig quantitau (amplificació lineal); *iii.* No dependent de PCR (detecció directa en DNA genòmic).

No obstant, requereix d'una quantitat considerable de DNA genòmic (>100 ng per reacció). És per això que, molts dels grups que utilitzen aquesta tècnica amplifiquen les regions genòmiques que contenen els SNPs d'interès per PCR simple o multiplicadora (Mein *et al*, 2000; Ohnishi *et al*, 2001; Nevilie *et al*, 2002).

La combinació de la tècnica Invader amb els *eTag reporter* es basa en la substitució de la *flap* per un *eTag reporter* específic (Figura 7.B). Aquesta nova aproximació rep el nom *eTag Multiplex Invader SNP Assay* ja que permet l'anàlisi de múltiples SNPs en una mateixa reacció. El desenvolupament, optimització i validació d'aquest mètode ha estat realitzat per la doctoranda a l'empresa ACLARA BioSciences Inc (vegeu Estudi 3).

## Hipòtesi

Tenint en compte els següents antecedents:

1. Que per anàlisi de lligament s'ha trobat que la regió 6p21 podria tenir un o uns gen/s de susceptibilitat per l'esquizofrènia.
2. Que el gen *DDR1*, que es localitza a 6p21, codifica per un receptor tirosina quinasa que s'expressa de forma important en el cervell durant el neurodesenvolupament en cèl.lules glials de la substància blanca.
3. Que recentment s'està observant que a part de les neurones, les cèl.lules de la substància blanca, principalment oligodendròcits, podrien estar involucrats en l'esquizofrènia.
4. Que l'esquizofrènia es considera una malaltia d'origen en el neurodesenvolupament malgrat que les manifestacions clíniques apareixen a partir de l'adolescència.

Ens plantegem la següent hipòtesi:

El gen *DDR1* està implicat en el neurodesenvolupament i, en particular, en el funcionament de l'oligodendroglia i el procés de mielinització. La presència de variants genètiques que alteren la seva expressió i/o funció afecta l'estructuració i funcionament del SNC i, per tant, confereixen susceptibilitat a l'aparició de l'esquizofrènia.

El nostre grup ha proposat l'estudi de *DDR1* com a gen candidat de susceptibilitat per l'esquizofrènia dins del marc de la hipòtesi de la disfuncionalitat de l'oligodendroglia i del procés de mielinització. En primer lloc, ens hem proposat identificar variants en el gen *DDR1* i estudiar la seva possible associació amb la malaltia mitjançant un estudi de casos i controls. En segon lloc, ens hem proposat estudiar la seva expressió durant el neurodesenvolupament i en cervell adult (en ratolins i humans). Aquest segon estudi és treball de tesi de les doctorandes Neus Franco i Bàrbara Roig del grup de la Dra. Elisabet Vilella.

## Objectius

La present tesi doctoral té com a objectiu general l'estudi de variants genètiques en el gen *DDR1* en una mostra de malalts d'esquizofrènia i individus controls i estudiar la seva possible implicació en l'esquizofrènia. Els objectius concrets de l'estudi són:

- Objectiu 1. Identificar variants genètiques relativament freqüents (>10%) en els exons i regions flanquejants del gen *DDR1* en una mostra de malalts d'esquizofrènia mitjançant la tècnica de seqüenciació en *pools* (barreges de mostres) de DNA.

Objectiu 2. Valorar si un grup de polimorfismes identificats en el gen *DDR1* estan associats a l'esquizofrènia en una mostra de població espanyola. Estudi d'associació cas-control per comparació de freqüències haplotípiques.

Objectiu 3. Investigar la possible estratificació genètica en la mostra poblacional estudiada.

Objectiu 4. Desenvolupar un nou mètode per l'anàlisi simultània de polimorfismes d'un únic nucleòtid o SNPs fent ús de la tecnologia dels *eTag reporter* d'ACLARA BioSciences Inc.

### 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)
<b>Relació homes/dones</b>	176 (59.3%)/121 (40.7%)	53/47	158 (54.9%) /130 (45.1)
<b>Edat<sup>a</sup></b>			
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
<b>Edat d'inici<sup>b</sup></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	
<b>Subtipus esquizofrènia<sup>c</sup> (Negativa/Positiva)</b>			
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%)	
<b>Antecedent familiars<sup>d</sup></b>			
Tots	75 (25.2%)	31 (31%)	
Homes	40 (13.5%)	18 (34%)	
Dones	35 (11.8%)	13 (28%)	

<sup>a</sup> Edat en el moment en que es va reclutar al pacient, en concret del dia en què es va fer la extracció de sang; <sup>b</sup> Edat d'inici en la que apareixen els primers símptomes psicòtics (referit pel propi pacient, pels familiars o per l'entorn assistencial); <sup>c</sup> 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); <sup>d</sup> 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 K<sub>3</sub>*

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

<sup>a</sup> C, cadena interrogada: AS, *antisense*, S, *sense*; <sup>b</sup> Per les sondes al.llel específiques, en negreta s'indica la posició polimòrfica i en cursiva la del nucleòtid biotinitat; <sup>c</sup> 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 MgCl<sub>2</sub>, 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 MgCl<sub>2</sub>.

##### *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<sub>m</sub> 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 <sup>a</sup>	Seqüència (5'-3')	L <sup>b</sup>	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		

<sup>a</sup>F indica encebador *forward*, R encebador *reverse* i el número indica la posició del primer respecte la seqüència AC004211.1; <sup>b</sup>Longitud 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](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 MgCl<sub>2</sub>, 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 <sup>a</sup>	Seqüència (5'-3')	L <sup>b</sup>	Enzim de restricció <sup>c</sup>
DDR1-exó 3 Leu 94 CTG/CTC rs2229933 <sup>d</sup>	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

<sup>a</sup>El número indica la posició del primer respecte la seqüència AC004211.1; <sup>b</sup>Longitud en parelles de bases del fragment de PCR; <sup>c</sup>Indica la longitud dels fragments digerits per l'enzim corresponent per cadascun dels al.lels; <sup>d</sup>Número 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 <sup>a</sup>	Seqüència (5'-3') (forward i reverse)	L <sup>b</sup>	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 $\gamma$ 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 TACCTGAGTTCAGTTCCGTC	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 CCATGTCAAACCGGTAGCCTCAAGC	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 <sup>c</sup>
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 $\alpha$ Leu162Val	22q13.31	GACTCAAGCTGGTGTATGACAAGT CGTTGTGTGACATCCCAGAGAAT	117	Hinf I	Vohl <i>et a</i> , 2000

<sup>a</sup>C, localització cromosòmica; <sup>b</sup>Longitud en parelles de bases del fragment de PCR; <sup>c</sup>Número 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 <sup>a</sup>	Enzim restricció	Referència
<i>APOE</i> 112 <sup>b</sup> 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 <sup>c</sup>
DDAH1 Intrònic	TGCCCAAGTACACTCTTCATTG CCCTCTCAGGCAGTGTTCAT	226	BsrF 1	rs761601

<sup>a</sup>Longitud en parelles de bases del fragment de PCR; <sup>b</sup>Els 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; <sup>c</sup>Número 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 MgCl<sub>2</sub>, 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<sub>m</sub> 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 Cl<sub>2</sub>Mg (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 (10<sup>7</sup> 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  $\chi^2$ .* 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](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.

## **ESTUDI 1: Rastreig de mutacions en el gen *DDR1* en una mostra de malalts d'esquizofrènia: seqüenciació en *pools* de DNA**

### **INTRODUCCIÓ**

Les estratègies d'identificació de nous SNPs més emprades són tres: *i.* seqüenciació automàtica directa de fragments de DNA amplificats per PCR; *ii.* tècniques basades en la separació diferencial i detecció de cadenes de DNA que difereixen en un o més nucleòtids (per exemple, SSCP, dHPLC, DGGE, HDA); *iii.* identificació d'SNPs *in silico* (*is*SNPs) per comparació de seqüències (EST i genòmiques) dipositades en les bases de dades. Aquesta última és simple, eficaç i de baix cost. No obstant, els virtuals *is*SNPs identificats han de ser confirmats o bé per seqüenciació directa o bé per algun dels mètodes comentats en l'apartat de 'Mètodes de genotipatge de polimorfismes d'un únic nucleòtid' en múltiples individus. D'altra banda, les tècniques que es basen en la migració diferencial de les cadenes mutants són emprades com a tècniques de rastreig, però no identifiquen la mutació en concret. Són tècniques simples i relativament barates, però, sovint, no són capaces de detectar totes les mutacions.

La millora dels sistemes de detecció de fluorescència i programes d'anàlisi de seqüències dels aparells d'EC, així com l'automatització del procés de seqüenciació, ha fet que la seqüenciació automàtica per EC sigui el mètode més emprat per la identificació i verificació d'SNPs. No obstant, es tracta encara d'un mètode costós per la identificació sistemàtica de nous SNPs. Una estratègia per abaratir el cost i accelerar el procés és el rastreig en barreges o *pools* de mostres de DNA de diferents individus (Sham *et al*, 2002). S'ha demostrat que aquesta és una aproximació efectiva tant per la identificació de nous SNPs per seqüenciació directa de fragments de DNA amplificats per PCR (Kwok *et al*, 1994; Taillon-Miller *et al*, 1999; Blazej *et al*, 2003), com també pel genotipatge de SNPs i microsatèl·lits (Barcellos *et al*, 1997; Le Hellard *et al*, 2002; Blazej *et al*, 2003). El factor limitant d'aquesta aproximació és la capacitat de detectar una mutació quan aquesta està present en almenys un dels membres del *pool*. En aquest cas, la sensibilitat està determinada per la reacció de seqüenciació i pel sistema de detecció de l'aparell d'electroforesi emprat.

L'objectiu d'aquest treball ha estat la utilització de la seqüenciació automàtica per EC en *pools* de mostres de DNA com a mètode de rastreig de mutacions per tal d'abaratir el cost de la identificació d'SNPs en el gen *DDR1* en una mostra de 100 malalts d'esquizofrènia. El nostre interès s'ha centrat en la identificació de variants genètiques en les regions codificants i exó-intró del gen *DDR1*.

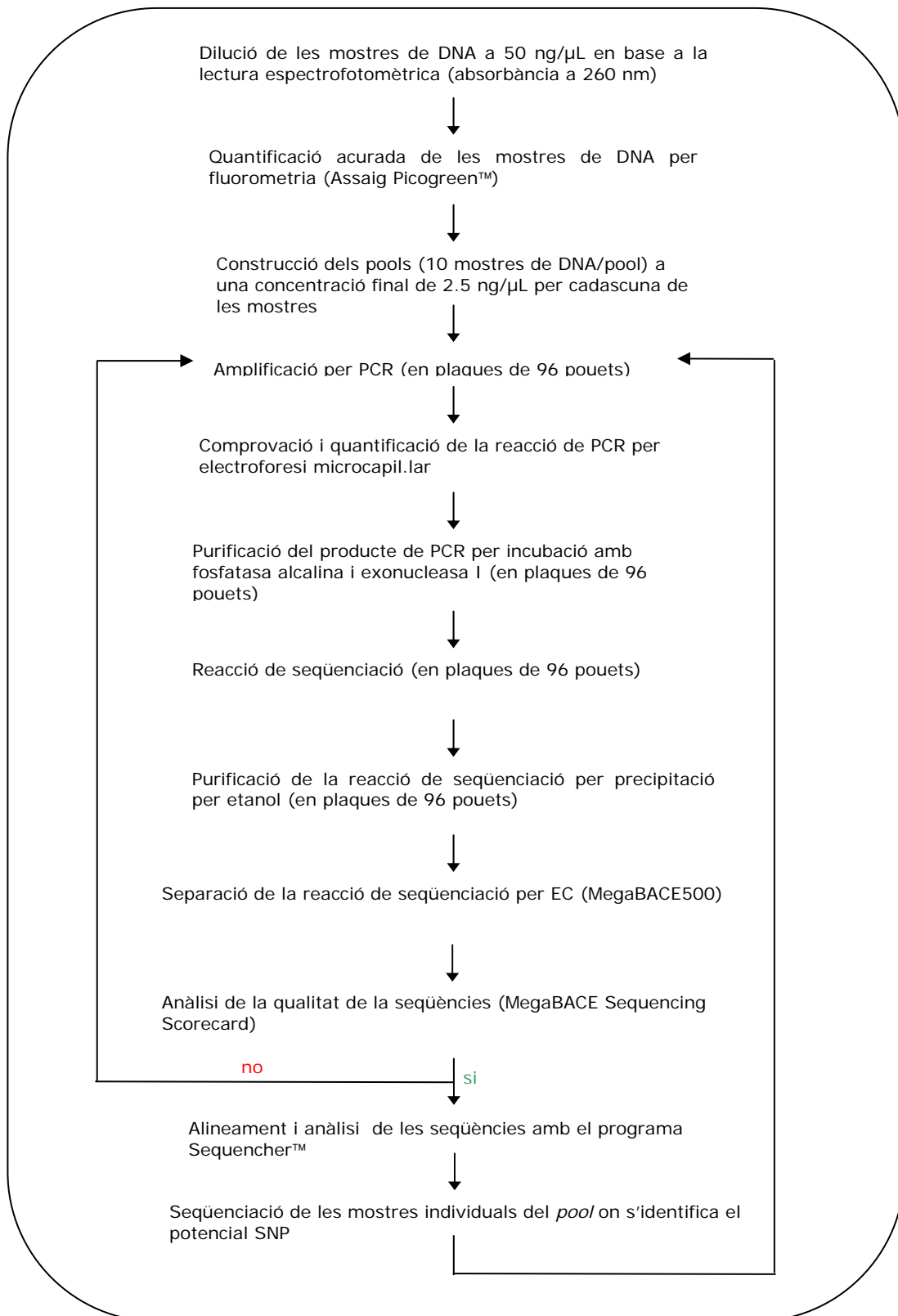
## RESULTATS

### Estudi preliminar: sensibilitat de la tècnica i estandarització dels protocols

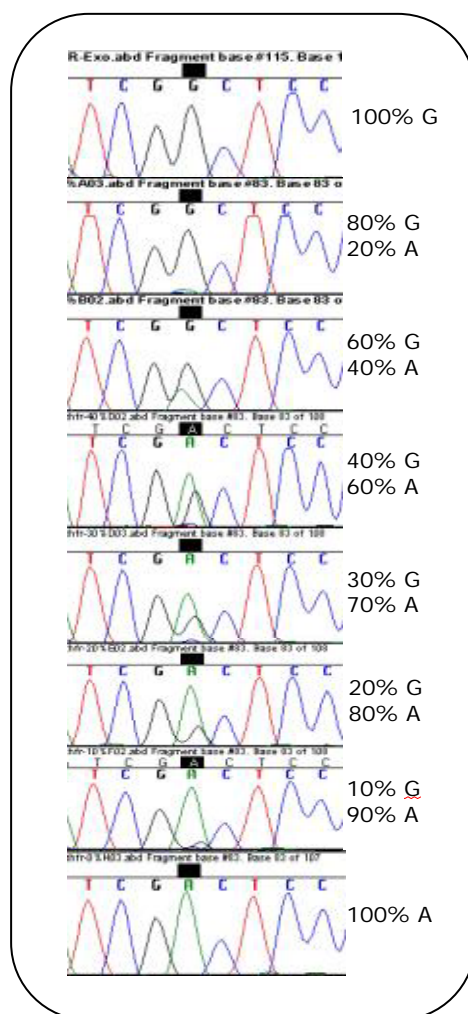
Amb l'objectiu d'estandaritzar els protocols de quantificació de DNA i seqüenciació automàtica i determinar el límit de detecció de la tècnica fent ús de l'aparell d'EC MegaBACE500 (APBiotech) disponible al laboratori, es va dur a terme un estudi preliminar. A la vegada, es volia minimitzar els errors de pipeteig i fer tot el procés amb el mínim nombre de plaques de 96 pouets. Per aquest motiu, vam considerar el mètode de purificació ExoSAP-IT (USB Corporation) pel producte de PCR i de precipitació per etanol per la reacció de seqüenciació. De manera que, en una única placa de 96 pouets es podia dur a terme la reacció de PCR, la purificació del producte de PCR, la reacció de seqüenciació, la purificació de la reacció de seqüenciació i la injecció de les mostres en l'aparell d'EC. El protocol establert per aquest estudi es resumeix en la Figura 8.

Per tal de determinar la sensibilitat de la tècnica de seqüenciació en barreges de DNA, es van construir *pools* amb diferent proporció dels al·lels del polimorfisme MTHFR C677T a partir de mostres prèviament genotipades segons Fross i col.laboradors. Vam observar que fent ús del sistema comercial de seqüenciació DYEnamic ET Dye Terminator Cycle Sequencing Kit (APBiotech) i el seqüenciador d'EC MegaBACE500 (APBiotech) i analitzant visualment els cromatogrames amb el programa Sequencher (GeneCodes Corporation), la sensibilitat de la tècnica era del ~10% per l'al·lel minoritari (Figura 9). Aquest és un resultat aproximat, ja que l'anàlisi dels cromatogrames es va fer visualment. A més, aquest programa no està optimitzat per la detecció de mutacions en *pools* de mostres de DNA. No obstant, el nostre objectiu era establir un mètode de rastreig qualitatiu i no pas quantitatiu considerant les eines de les quals disposàvem al laboratori.

Tenint en compte la sensibilitat observada de la tècnica i el nombre de reaccions de seqüenciació a realitzar (100 mostres x 16 fragments de PCR x 2 reaccions encebador *forward* x 2 reaccions encebador *reverse* = 6.400 reaccions de seqüenciació), es va decidir agrupar les mostres en grups o *pools* de 10 individus. De manera que, el nombre d'assaigs quedaria reduït en una desena part, sense considerar les reaccions de re-seqüenciació dels individus d'un dels *pools* on es detectés la potencial variant per tal de confirmar-la. Això volia dir que aquelles mutacions poc freqüents o rares passarien desapercibudes (freqüència <10%). No obstant, el nostre interès se centrava principalment en els polimorfismes o variants genètiques freqüents (informatives) per tal de dur a terme posteriorment un estudi d'associació de casos i controls.



**Figura 8.** Protocol establert per la detecció de mutacions per seqüenciació automàtica en *pools* de mostres de DNA.



**Figura 9.** Detecció del polimorfisme MTHFR C677T per seqüenciació automàtica en *pools* de mostres de DNA. Seqüència d'electroforogrames on s'indica la proporció per cada al.lel.

### Construcció dels *pools* de DNA

Per tal garantir l'amplificació paral.lela per PCR de cada mostra en el *pool*, calia posar la mateixa quantitat de DNA genòmic de cada mostra en el *pool*. D'altra banda, s'havia de comprovar que totes les mostres podien ser amplifiades individualment per PCR prèviament a la construcció dels *pools*.

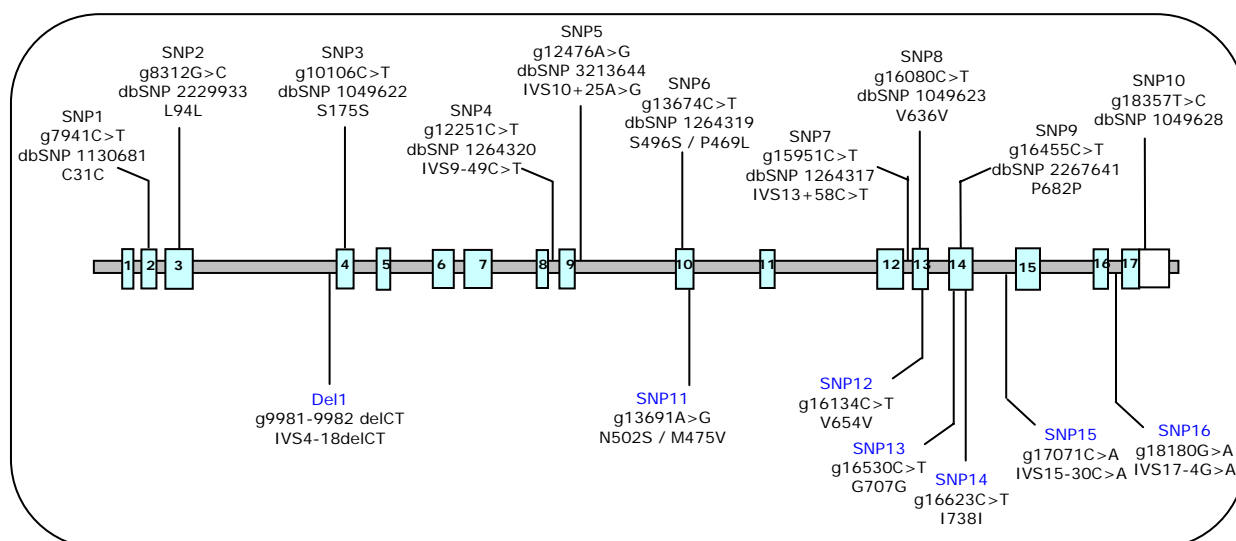
La concentració del DNA genòmic es va mesurar per fluorimetria fent ús del tenyiment per PicoGreen. Aquest és el mètode de quantificació de DNA recomanat per la construcció de *pools* de DNA (Barcellos *et al*, 1997). D'acord amb la mesura fluorimètrica, es van preparar els 10 *pools* afegint la mateixa quantitat de DNA (600 ng) de 10 mostres diferents. La concentració final de cadascuna de les mostres en el *pool* va ser de 2.5 ng/ $\mu$ L.



Per últim, mencionar que les mostres havien estat prèviament amplificades per PCR de forma satisfactòria en d'altres estudis (Virgos *et al*, 1999; Virgos *et al*, 2001).

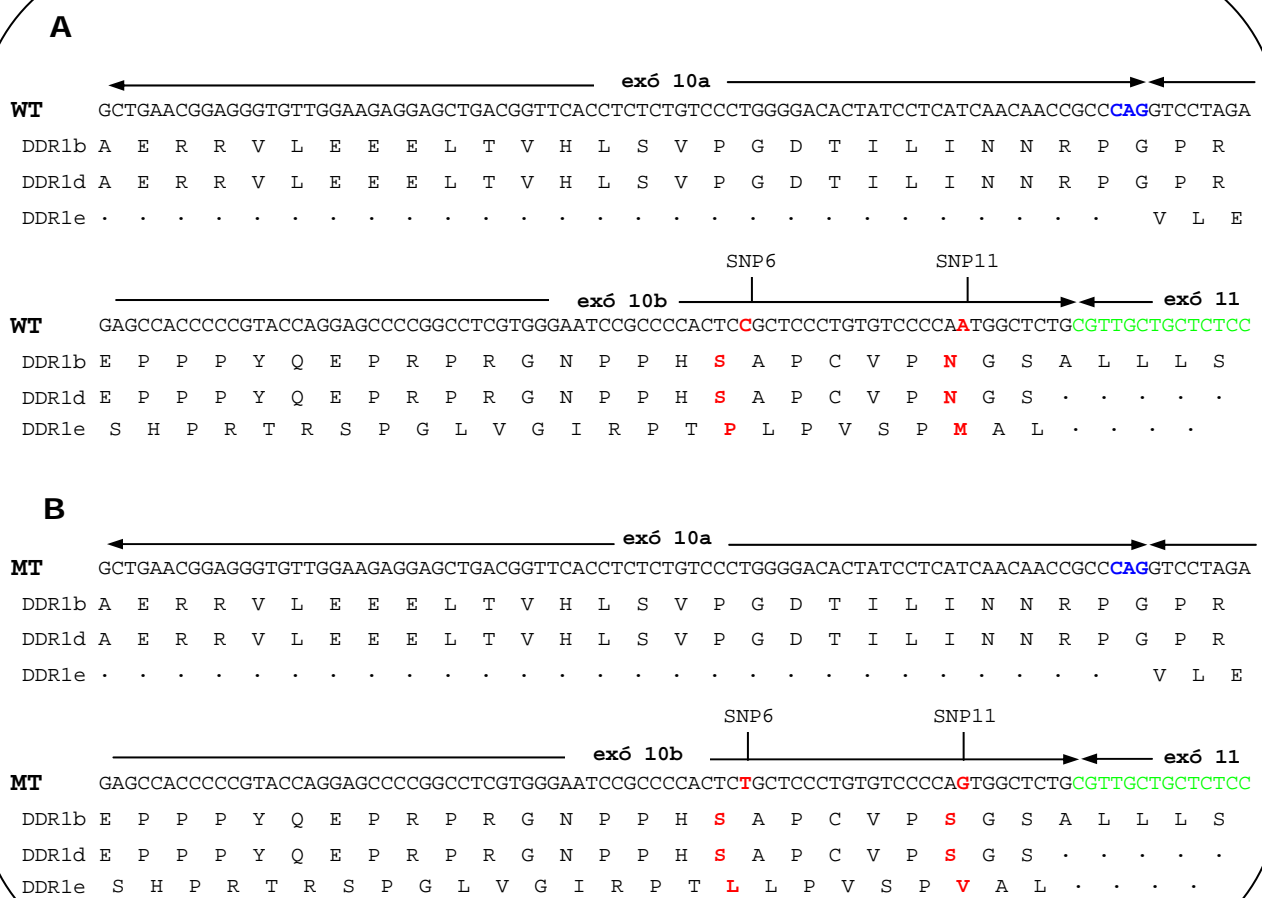
### Identificació de variants genètiques en les regions codificant i exó-intró del gen *DDR1* per seqüenciació en *pools* de mostres de DNA

Seguint el protocol establert en l'estudi preliminar, vam detectar un total de 17 variants genètiques, 16 SNPs i una deleció de 2 pb en l'intró 4, en una mostra de 100 malalts d'esquizofrènia no emparentats (Figura 10). Les substitucions nucleotídiques més comunes van ser les transicions (14 de 16, 87.5%), en comparació amb les transversions (2 de 16, 12.5%).



**Figura 10.** Estructura del gen *DDR1* amb les variants identificades per seqüenciació en *pools* de mostres de DNA. Exons representats pels blocs verds verticals i enumerats de l'1 al 17. El bloc blanc correspon a la regió 3'UTR. La posició de les variants corresponen a la seqüència OTTHUMG00000016356 de la base de dades Vega. Les variants marcades amb blau indiquen que són noves o no han estat dipositades a la dbSNP, per calcular la posició dels aa s'ha emprat la isoforma *DDR1b*. En negre s'indiquen aquells SNPs que es troben dipositats a la dbSNP.

Deu SNPs es localitzaren en regions codificants. D'aquests, 8 són silenciosos i 2 resulten en un canvi d'aminoàcid. L'SNP6 (g13674C>T), transició C→T a 140 pb del començament de l'exó 10, resulta en la substitució d'una prolina per leucina en la posició 469 de la isoforma e, però es tracta d'un SNP silencios en la resta d'isoformes (Figura 11). L'SNP11 (g13691A>G), transició A→G a 157 pb de l'exó 10, resulta en la substitució d'una asparagina per una serina en la posició 502 de les isoformes a, b, c i d i d'una metionina per una valina en la posició 475 de la isoforma e (Figura 11). Les seqüències humanes, de ratolí i rata dipositades al NCBI suggereixen la conservació de l'aminoàcid asparagina en aquesta posició. Ara bé, els models computacionals de predicció de l'estructura secundària (Combet *et al*, 2000) no apunten a cap alteració rellevant causada per aquesta mutació.

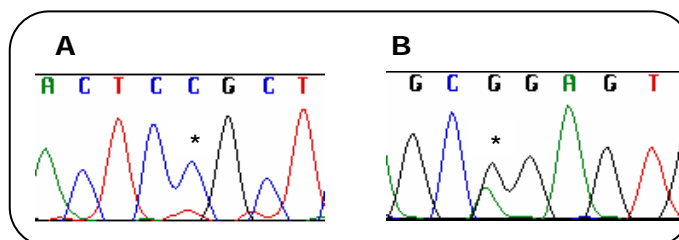


**Figura 11. Mutacions de canvi de sentit identificades en l'exó 10 del gen *DDR1*.** El cDNA de *DDR1* corresponent a l'exó 10 s'indica com a WT o MT, corresponent a la seqüència salvatge (A) o mutant (B). Els productes proteics s'indiquen per les isoformes b, d i e. En blau s'indica el punt críptic acceptor d'*splicing* alternatiu en l'exó 10 que dona lloc a la isoforma e (Alves *et al*, 2001).

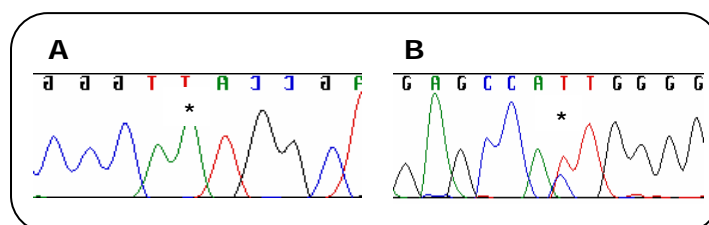
Cinc SNPs i la deleció IVS4-18delCT són intrònics. Mitjançant la utilització de models de predicció de llocs de *splicing* en humans, cap dels anteriors SNPs afectarien o introduirien nous llocs de *splicing* (NetGene2 Server, Brunak *et al*, 1991; Splice Site Prediction by Neural Network, Reese *et al*, 1997). Per últim, en la regió 3'UTR s'ha identificat un SNP a 33 pb de l'exó 17.

De les 17 variants, 16 es van detectar en almenys un dels *pools* a excepció de l'SNP11 en l'exó 10. En aquest cas particular, es van seqüenciar individualment les mostres del *pool* 6, ja que en aquest *pool* s'observava la presència d'un potencial SNP (SNP6) només en les reaccions de seqüenciació corresponents a l'encebador Ex10 R29358DDR1. En seqüenciar individualment les mostres d'aquest *pool*, es va confirmar la presència de l'SNP6 (dbSNP rs1264319) en una de les mostres (Figura 12). Casualment, també es va detectar en una única mostra una variant addicional

(SNP11: g13691A>G, N502S) la qual havia passat completament desapercibuda en la seqüenciació del *pool* (Figura 13). Aquestes dues variants, SNP6 i SNP11, van ser solament detectades en el *pool* 6.



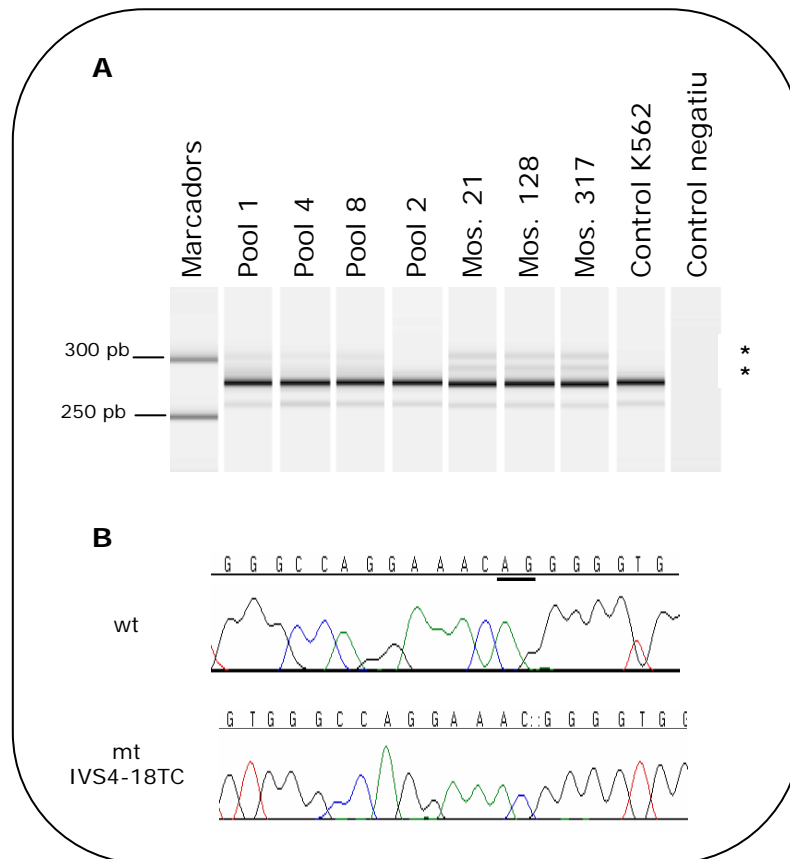
**Figura 12.** Electroforogrames parcials de l'exó 10 corresponents al *pool* 6 (A) i mostra número 279 (B). L'asterisc indica la posició de l'SNP6 (dbSNP rs1264319).



**Figura 13.** Electroforogrames parcials de l'exó 10 corresponents al *pool* 6 (A) i mostra número 289 (B). L'asterisc indica la posició de l'SNP11 (g13691A>G; N502S).

La presència d'un parell de bandes de major pes molecular en l'electroforesi dels productes de PCR del fragment Ex4 en tres *pools*, suggeria la presència d'una deleció o inserció d'unes poques parelles de bases. Les mostres dels tres *pools* es van amplificar per PCR individualment i es van analitzar per electroforesi per tal d'identificar aquelles mostres amb un patró electroforètic anòmal (Figura 14.A.). Es van identificar tres mostres, cadascuna inclosa en un *pool* diferent. Per tal de determinar la variant genètica, es van clonar els fragments Ex4 d'aquestes tres mostres en un vector plasmidi. D'aquesta manera vam confirmar la presència en heterozigosi de la deleció IVS4-18delCT (Figura 14.B.).

En la regió del gen *DDR1* hi ha un total de 117 SNPs dipositats a la base de dades dbSNP. D'aquests, 40 es localitzen en les regions analitzades en el present treball. En la mostra estudiada, només se n'han identificat 10 d'aquests (Figura 10). Les 7 variants genètiques restants identificades són noves o, si més no, no han estat reportades a la dbSNP.



**Figura 14.** Detecció de la mutació IVS4-18delCT. (A) Separació electroforètica dels productes de PCR Ex4 per electroforesi microcapil·lar (Agilent 2100 BioAnalyzer). Els asteriscs indiquen la presència de dues bandes de major pes molecular que el fragment Ex4 en aquells *pools* (*pools* 1, 4 i 8) o mostres (mostres 21, 128, 317) amb la mutació IVS4-18delCT. (B) El producte de PCR de les mostres 21, 128 i 317 es va clonar en un vector pGEM<sup>®</sup>-T Easy (Promega, Madison, WI) i 6 colònies es van seqüenciar fent ús de l'encebador M13 *forward*. La mateixa mutació es va detectar en les tres mostres. Els electroforogrames parcials del fragment Ex4 corresponents a l'al·lel salvatge (wt) i mutant (mt) de la mostra 128 es mostren com a exemple. La deleció CT (o GA en la cadena complementària) està subratllada.

## DISCUSSIÓ

El present estudi va explorar la tècnica de seqüenciació automàtica per EC en *pools* de 10 mostres de DNA com a mètode de rastreig de mutacions per tal d'abaratir el cost i accelerar el procés d'identificació d'SNPs en el gen *DDR1* (regions codificants i exó-intró) en una mostra de 100 malalts d'esquizofrènia. Un total de 16 SNPs i una deleció intrònica de 2 pb van ser identificats en almenys un dels *pools*, a excepció de l'SNP11 el qual va ser detectat per seqüenciació individual del fragment Ex10 en una de les mostres del *pool* 6 (Figura 10).

Considerant les eines emprades, els resultats presentats demostren que aquesta aproximació és vàlida per la identificació d'SNPs amb una freqüència major al 10% per l'al·lel minoritari. Aquests resultats són similars als obtinguts per d'altres autors (Wolford *et al*, 2000). Donada la sensibilitat de la tècnica i la mida del *pool* emprat, aquest mètode resulta solament eficaç per la detecció de polimorfismes comuns o freqüents. Per la detecció de mutacions rares, caldria fer barreges de 5 o menys mostres de DNA. L'anàlisi realitzat ha estat qualitatiu ja que el software emprat per l'anàlisi de cromatogrames (Sequencher) no està dissenyat per la detecció de mutacions en *pools* de mostres. Una anàlisi quantitativa dels resultats fora possible fent ús, per exemple, del programa PHRED, el qual permet d'avaluar les àrees o alçades dels pics electroforètics (Ewing *et al*, 1998).

Les 100 mostres d'estudi es van agrupar en *pools* o barreges de 10, reduint el nombre de reaccions inicials a realitzar per 10. De manera que, per la detecció d'una variant determinada es requeria que aquesta fos present en el *pool*, com a mínim, en heterozigosi en dues mostres o en homozigosi en una. Així doncs, 14 dels SNPs identificats es trobaven amb una freqüència major al 10% en les mostres estudiades. La detecció de l'SNP6 (g13674C>T) se'ns fa difícil d'explicar ja que només es va observar en el *pool* 6 i en les dues reaccions de seqüenciació on es feia ús de l'encebador *revers*. Inicialment es va pensar que es tractava de soroll de fons, però ens va sorprendre que aquest s'observava en els dos cromatogrames duplicats. A més, aquesta mutació estava descrita a la dbSNP (rs1264319) i, per això, es va voler comprovar la seva presència seqüenciant individualment les 10 mostres del *pool* 6. Com era d'esperar, es va confirmar la presència d'aquest SNP en heterozigosi en una única mostra. Addicionalment, es va detectar en una altra mostra una nova mutació (SNP11, g13691A>G), també en heterozigosi. Curiosament, aquestes són les dues úniques mutacions identificades que comporten un canvi d'aminoàcid (Figura 11).

Malgrat les variants identificades en el domini discoidina són sinònimes, és interessant el fet que dues d'aquestes, l'SNP1 C31C i SNP3 S175S, es localitzin en posicions crítiques per l'activitat de DDR1. En l'estudi recent de mutagènesi dirigida d'Abdulhussein i col.laboradors, els autors demostren que l'aminoàcid serina en la posició 175 té un paper clau en la unió a col·lagen. La seva mutació per una alanina anul·la la unió del domini discoidina a col·lagen. A més, mitjançant models de predicció de l'estructura tridimensional del domini discoidina, demostren que els

residus cisteïna 31 i cisteïna 185, dels extrems N i C terminal del domini discoidina, formen un pont disulfur (Abdulhussein *et al*, 2004).

L'SNP6 (g13674C>T) es va identificar en una pacient diagnosticada amb esquizofrènia paranoide i edat d'inici als 24 anys. Aquesta mutació resulta en la substitució d'una prolina per leucina en la posició 469 de la isoforma e, però es tracta d'un SNP silenciós en la resta d'isoformes. La isoforma e ha estat identificada en línies cel·lulars tumorals (Alves *et al*, 2001) i se'n desconeix la seva funció.

L'SNP11 (g13691A>G) es va identificar en una pacient diagnosticada amb trastorn esquizofreniforme i edat d'inici als 18 anys. Aquesta mutació resulta en la substitució d'una asparagina per una serina en la posició 502 de les isoformes a, b, c i d i d'una metionina per una valina en la posició 475 de la isoforma e. Les seqüències humanes, de ratolí i rata dipositades a la base de dades Entrez Nucleotides Database del NCBI, suggereixen la conservació de l'aminoàcid asparagina en aquesta posició. La repercussió d'aquesta mutació és en aquest moment especulativa. El nostre grup s'ha plantejat de realitzar l'estudi bioinformàtic i funcional per tal de determinar la implicació d'aquesta variant. Les freqüències observades per l'al·lel mutant S502 en les mostres de malalts d'esquizofrènia i individus controls estudiades fou baixa (0.02) (vegeu Estudi 2).

La comprovació dels productes de PCR per electroforesi prèviament a la reacció de seqüenciació ens va permetre d'identificar la deleció IVS4-18delCT en tres *pools*, ja que no era detectada per seqüenciació. Aquesta deleció es presentava en heterozigosi en tres mostres. Tots tres casos eren homes amb diagnòstics d'esquizofrènia catatònica (edat d'inici als 21 anys), paranoide (edat d'inici als 19 anys) i residual (edat d'inici als 20 anys). Com s'ha comentat anteriorment, no sembla que aquesta i la resta de variants intròniques identificades en aquest estudi afectin els mecanismes de *splicing*. A la dbSNP, hi ha una única mutació reportada que afectaria el mecanisme de *splicing* (rs2855546). Aquesta comporta un canvi de la seqüència consens del lloc acceptor AG per GG de l'intró 5. Es tracta d'un *is*SNP el qual no ha estat validat o estudiat en mostres poblacionals. En aquest treball tots els cromatogrames foren homozigots per la seqüència salvatge AG. Ara bé, no es pot descartar la presència d'aquesta variant en una freqüència molt baixa o en mostres amb bagatge genètic diferent.

D'acord amb diferents treballs, s'estima que entre un 6-12% (Reich *et al*, 2003) o 28-35% (Carlson *et al*, 2003) dels SNPs dipositats a les bases de dades serien falsos SNPs. En concret, en la dbSNP, hi ha dipositats un total de 117 SNPs en la regió del *DDR1*, dels quals 40 es troben en les regions analitzades en aquest treball. D'aquests, només vam detectar-ne 10 (25%). Això suggereix que o bé la resta d'SNPs són poc freqüents (freqüència menor al 10% per l'al·lel minoritari) i no han estat detectats donada la sensibilitat de la tècnica emprada o bé no estan presents en les mostres estudiades. D'altra banda, cal recordar que molts d'aquests SNPs han

estat identificats *in silico* i no han estat validats per genotipatge en múltiples mostres. Així és que està per determinar si es tracta de veritables SNPs. Curiosament, cap de les 3 variants descrites en la dbSNP que comporten un canvi d'aminoàcid (Leu94Val, Leu839Val, Arg873Trp) han estat identificades en les mostres estudiades.

El protocol desenvolupat en aquest treball permet de dur a terme tot el procés d'identificació de noves mutacions (des de la PCR inicial fins la injecció en l'aparell de EC) en única placa de 96 pouets, sense la necessitat de transferir les mostres i reaccions, i evitar així possibles errors de pipeteig. Aquesta simplificació del protocol, conjuntament amb l'anàlisi dels *pools* o mescles de DNA, va suposar una reducció en cost i temps considerable. Si l'interès rau en la identificació de polimorfismes freqüents, aquesta pot ser una aproximació vàlida com a mètode de rastreig inicial.

Universitat ROVIRA I VIRGILI  
Facultat de Medicina  
Departament de Ciències Mèdiques Bàsiques

**Anàlisi mutacional i estudi  
d'associació del gen receptor  
domini discoidina 1 (*DDR1*)  
en l'esquizofrènia**



Carmen Virgos Matilla  
Tesi doctoral  
Reus, 2004