

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

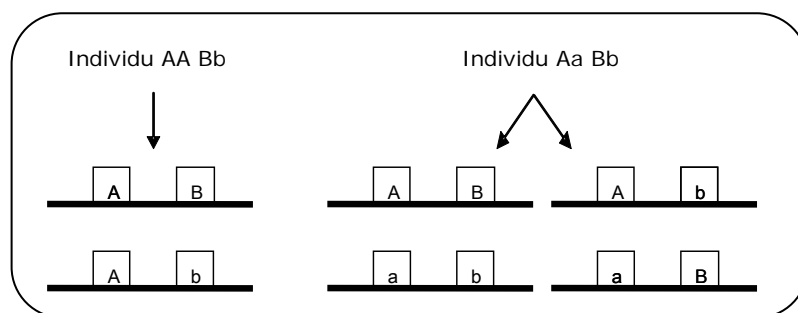
### INTRODUCCIÓ

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

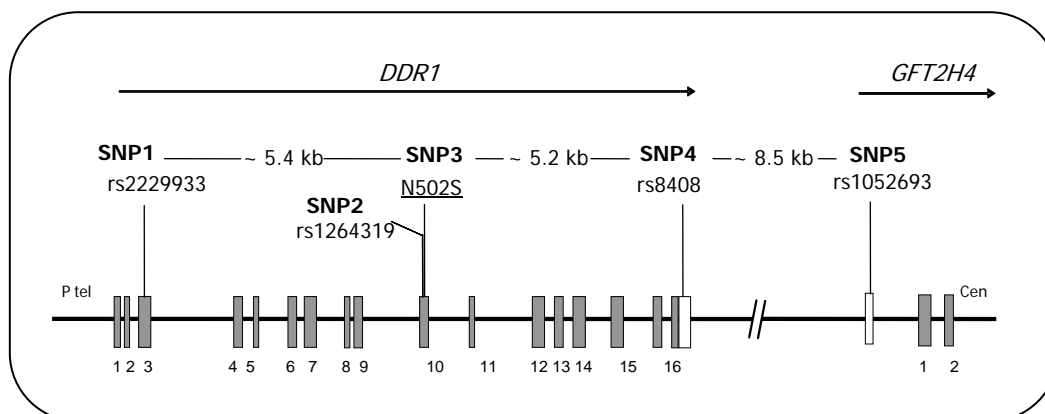
Per l'estudi d'associació, es van analitzar els dos SNPs identificats prèviament en l'Estudi 1 que comportaven un canvi d'aminoàcid. Aquests foren la variant SNP11 g13691A>G que resulta en el canvi d'una asparagina per una serina en la posició 502 i la variant SNP6 g13674C>T que resulta en el canvi d'una 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).

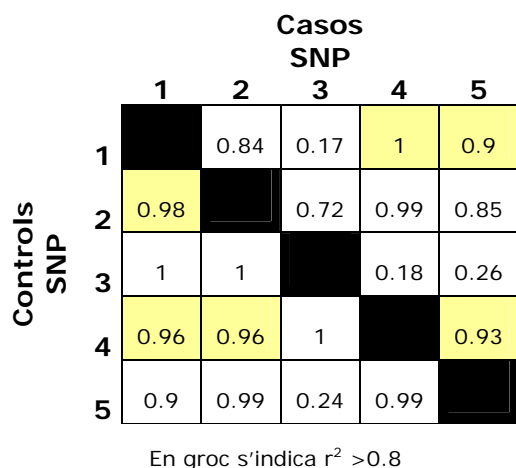
**Taula XVIII.** Distribució de les freqüències dels haplotips estimats mitjançant els algorismes E-M i Phase implementats en els programes Arlequin i Phase, respectivament.

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.

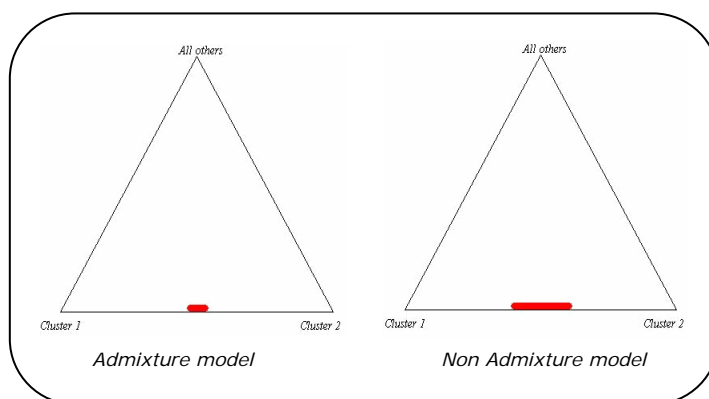


**Figura 17.** Sumari del test de desequilibri de lligament entre les parelles de SNPs.

### 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.