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: 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*.

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

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 sequenciació 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 desapercebudes (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.

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

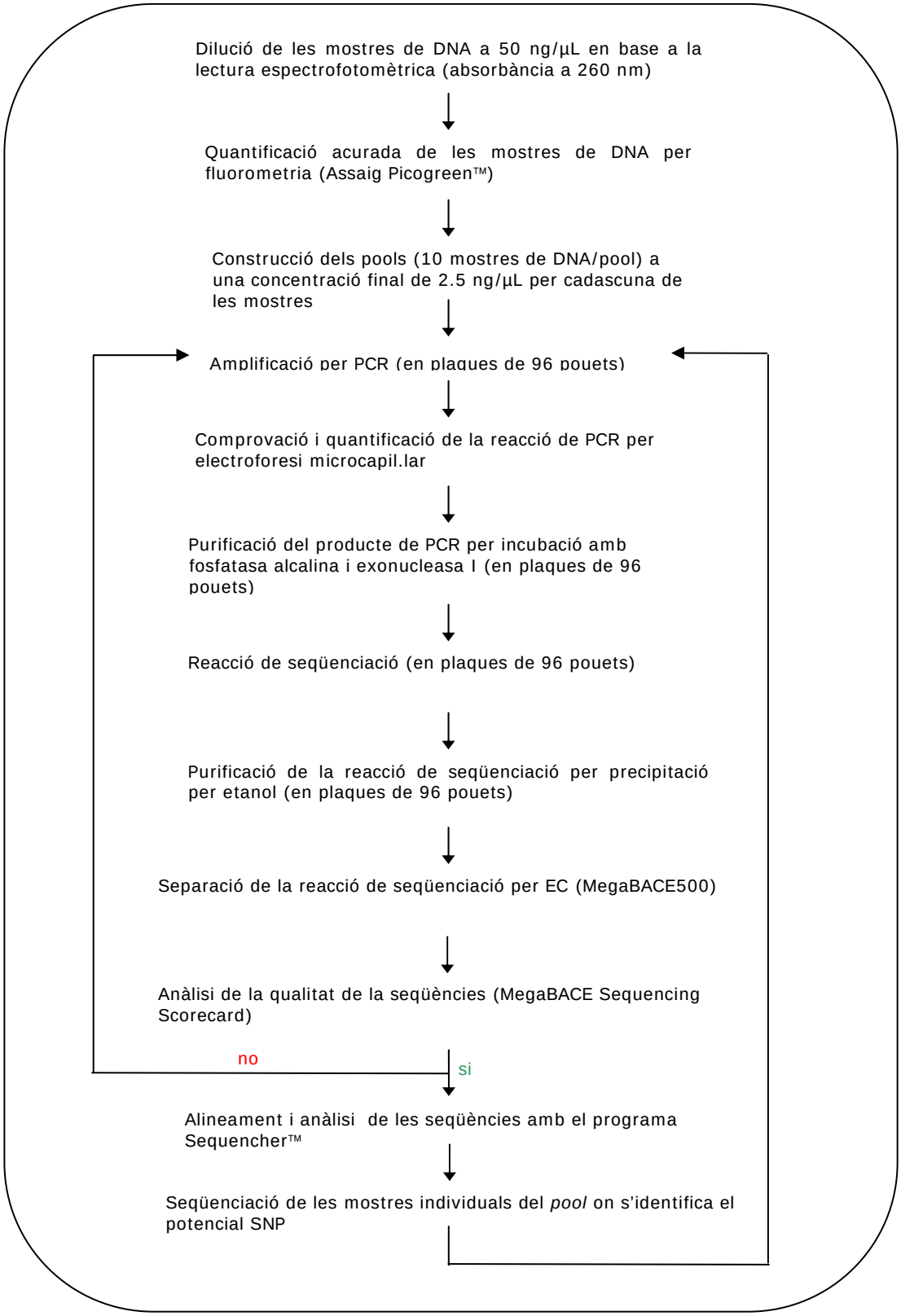

Figura 8. Protocol establert per la detecció de mutacions per seqüenciació automàtica en *pools* de mostres de DNA.

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

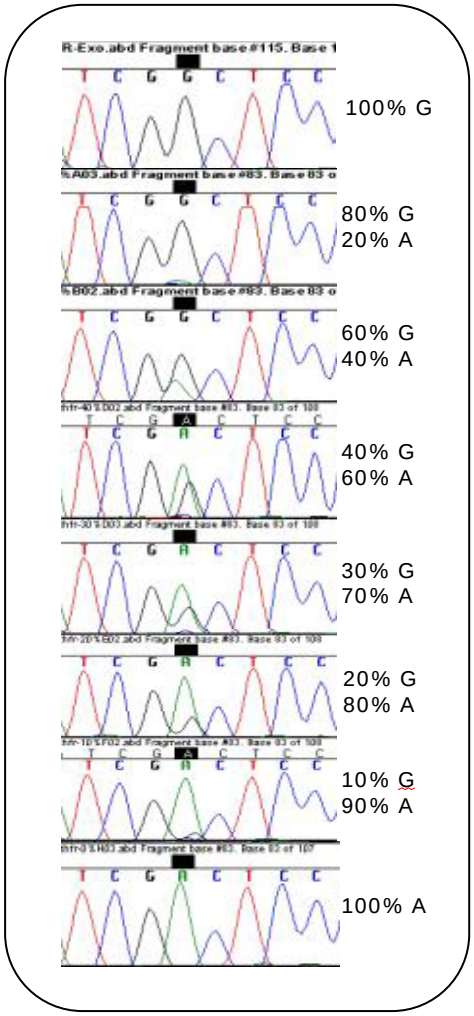

Figura 9. Detecció del polimorfisme MTHFR C677T per seqüenciació automàtica en *pools* de mostres de DNA. Seqüència d'electroforogrames on s'indica la proporció per cada al.lel.

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 amplificades individualment per PCR prèviament a la construcció dels *pools*.

La concentració del DNA genòmic es va mesurar per fluorometria 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/µL.

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

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%).

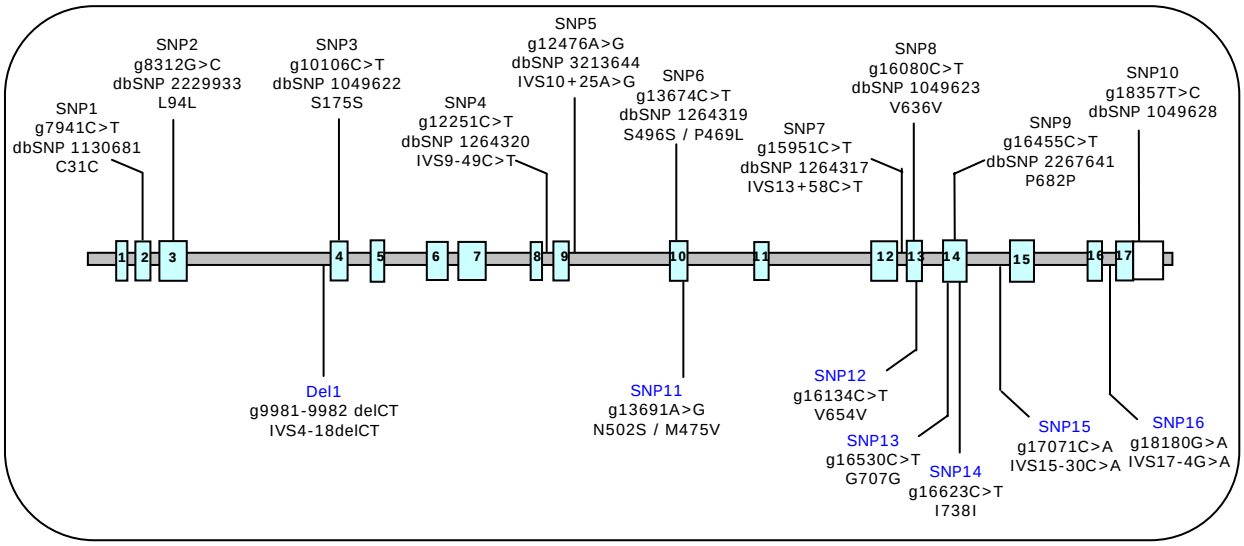

Figura 10. Estructura del gen *DDR1* amb les variants identificades per seqüenciació en *pools* de mostres de DNA. Exons representats pels blocs verds verticals i enumerats de l'1 al 17. El bloc blanc correspon a la regió 3'UTR. La posició de les variants corresponen a la seqüència OTTHUMG00000016356 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 silenciós 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 seqüencies 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.

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

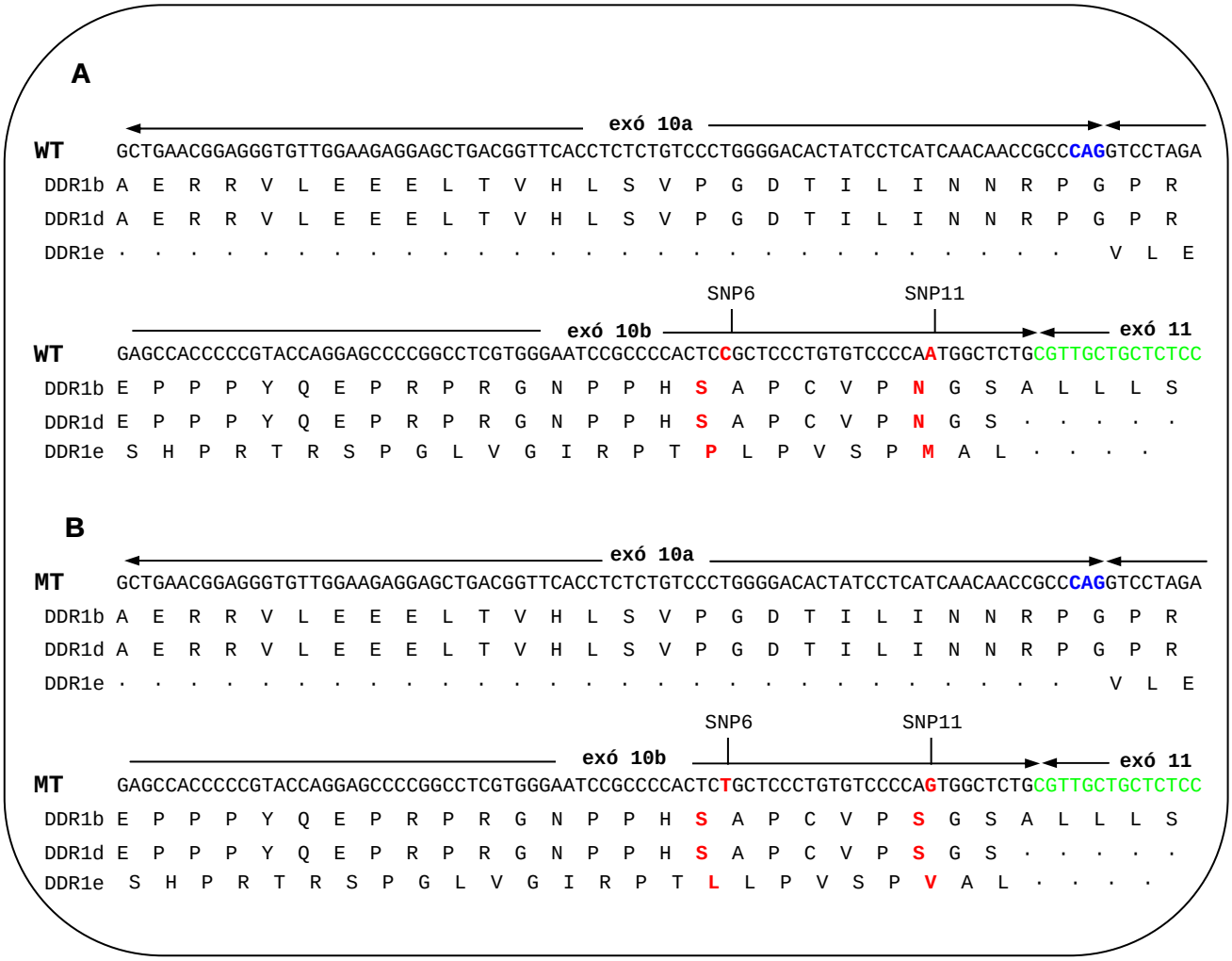

Figura 11. Mutacions de canvi de sentit identificades en l'exó 10 del gen *DDR1*. El cDNA de DDR1 corresponent a l'exó 10 s'indica com a WT o MT, corresponent a la seqüència salvatge (A) o mutant (B). Els productes proteics s'indiquen per les isoformes b, d i e. En blau s'indica el punt críptic acceptor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

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

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

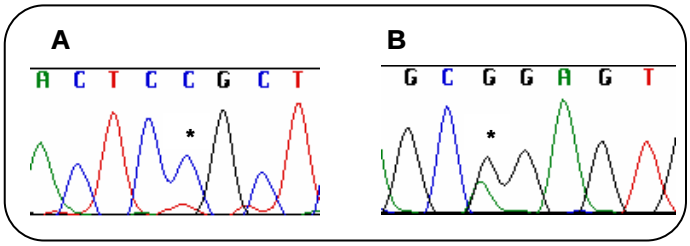

Figura 12. Electroforogrames parcials de l'exó 10 corresponents al *pool* 6 (A) i mostra número 279 (B). L'asterisc indica la posició de l'SNP6 (dbSNP rs1264319).

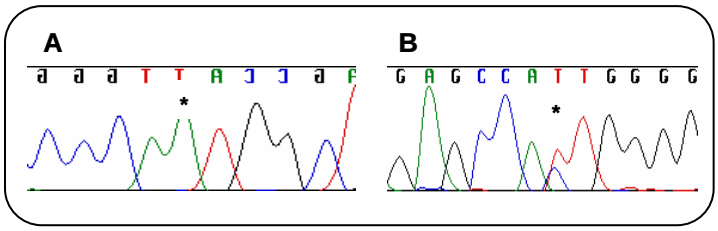

Figura 13. Electroforogrames parcials de l'exó 10 corresponents al *pool* 6 (A) i mostra número 289 (B). L'asterisc indica la posició de l'SNP11 (g13691A>G; N502S).

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.

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

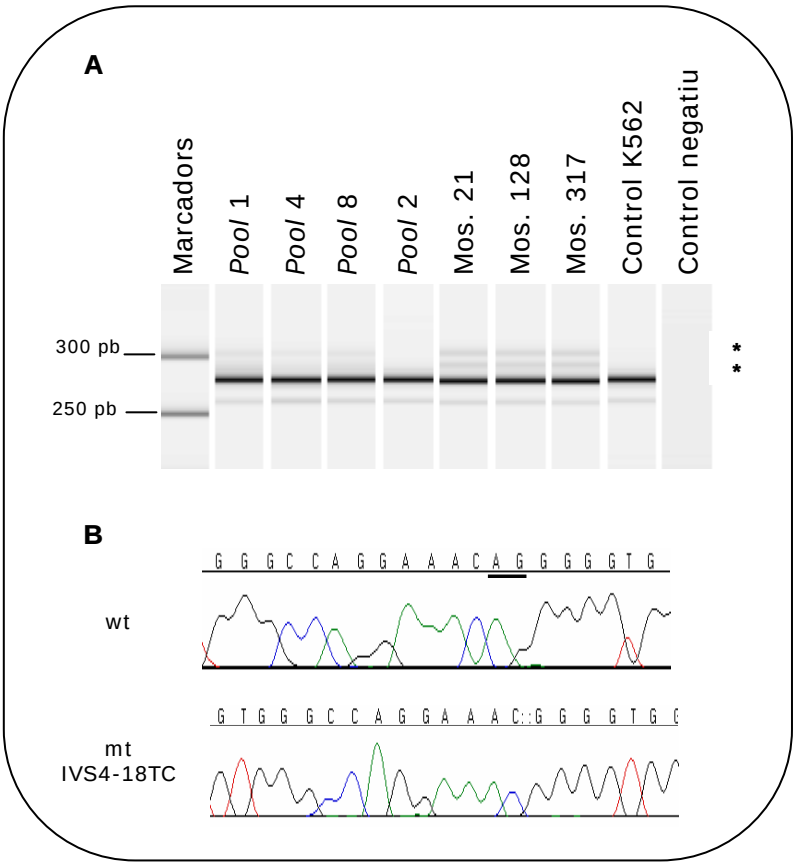

Figura 14. Detecció de la mutació IVS4-18delCT. (A) Separació electroforètica dels productes de PCR Ex4 per electroforesi microcapil.lar (Agilent 2100 BioAnalyzer). Els asteriscs indiquen la presència de dues bandes de major pes molecular que el fragment Ex4 en aquells *pools* (*pools* 1, 4 i 8) o mostres (mostres 21, 128, 317) amb la mutació IVS4-18delTC. (B) El producte de PCR de les mostres 21, 128 i 317 es va clonar en un vector pGEM®-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.

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

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 trobarien 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

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

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 seqüencies 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 p*ools*, 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

ISBN: 978-84-690-7776-4 / DL: T.1209-2007

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.