

Discussió

Science... never solves a problem without creating ten more.

George Bernard Shaw

L'objectiu central d'aquest treball ha estat la cerca d'un nou gen de retinitis pigmentària autosòmica recessiva en el locus RP26. La retinitis pigmentària és una malaltia molt heterogènia genèticament, de la qual han estat identificats més d'una trentena de gens i loci causals, des de començaments de la dècada dels noranta fins a l'actualitat. El locus RP26, un dels darrers locus en ser caracteritzat –entre mitjans i finals dels anys noranta– va ser identificat al nostre grup a partir d'una anàlisi de lligament realitzada en una única família de 15 fills, cinc dels quals amb la malaltia (Bayés et al. 1998). Aquest tipus d'aproximació a la cerca del gen causal d'una malaltia hereditària, que implica una anàlisi de lligament genètic, requereix l'escrutini d'un nombre mínim de meiosis per tal d'assolir significació estadística. Generalment, la significació estadística s'assoleix més fàcilment en malalties mendelianes d'herència dominant, perquè la presència d'individus afectats en cada generació incrementa considerablement el nombre de meiosis que podem analitzar. En canvi, en malalties d'herència recessiva, la major part dels individus afectats es concentren en una sola generació i, únicament si la família ha tingut molta descendència, o si disposem de diverses famílies que comparteixen una mateixa alteració, podrem analitzar un nombre elevat d'individus afectats que ens permeti aplicar el lligament genètic per tal de localitzar el gen de la malaltia. En la forma recessiva de la retinitis pigmentària, la no disponibilitat de famílies grans en dificulta molt l'estudi, i és una de les raons principals per la qual encara resten 4 loci orfes (VEGEU APARTAT 4.1 DE LA INTRODUCCIÓ). Tot i aquestes dificultats, inherents a l'estudi de la retinitis pigmentària autosòmica recessiva, al nostre grup, al llarg dels darrers quinze anys, han estat localitzats dos nous loci: el locus RP26, al cromosoma 2, i el locus RP19, al cromosoma 1. En aquest darrer, es va identificar el gen *ABCA4* –fins llavors responsable únicament de la malaltia d'Stargardt– com a gen causal de retinitis pigmentària (Martínez-Mir et al. 1997a, Martínez-Mir et al. 1998); descoberta que va obrir la porta a la implicació d'aquest gen en un ventall creixent de distròfies retinals, tant centrals com perifèriques.

Cerca “pregenòmica” del gen causal d'RP en el locus RP26

En el cas del locus RP26, tant les grans dimensions de la regió definida per cosegregació com l'absència d'un candidat clar, no van permetre identificar d'entrada el gen causal. La mida original del locus –que actualment correspon a 17.4 Mb del mapa físic– no va poder ser reduïda perquè no disposàvem d'altres famílies en què la malaltia cosegregués amb marcadors a 2q31-q33 i que fossin útils per a localitzar noves recombinacions i estrènyer els marges del locus. En aquest punt, una opció hagués estat la construcció d'un mapa físic detallat del locus a partir de la seqüenciació de clons genòmics de la regió per a localitzar els gens i situar els marcadors genètics sobre els fragments. La mida del locus era, també en aquest cas, un impediment major per a acometre aquest tipus d'estratègia, ja que el mapatge i la seqüenciació posterior de 17 milions de parells de bases de seqüència genòmica únicament es podia realitzar amb la infraestructura d'un centre de seqüenciació com els que en aquells moments s'estaven establint per a la seqüenciació completa del genoma humà. La construcció del mapa físic d'una regió cromosòmica un cop definit un nou locus d'una

malaltia hereditària –que equivalia a un projecte genoma a petita escala– es podia dur a terme en un grup de recerca petit si les dimensions del locus eren inferiors a dues megabases, però aquest no era, ni molt menys, el cas del locus RP26. Així, per tal d’obtenir informació posicional addicional del locus ens havíem d’esperar que comencessin a incorporar-se a les bases de dades seqüències de la regió generades pel Projecte Genoma Humà en curs.

Els projectes de seqüenciació d’EST: una eina per a la cerca dels gens de malalties?

El projecte de seqüenciació del genoma va néixer envoltat per la discussió sobre si era necessari seqüenciar totalment el complement haploide dels humans, tenint en compte que únicament una petita fracció de tota la seqüència genòmica codifica proteïnes. Els detractors de la seqüenciació completa argumentaven que el projecte genoma –motivat d’entrada per l’horitzó de l’estudi de les malalties humanes i, per tant, encaminat a la localització de tots i cadascun dels gens causals d’aquestes patologies– havia de consistir exclusivament en la seqüenciació de les porcions codificants del genoma. Així per exemple, *The Institute of Genome Research* (TIGR) va ser un dels centres en els quals van començar a ser generades per retrotranscripció a partir d’mRNA i seqüenciades col·leccions de cDNA sencers, i fragmentaris (EST), per a disposar de –i poder patentar– les seqüències de possibles gens responsables de malalties humanes.

És en aquest moment, cap a 1998, que, al nostre grup, a partir de les dades sobre EST contingudes en les bases de dades i de la informació sobre els teixits en els quals havien estat aïllades, es va iniciar una cerca *in silico* d’EST de retina que mapessin dins del locus RP26 per tal de poder caracteritzar gens candidats. La caracterització de gens, mitjançant aquest tipus d’estratègia, en una regió tan vasta no deixa de ser una aproximació amb un component atzarós alt. A la retina, com en tot teixit neural, s’expressa un nombre molt alt de gens, major que en altres tipus de teixits menys complexos estructuralment i funcional. Per tant, la cerca de candidats a través de l’escrutini d’EST hauria d’estar encaminada a la detecció d’aquelles EST que s’expressen exclusivament a la retina, indicatives de gens que codifiquen proteïnes específiques de la retina. Molts dels gens de la retinitis pigmentària cauen dins d’aquesta categoria: els gens de la cascada de fototransducció, els del cicle visual i els dels factors de transcripció en són exemples. Ara bé, aplicant aquest tipus d’estratègia, podem deixar de banda gens que també s’expressen en altres teixits i, fins i tot, aquells que ho fan de forma ubíqua. Aquest és el cas de gens que, tot i expressar-se arreu o en teixits concrets a banda de a la retina, provoquen únicament la disfunció del teixit retinal. Exemples d’aquest segon grup són els gens *PRPF8*, *PRPF31*, i *HPRP3*, que codifiquen factors de processat de premRNA d’expressió ubíqua i que, un cop alterats, causen exclusivament retinitis pigmentària autosòmica dominant, o bé, el gen *CERKL*, descrit en aquest treball, entre d’altres. En resum, la cerca d’EST en una regió tan àmplia s’assembla força a buscar una agulla en un paller i només en aquells casos en què s’identifica una EST que correspon a un homòleg d’un gen RP ja descrit –com per exemple va succeir en el locus *ACHM3* d’acromatòpsia– aquest tipus d’estratègia pot conduir a la identificació del gen causal.

Identificació d'un nou gen humà en el locus RP26: el gen ORMDL1

La cerca d'EST en el locus RP26 ens va permetre identificar un nou gen, *ORMDL1*, i caracteritzar que forma part d'una família gènica conservada evolutivament als eucariotes, constituïda per dos altres membres, *ORMDL2* i *ORMDL3*, en humans i a la resta de vertebrats, i que presenta còpies duplicades en organismes com ara *Saccharomyces cerevisiae* i *Arabidopsis thaliana*. En aquest treball, hem demostrat que en llevat, la presència intacta d'una de les dues còpies dels gens *ORM* és necessària per a la resistència contra agents com ara el ditiotreitòl (DTT) o la tunicamicina, que indueixen la resposta a proteïnes malplegades (*unfolded protein response*, UPR) al reticle endoplasmàtic. La pressumpta funció dels gens *ORMDL* en el plegament de proteïnes al reticle endoplasmàtic –sugerida a partir de les dades del fenotip del doble *knockout* *ORM1* *ORM2* de llevat i de la localització subcel·lular de les proteïnes *ORMDL*– feien del gen *ORMDL1* un bon candidat com a gen *RP26*. L'alteració dels processos de plegament de proteïnes al reticle endoplasmàtic ha estat implicada en la patogènesi de determinades distròfies retinals, i és en aquest escenari que podríem situar *ORMDL1* com a candidat. Per exemple, s'ha vist que molts dels efectes de les mutacions en el gen de la rodopsina provoquen un plegament defectuós de la proteïna que no pot col·locar-se correctament a la membrana dels discos del segment extern, n'impedeix la biogènesi i causa, així, la degeneració dels fotoreceptors (Chapple et al. 2001). En aquest sentit, és creu que un dels gens responsables de l'amaurosi congènita de Leber, el gen *A1PL1* –que codifica una proteïna similar a la proteïna que interacciona amb el receptor d'arilhidrocarboni (*aryl hydrocarbon receptor-interacting protein-like*)– podria actuar com a xaperona o coxaperona (Chapple et al. 2001). Recentment, s'ha vist que la mutació del gen de l'anhidrasa carbònica IV (RP17) provoca el plegament incorrecte de la proteïna, la inducció de l'UPR i, finalment, l'apoptosi en cèl·lules en cultiu transfectades amb la còpia mutada del gen, i que els efectes de la mutació es redueixen en utilitzar un agent químic que actua com a xaperona (Bonapace et al. 2004, Rebello et al. 2004).

En aquest escenari, les evidències a favor de la implicació en distròfies retinals d'alteracions de la resposta a proteïnes malplegades apuntaven cap a *ORMDL1*. No obstant això, la consideració d'*ORMDL1* com a candidat podia ser posada en dubte si consideràvem dues qüestions. D'una banda, el fet que, tot i expressar-se a la retina, *ORMDL1* és un gen d'expressió ubíqua. Per tant, la seva alteració podria afectar altres teixits. De l'altra, la presència al genoma dels dos homòlegs d'aquest gen, *ORMDL2* i *ORMDL3*, i la redundància funcional de les còpies duplicades. L'existència d'una certa conservació evolutiva de la funció de les proteïnes *ORMDL* quedava palesa als assajos de rescat fenotípic en llevat –per complementació del doble *knockout* amb el gen *ORMDL3* humà–. En aquest sentit, *ORMDL1* no era un bon candidat si l'efecte d'una alteració en aquest gen podia ser contrarestada pels altres dos gens. Contràriament, podríem pensar en una especialització de les còpies duplicades i en una redundància parcial de funcions, situació en la qual l'alteració d'*ORMDL1* no seria rescatada completament per cap dels altres dos gens i, progressivament, donaria lloc a un procés de degeneració retinal. Tot i aquestes especulacions, el gen *ORMDL1* va ser descartat com a

gen *RP26* després de seqüenciar-lo als pacients de la família P2 i no trobar cap mutació en la seqüència codificant.

No obstant, la caracterització d'aquest gen va posar de relleu que la regió del cromosoma 2 en la qual es troba situat el locus *RP26*, a 2q31-q32, és paràloga a dues altres regions dels cromosomes 12 (a 12q13) i 17 (a 17q21), en les quals vam identificar els gens paràlegs *ORMDL2* i *ORMDL3*, respectivament. Es creu que la complexitat observada als genomes dels vertebrats –i, per extensió, al genoma humà– es va originar, en part, a través de duplicacions genòmiques, totals o segmentals, que van tenir lloc en l'ancestre dels vertebrats (Sidow 1996). Aquestes regions, com les dels gens *ORMDL* i els complexos gènics *HOX*, són el vestigi de fenòmens genòmics a gran escala que van permetre als vertebrats un gran increment funcional mitjançant la generació de múltiples gens nous per duplicació i divergència diferencial de les còpies duplicades. Així, actualment reconeixem que molts dels gens dels vertebrats, com també passa en molts altres organismes, formen part de famílies gèniques. Els membres d'una família gènica poden fer funcions lleugerament diferents i parcialment solapades o, per contra, poden haver adoptat nous papers. En alguns casos una de les còpies duplicades s'expressa de forma específica en un teixit o en un tipus cel·lular concret i, en canvi, l'altra ho fa de manera ubíqua, i per tant, tot i ser pràcticament idèntiques, les còpies duplicades poden fer funcions diferents segons l'instant o el lloc en què s'expressen. L'estudi de les famílies gèniques aplicat a la genètica humana ha servit en determinades situacions per identificar nous gens responsables de patologies en les quals havia estat caracteritzat prèviament com a gen causal un dels altres membres de la família gènica. En aquests casos, els altres membres de la família gènica passen a ser considerats automàticament gens candidats. En la retinitis pigmentària, els gens que codifiquen les subunitats α i β dels canals catiónics dels bastons –causals de retinitis pigmentària autosòmica recessiva– i els dels cons –causals d'acromatòpsia– formen part d'una mateixa família gènica. El mateix succeeix amb els gens *GNAT1* i *GNAT2* que codifiquen, respectivament, la subunitat α de la transducina dels bastons i la dels cons. Aquestes dues famílies gèniques exemplifiquen com el coneixement de l'existència de gens homòlegs pot ser útil per a la caracterització de nous gens patogènics. En el nostre cas, ens hagués estat útil la identificació, per part d'un altre grup, d'un nou gen *RP* en una de les altres regions cromosòmiques que formen part del mateix grup de paralogia que inclou la regió del cromosoma 2 on hi ha el locus *RP26*. Aquesta informació ens hagués permès identificar el corresponent gen paràleg del cromosoma 2 i considerar-lo com a candidat.

Cerca “genòmica” del gen causal d'RP en el locus RP26

A mesura que els desenvolupaments tècnics en seqüenciació automàtica s'anaven succeint, cada cop van ser més les veus que defensaven la seqüenciació completa del genoma humà. A banda de raons tècniques, la seqüenciació completa del genoma també quedava justificava per la importància creixent que començaven a prendre les regions intergèniques del genoma i els segments

no codificants dels gens, que havien deixat d'ésser considerats DNA escombraria. Tot i que la contrucció de grans col·leccions d'EST i de cDNA ha permès identificar un gran nombre de gens humans nous, en aquestes col·leccions hi són poc representats aquells gens que s'expressen poc, o que són específics d'un teixit o d'un estadi del desenvolupament que no han estat inclosos a la col·lecció. Així doncs, amb aquesta estratègia no era possible identificar tots i cadascun dels gens humans. Per tant, la seqüenciació completa del genoma –en combinació amb la informació derivada de les dades d'expressió (EST, cDNA) i amb algorismes cada cop més refinats de predicció de gens– era absolutament indispensable per assolir una cartografia completa de tots els gens, inclosos els que presenten nivells baixos d'expressió i entre els quals hi ha també gens causals de malalties. D'altra banda, la seqüència completa del genoma humà també seria d'utilitat per a disposar d'un mapa físic en el qual poder situar amb exactitud elements no gènics, com ara els marcadors polimòrfics del mapa genètic.

A començaments de l'any 2002, les dades de seqüència genòmica de la regió del cromosoma 2 on havia estat definit el locus *RP26* no eren encara completes. Tot i això, podíem aventurar-nos a situar sobre un mapa físic fragmentari els marcadors polimòrfics que havien estat emprats per al mapatge del locus per tal de visualitzar-ne les dimensions reals. D'aquesta manera, podíem identificar també la posició dels marcadors homozigots, prop dels quals hauria d'estar localitzat el gen *RP26*, i dels gens coneguts.

El locus abarcava una regió de més de 17 milions de parells de bases, entre els marcadors D2S148 i D2S117, en la qual hi havia una cinquantena de gens coneguts i que, probablement, si teníem en compte les dades d'EST i mRNA de les bases de dades, es podria ampliar a un centenar de gens. D'entrada, vam llistar els gens coneguts del locus i vam fer cerques bibliogràfiques per a disposar de dades que ens permetessin trobar una connexió entre algun d'aquests gens i la funció/disfunció retinal. Aquí cal apuntar que, a causa de la complexitat de les cèl·lules retinals, de la quantitat de gens que s'expressen a la retina, i de la varietat de processos cel·lulars alterats en la retinitis pigmentària, conjuntament amb l'alt nombre de gens del locus, era relativament fàcil establir connexions entre la patologia i determinats gens del locus. Com veurem més endavant, aquesta concentració de gens candidats en el locus, probablement, no és atzarosa. Aplicant el criteri funcional, podíem quedar-nos amb un llistat més reduït de gens, amb el qual començar a treballar, però havíem de tenir en compte que el gen *RP26* podia no ser cap dels gens inclosos al llistat: podia no haver estat caracteritzat encara, o podia haver estat identificat, però ser un gen de funció desconeguda i, per tant, d'adscripció difícil com a candidat, o podia ser que alguns dels gens del llistat es trobessin en regions d'heterozigositat i que, per tant, poguessin ser descartats aplicant un criteri posicional. Així doncs, calia combinar les dades del nombre de gens del locus i de la funció i expressió de cadascun d'ells amb informació de tipus posicional per tal d'identificar un subgrup de gens idoni com a candidats. La clau per identificar el gen *RP26* era la informació posicional: havíem de reduir la mida del locus, ja fos, a partir de noves famílies en les quals la malaltia cosegregués amb marcadors del locus i que ens permetessin identificar noves recombinacions i definir una regió

crítica, o a partir d'una estratègia de mapatge d'homozigositat que ens conduís a la identificació d'una regió d'homozigositat candidata. Aquesta darrera estratègia es pot aplicar exclusivament per al mapatge de gens de malalties autosòmiques recessives, partint de la premisa que la família estudiada és consanguínia, com en el nostre cas.

Inicialment, la segona de les estratègies posicionals –el mapatge d'homozigositat– era prematura, perquè l'ordre dels marcadors microsatèl·lits en el mapa físic no era encara definitiu i hi havia alguns buits de seqüència, contigs mal mapats i marcadors no situats en el mapa. D'altra banda, la mida immensa del locus feia indispensable l'anàlisi d'un nombre molt alt de marcadors per poder cobrir tot el locus i realitzar un mapatge d'homozigositat en profunditat, que no podia ser dut a terme exclusivament amb marcadors microsatèl·lits. Havíem d'extendre aquesta anàlisi a algun altre tipus de marcadors polimòrfics, molt més abundants i que, per tant, ens permetessin analitzar l'homozigositat/heterozigositat en posicions situades a menys d'1 Mb de distància. Cap a finals dels anys noranta, la seqüenciació incipient del genoma havia posat de manifest l'existència d'un nou tipus de marcadors polimòrfics, els marcadors d'un únic nucleòtid (SNP), molt més abundants que qualsevol dels altres marcadors genètics emprats fins llavors i distribuïts arreu del genoma. Ara bé, el mapatge d'homozigositat amb SNP en el locus *RP26* hauria d'esperar, encara uns mesos, fins a l'estiu d'aquell any, que el nombre d'SNP i, sobretot, les dades sobre la seva heterozigositat fossin més abundants.

Definició d'una regió crítica a partir de la informació d'una segona família RP

Així doncs, la cerca del gen *RP26* va començar amb la primera de les estratègies posicionals: la definició d'una regió crítica dins del locus a partir de les dades de cosegregació derivades de l'estudi d'altres famílies. En aquest sentit, únicament disposàvem de la informació provinent d'una família xinesa que havia estat analitzada per un grup de la Universitat de Fudan a la Xina. Tot i que aquesta família no contenia el nombre d'individus afectats de retinitis pigmentària necessari per a poder-la analitzar per lligament genètic, una anàlisi de cosegregació amb marcadors del cromosoma 2 va revelar que la malaltia en aquesta família cosegregava amb el locus *RP26*. En concret, en l'interval comprès entre els marcadors D2S335 i D2S2366 (FIGURA 1 DE RESULTATS). En considerar aquestes dades, havíem de tenir en compte que provenien d'una anàlisi de cosegregació; per tant, no tenien la significació estadística d'una anàlisi de lligament i es podia tractar d'una cosegregació espúria i que realment el gen responsable de la malaltia en aquesta família es trobés en una altra localització del genoma. D'altra banda, podia ser també que, tot i cosegregar amb marcadors del locus *RP26*, el gen causal en aquesta família no fos el mateix que el de la família P2. En aquests dos casos, la interpretació conjunta de les dades de la família xinesa i de la família espanyola ens estaria conduint a conclusions equivocades, però si, per contra, suposàvem que ambdues havien de tenir el mateix gen alterat, ens podria permetre reduir considerablement la regió en la qual cercar el gen *RP26*.

El solapament dels dos intervals de cosegregació –el de la família P2 i el de la família xinesa– definia una regió crítica d’unes 6 Mb entre els marcadors D2S148 i D2S2366. Aquesta regió contenia una cinquantena de gens, entre els quals, gens de grans dimensions, com el que codifica la proteïna muscular titina, i altres gens, com els que formen el complex D de gens HOX. L’anàlisi de marcadors microsatèl·lits addicionals en la família P2 en el marge centromèric del locus RP26 ens va permetre identificar tres nous marcadors pels quals el pacient II.5 era recombinant –els microsatèl·lits D2S324, D2S2978 i D2S2261–. De tots ells, el més distal era D2S2261, que esdevenia el nou marge centromèric del locus. D’aquesta manera la regió crítica quedava reduïda a un interval d’aproximadament 3 Mb, entre D2S2261 i D2S2366, que contenia únicament vuit gens. La conjunció de les dades de les dues famílies i la informació posicional obtinguda de l’anàlisi dels microsatèl·lits ens havia permès reduir substancialment la quantitat de gens per analitzar. Aquest nombre de gens era assumible per a ser seqüenciats directament, tot i que algun d’ells presentava un nombre d’exons considerable, i podien ser prioritzats aplicant un criteri funcional, com ara dades sobre la seva expressió a la retina. L’escrutini en detall dels gens de la regió crítica també es va veure afectat per l’estadi en què es trobava la seqüenciació d’aquesta regió distal del cromosoma 2. Vam començar a treballar amb les dades de seqüència de desembre de 2001, que havien servit per construir l’ordenament del genoma UCSC hg10 (*Human Draft Genome Assembly UCSC hg10*), en el qual encara hi havia molts buits de seqüència que ens van dificultar la col·locació d’algun dels gens de la regió, com ara *NCKAP1* i *FRZB*. En ordenaments subseqüents, aquestes discontinuïtats de seqüència van ser corregides, i també van ser mapats a la regió crítica nous mRNAs de GenBank, fet que va ampliar el nombre de gens per analitzar. Tot i això, algun dels gens que vam incloure al nostre escrutini, com ara *elf4G* i *ZNF533* –que formaven part de la regió crítica a l’ordenament d’abril de 2002– en ordenaments posteriors, a partir del de juny del 2002, van ser remapats fora de l’interval. Aquest fet il·lustra la dificultat de treballar amb dades encara parcials de seqüència i amb ordenaments del genoma no definitius. Finalment, i descomptant els dos gens remapats, a finals de l’estiu de 2002, havíem analitzat 10 gens de la regió crítica sense identificar cap mutació, encara que alguns d’ells, com *NEUROD1* o *FRZB*, eren molt bons candidats funcionals: el factor de transcripció *NEUROD1*, pel seu paper en la determinació cel·lular a la retina i pel fenotip de degeneració retinal d’una soca de ratolí *knockout*, i el gen *FRZB* –que forma part de la via Wnt/Frizzled– per la seva contribució a l’establiment de la polaritat i la diferenciació de determinades poblacions de neurones de la retina. Fins i tot el gen *ERdj5* (*PDIP5*), de presumpta funció xaperona, podia ser un candidat funcional tenint en compte les dades sobre la implicació, en determinades distròfies retinals, d’altres xaperones i de l’alteració dels processos de plegament de proteïnes al reticle, que hem esmentat anteriorment. El fet que cap dels gens analitzats fos el gen *RP26*, ens va fer pensar que, probablement, la hipòtesi que en les dues famílies l’alteració genètica causal de retinitis pigmentària havia de ser la mateixa no era certa, i que les dades de la família xinesa ens havien conduït erròniament a la definició d’una regió crítica en un dels extrems del locus, quan potser el gen *RP26* es trobava en una altra banda.

Mapatge d'homozigositat: els polimorfismes d'un únic nucleòtid

En aquest punt del nostre estudi, les bases de dades d'SNP ja començaven a contenir un alt nombre de marcadors i, el que era més important, informació sobre l'heterozigositat d'alguns d'ells. D'altra banda, s'havien desenvolupat motors de cerca d'SNP, vinculats a les bases de dades existents (NCBI dbSNP o *The SNP Consortium*), que permetien definir les posicions físiques entre les quals volíem establir la cerca –en bp, o emprant microsatèl·lits o gens com a punts de referència–. Tant la informació continguda en les bases de dades com les noves eines bioinformatiques, ens van fer plantejar que era un bon moment per a iniciar una aproximació de mapatge d'homozigositat per tal de localitzar el gen *RP26*, ja que l'escrutini de gens de la regió crítica no havia donat els fruits esperats i perquè necessitàvem una estratègia que ens permetés definir una subregió per analitzar. El mapatge d'homozigositat es fonamentava en les dades existents sobre la consanguinitat de la família P2 i sobre l'existència d'un presumpte interval d'homozigositat definit en l'estudi inicial del locus (Bayés et al. 1998). Tenint en compte la consanguinitat de la família, el gen responsable de la malaltia havia d'estar situat en una regió per a la qual els haplotips matern i patern dels malalts RP fossin iguals; perquè haurien hereditat, tant per via materna com paterna, la mateixa mutació responsable de la malaltia i els mateixos al·lels dels marcadors propers al gen *RP26*.

En aquest cas, el que tractàvem de determinar era l'AUTOZIGOSITAT d'una regió interna del locus; el terme autozigositat fa referència, segons (Hartl i Clark 1997), a l'existència d'un segment idèntic derivat per descendència, i no per atzar –en el cas que la freqüència de determinats al·lels dels marcadors d'aquest segment fos molt alta a la població, i, per tant, fos relativament comú ser homozigot per aquests marcadors–. Així, un individu és autozigot per un locus si les dues còpies del locus provenen d'un mateix al·lel ancestral, i.e. si els dos al·lels són idèntics per filiació (Hartl i Clark 1997). L'autozigositat es produeix en el cas que els dos progenitors comparteixin un segment cromosòmic idèntic per filiació i que els dos segments idèntics siguin hereditats per algun dels fills. Només en aquest cas, l'homozigositat observada és un criteri útil per al mapatge de gens causals de malalties.

Un cop situats en el mapa físic els 8 marcadors emprats en l'anàlisi de lligament, i tenint en compte l'heterozigositat del marcador D2S350 als pacients de la família P2, el locus quedava dividit en dues regions: una primera regió de 2.5 Mb, entre D2S2261 i D2S350, que contenia un marcador homozigot central (D2S364) i que corresponia a grans trets amb la regió crítica (FIGURA 16 DE RESULTATS); i una segona regió que corresponia a la resta del locus, entre D2S350 i D2S117, que contenia els altres tres microsatèl·lits homozigots (D2S118, D2S389, D2S318). A la primera de les regions, havíem analitzat tots els gens coneguts, i a la segona, molt més gran, els tres marcadors homozigots estaven situats molt propers en un segment de gairebé una megabase, entre D2S118 i D2S318. De manera que calia determinar l'abast de l'homozigositat en les dues regions: si la gran regió distal contenia una subregió d'homozigositat centrada als tres marcadors homozigots o, pel contrari, en alguna altra banda de l'interval, o si l'homozigositat del marcador D2S364 s'extenia al llarg de la regió crítica.

L'abordatge inicial d'aquesta qüestió va ser realitzat mitjançant l'anàlisi de 16 marcadors microsatèl·lits addicionals. D'entrada, no vam poder localitzar nous microsatèl·lits a la regió crítica – que caldria analitzar mitjançant SNP–, però l'anàlisi de nous marcadors a la meitat distal del locus va posar de manifest l'existència d'una successió de punts d'homozigositat i d'heterozigositat. El microsatèl·lit D2S318 quedava separat de D2S118 i D2S389 per un punt d'heterozigositat, el marcador D2S2246 (FIGURA 16 DE RESULTATS). Prop de D2S118 i D2S389, vam localitzar un altre marcador homozigot (D2S1775). De resultes, D2S118, D2S389 i D2S1775 quedaven dins d'un segment d'homozigositat d'1.1 Mb delimitat pels marcadors heterozigots D2S2262 i D2S2246. Altres punts d'homozigositat eren el marcador D2S1391 en un segment de 2 Mb, entre D2S2281 i D2S1361; i els microsatèl·lits D2S280 i D2S315 en un segment de 2.9 Mb, entre D2S161 i D2S117. De fet, aquest darrer segment va quedar reduït a 857 kb, perquè l'individu II.15 –un dels fills amb retinitis, de la família P2– era recombinant pel marcador D2S273, microsatèl·lit que passava a esdevenir el marge distal del locus. L'anàlisi de microsatèl·lits addicionals ens va permetre, doncs, també, definir amb exactitud l'interval de cosegregació del locus RP26 que comprèn 12 Mb, entre els marcadors D2S2261 i D2S273.

La visió global del locus, després de l'anàlisi de tots aquests microsatèl·lits, mostrava l'absència d'una regió clara d'homozigositat que s'extengués més enllà de les 3 Mb. Així, si existia tal regió d'homozigositat havia d'abarcabar un segment menor de 3 Mb. Per tal de determinar si això darrer era cert, vam realitzar tres tandes de genotipatge d'SNP a la família P2 que, globalment, van suposar l'anàlisi de 101 marcadors addicionals i que ens van permetre cobrir el locus a intervals de 500 kb. Després de l'anàlisi de microsatèl·lits i de la primera tanda de genotipatge d'SNP, teníem dubtes seriosos sobre si el locus contenia realment una regió d'homozigositat, ja que continuàvem trobant arreu una alternança de marcadors homozigots i heterozigots i cap segment en què l'homozigositat fos contínua.

Les dades no concloents del genotipatge de marcadors en el locus RP26 ens van fer plantejar l'anàlisi de gens candidats, independentment de l'homozigositat/heterozigositat del segment on es trobaven. Les 12 Mb de cosegregació del locus contenen, probablement, més d'una seixantena de gens, entre els quals hi ha diversos bons candidats com a gens causals de distròfies retinals. Hem vist que la regió crítica contenia gens com ara *NEUROD1* o *FRZB*, que s'expressen a la retina i regeixen aspectes com la determinació, la diferenciació i la polaritat cel·lular –processos en els quals intervenen altres gens implicats en la retinitis–, i que en el locus hi ha gens com *ERdj5* o *ORMDL1*, relacionats amb el plegament de proteïnes al reticle endoplasmàtic, que també ha estat associat a la patogènesi de la malaltia. A banda d'aquests, l'interval de cosegregació també conté altres excel·lents candidats: el gen *ITGAV* –que codifica la subunitat alfa de la integrina V que participa en el procés de fagocitosi dels segments externs dels fotoreceptors– i el gen *GLS* –que codifica l'enzim glutaminasa, important per a la generació d'un dels neurotransmissors principals de la retina, el glutamat–.

¿Per què aquesta acumulació de gens candidats en una mateixa regió genòmica? Probablement, alguns d'aquests gens, que tenen un paper demostrat o presumpte en la biologia

retinal, comparteixen una regulació transcripcional comuna i aquesta regulació condiona la seva localització en un mateix segment cromosòmic. L'acumulació de gens candidats en el locus RP26 subratlla la dificultat d'aplicar en un segment d'aquestes característiques, definit per lligament, una aproximació com la del candidat posicional. Aquesta estratègia únicament és útil en els casos en què podem reduir considerablement l'interval de cosegregació, o bé, en d'altres en els quals, dins d'un interval gran, hi ha un únic candidat clar. En el primer cas però, encara que acotem una regió de menys de 3 Mb, el nombre de gens continguts pot ser elevat i, finalment, acabarem seqüenciat diversos candidats posicionals, no sempre de manera exitosa i, sovint, el gen causal resulta ser el més inesperat de tots.

En el nostre cas, l'anàlisi dels gens *ITGAV* i *GLS* tampoc va ser fructífera. Malgrat això, alguns d'aquests gens continuen sent excel·lents candidats com a gens causals de retinitis per a altres famílies en què la malaltia mostri lligament amb marcadors a 2q31-q32. Posem per cas que, després d'analitzar el gen *CERKL* en aquestes hipotètiques famílies RP26, no hi haguéssim identificat cap mutació. La retinitis en aquestes famílies podria ser causada, doncs, per l'alteració d'algun dels altres candidats (*NEUROD1*, *FRZB*, *ITGAV*, etc.). En aquest sentit, hi ha diversos exemples de regions cromosòmiques en les quals es concentren gens implicats en distròfies retinals, com ara la regió a 6p21 que conté els gens *TULP1*, *RDS*, i *GUCA1A*, i no seria d'estranyar que amb la regió del cromosoma 2 on hi ha el locus RP26 passés el mateix.

En les seqüències intròniques dels gens *ITGAV* i *GLS*, vam poder caracteritzar diversos SNP, pels quals els pacients eren heterozigots. L'exclusió per seqüenciació d'aquests candidats, conjuntament amb l'evidència que es trobaven situats en segments no homozigots, ens va fer plantejar seriosament altre cop la possibilitat que el locus RP26, a la família P2, contingués una regió d'homozigotitat; és a dir, que la causa de la malaltia en aquesta família fos deguda a la seva consanguinitat, i que els individus afectes haguessin hereditat el mateix al·lel mutat a partir de l'ancestre comú portador.

El genotipatge d'SNP va ser realitzat per seqüenciació de fragments –d'entre 200 i 800 bp, aproximadament– amplificats per PCR, que contenien l'SNP en una posició central. En els casos en què era possible, vam tractar d'amplificar fragments que continguessin més d'un SNP per tal d'assegurar-nos la informativitat del fragment analitzat a la nostra família. Generalment, els SNP van ser seleccionats en funció de la seva heterozigotitat, i.e. triàvem aquells SNP l'heterozigotitat dels quals era superior al 30% i aquells pels quals disposàvem de dades sobre l'heterozigotitat en població caucàsica. D'aquesta manera, incrementàvem la probabilitat que algun dels dos progenitors de la família fos heterozigot i, per tant, que l'SNP fos informatiu per a l'anàlisi d'homozigotitat. El fet que es tracti de marcadors dial·lèlics repercuteix enormement en la baixa informativitat dels SNP, per això és tan essencial conèixer-ne l'heterozigotitat. Abans de genotipar els fills, però, sempre genotipàvem els pares per determinar si els SNP que havíem triat eren informatius a la família P2. En algunes ocasions, per tal de poder fragmentar una subregió del locus, aproximadament cada 500 kb,

havíem de triar SNP dels quals no disposàvem d'informació sobre la seva heterozigositat, perquè no n'hi havia d'altres, i confiar que el genotipatge dels pares revelés que eren informatius.

Com hem vist, l'anàlisi d'SNP ens va permetre assolir l'elevada densitat de marcadors necessària per a determinar amb un grau de certesa alt la presència d'un interval continu d'homozigositat. Després de les tres tandes de genotipatge, vam assolir la densitat informativa idònia i vam determinar que, efectivament, el locus RP26 contenia una regió d'homozigositat. Vam poder fragmentar el locus en segments de mida inferior a una megabase –flanquejats per marcadors heterozigots– i vam caracteritzar una única regió en la qual l'homozigositat abarcava més d'una megabase –en concret 2.5 Mb, entre els marcadors D2S2261 i rs7510– i que contenia un dels marcadors homozigots de l'anàlisi de lligament, el microsatèl·lit D2S364. L'altre punt d'homozigositat, al voltant del marcadors D2S118, D2S389 i D2S1775, quedava fragmentat per la presència d'un SNP heterozigot (rs1263100) entre D2S118 i D2S389. Aquesta darrera dada visualitzava com podia variar la nostra concepció sobre l'homozigositat/heterozigositat d'una regió concreta dins del locus depenent de la profunditat de l'anàlisi. Inicialment, l'existència de 3 marcadors homozigots a la regió distal ens podia fer sospitar que es tractava d'una regió homozigota *de facto*, però en incrementar el nivell d'anàlisi fins a 500 kb va resultar que aquests microsatèl·lits estaven flanquejats per microsatèl·lits heterozigots molt propers (D2S2262 i D2S2246) i que, entre ells, hi havia un SNP heterozigot. Aplicant el criteri de l'homozigositat, el gen *RP26* no podia estar localitzat en aquesta part del locus.

La regió d'homozigostat (regió candidata) coincidia en gran part amb la regió crítica que havíem definit a partir del solapament dels intervals de cosegregació de la família P2 i de la família xinesa. En quedaven exclosos, però, els gens *NCKAP1*, *SKRP1*, i *MP44*, situats més enllà del marge distal de la regió candidata, l'SNP rs7510. Finalment, les dades del mapatge d'homozigositat i les derivades de l'anàlisi de més famílies RP (cosegregació xinesa) convergien en una mateixa regió. De les 17.4 Mb de cosegregació sobre el mapa físic de començaments del 2002 havíem passat a 12 Mb de cosegregació –després d'haver establert finament els marges proximal i distal del locus– i en darrer terme, a la identificació d'una regió candidata de 2.5 Mb a finals del 2002. Tenint en compte que la recombinació meiótica fragmenta els haplotips, la longitud d'un segment d'homozigositat (autozigositat) reflecteix l'antiguitat del segment i, per tant, un segment autozigot curt indica un origen llunyà. Si extenem aquest raonament a tot el genoma, podem veure com la proporció d'autozigositat que conté el genoma d'un individu és una mesura del grau de parentesc dels seus progenitors (Broman i Weber 1999). En el cas de la família P2, les dades genealògiques apunten que els progenitors comparteixen un avantpassat comú quatre generacions enrera, un rebesavi, i que per tant són cosins tercers. La descendència d'aquest matrimoni entre cosins tercers comparteix l'haplotip al llarg de 12 Mb (interval de cosegregació), de les quals únicament 2.5 Mb (regió d'autozigositat) s'han anat heredant conjuntament com a haplotip, sense ser fragmentades per recombinació meiótica, al llarg de les cinc generacions que separen els fills de la família P2 de l'avantpassat comú dels progenitors.

Identificació del gen RP26

La regió d'homozigositat contenia 7 gens indexats en RefSeq (*UBE2E3*, *ITGA4*, *NEUROD1*, *SSFA2*, *PDE1A*, *ERdj5* i *FRZB*) i 2 mRNAs de GenBank (PRO1316 i AF494535). D'aquests, n'havíem seqüenciat 7 i restaven per analitzar *ITGA4* i AF494535. La seqüenciació d'aquests dos gens no ens va permetre identificar cap mutació i, per tant, vam cercar a les bases de dades evidències de l'existència d'altres mRNA a la regió. Mitjançant aquesta cerca, vam localitzar diversos mRNA dels gens *UBE2E3* i *SSFA2* que contenien exons addicionals que allargaven les pautes de lectura d'ambdós. L'anàlisi dels exons addicionals d'aquests gens tampoc ens va permetre identificar el gen *RP26*.

En aquest punt del projecte, en el qual ja havíem seqüenciat tots els gens cartografiats a la regió candidata, ens vam plantejar la possibilitat que el gen *RP26* fos un gen nou, encara no descrit, i a tal efecte, vam iniciar una cerca d'EST en les bases de dades. Per tal de caracteritzar nous gens, ens vam fixar en aquelles EST mapades a la regió candidata que estiguessin compostes per més d'un exó, indicatives que podien provenir de gens *de facto* i que no constituïen seqüències genòmiques curtes que es transcrivien. La cerca d'EST i cDNA parcials va donar fruit i ens va portar a la identificació d'un nou gen amb una pauta de lectura oberta de 1596 bp que s'extenia al llarg de 13 exons i que codificava una proteïna de 532 aminoàcids.

El gen CERKL codifica una proteïna similar a la ceramida quinasa

Aquest nou gen, identificat a la regió candidata del locus RP26, contenia un domini homòleg al que presenten les proteïnes de la família de quinases lipídiques que agrupa les diacilglicerol quinases, les esfingosina quinases i les ceramida quinases. És amb aquestes darreres, amb les quals la proteïna codificada per aquest gen mostrava el grau més alt de similitud; en concret, amb la ceramida quinasa descrita en humans, CERK (Sugiura et al. 2002), amb la qual compartia un 29% d'identitat i un 50% de similitud aminoacídiques. Per aquesta raó, vam anomenar aquesta proteïna, proteïna similar a la ceramida quinasa –*ceramide kinase-like protein* (CERKL)– segons les directrius terminològiques del Comitè de nomenclatura de la Organització del genoma humà (HUGO). L'anàlisi filogenètica de diverses seqüències de quinases lipídiques situava clarament CERKL dins de la subfamília de les ceramida quinases.

La seqüenciació dels exons de *CERKL* ens va permetre identificar una mutació en l'exó 5 que introduïa un codó prematur de terminació. La mutació (R257X) es trobava en homozigosi en tots els pacients de la família P2, en la qual cosegregava amb la malaltia, i no va poder ser identificada en cap dels 340 cromosomes analitzats que provenien d'individus control –nombre de cromosomes que permet identificar un polimorfisme amb una freqüència de l'1% amb una significació estadística del 95% (Collins i Schwartz 2002). Les dades de cosegregació, l'absència del canvi en població control i la natura d'aquest canvi –la introducció d'un codó de terminació– demostren de manera concloent que l'alteració identificada en aquest nou gen és la causa de la retinitis dels pacients de la família P2.

L'anàlisi de cosegregació realitzada amb els microsatèl·lits D2S2310 i D2S364, que flanquegen el gen, en sis de les 55 famílies espanyoles amb retinitis pigmentària autosòmica recessiva –que han estat estudiades al nostre grup– ens va permetre localitzar una altra família en la qual la malaltia cosegregava amb aquests marcadors. Les sis famílies que es van emprar en aquesta anàlisi corresponien a famílies en les quals el locus RP26 no havia pogut ser descartat en les anàlisis de cosegregació realitzades anteriorment. En aquesta nova família (E1) vam seqüenciar tots els exons del gen i vam identificar la mateixa mutació R257X.

Orígen únic o punt calent mutacional?

La identificació de la mateixa mutació en les dues famílies ens va fer plantejar si d'alguna manera estaven emparentades i, per tant, la mutació R257X tenia un origen únic o si, per contra, es tractava d'esdeveniments mutacionals independents. Per a comprovar aquesta qüestió, vam analitzar en els pacients de la família E1 l'haplotip en una regió de 600 kb al voltant del gen (12 marcadors) per tal de determinar si era compartit amb el dels pacients de la família P2. L'anàlisi haplotípica va demostrar que els pacients de les dues famílies no compartien el mateix haplotip i que, per tant, molt probablement, la mutació R257X no tenia un origen únic. La patronímia i la procedència geogràfica de les famílies –una procedent d'un poble del nord de la província de Màlaga i l'altra d'Euskadi– tampoc sostenen un parentesc entre ambdues, tot i que no podem descartar que l'origen de la mutació es remunti moltes més generacions enllà dels avantpassats comuns de les dues famílies i que, per tant, després de tantes generacions no s'hagi conservat l'haplotip proper al gen.

Així doncs, si la mutació no té un origen comú, com és que les dues famílies la comparteixen? En aquest sentit, el codó 257 podria ser un punt calent mutacional: l'arginina 257 està codificada pel codó CGA –un dels codons de poc ús en humans– en el qual és relativament fàcil que es pugui produir una transició a la primera posició del codó (C→T) que el transformi en un codó de terminació. Precisament per aquest fet, l'ús d'aquest codó és baix i, generalment, l'aminoàcid arginina és especificat per un dels altres cinc codons que el codifiquen. En CERKL, de les 38 R que conté la proteïna, únicament 4 estan codificades pel codó CGA. El codó R257 es troba situat en l'exó 5, que no forma part dels transcrits curts del gen i potser, per això, la transició en aquest codó no presenta una gran pressió selectiva en contra; tenint en compte, a més, que l'efecte fenotípic que causa la mutació és una malaltia progressiva, circumscrita a la retina, que no impedeix tenir descendència a les persones que la pateixen. En canvi, les transicions en les arginines dels exons E1 (R66), E7 (R338) i E11 (R435) probablement estan seleccionades en contra perquè, o bé estan situades molt a l'inici de la proteïna –el cas d'R66– o bé formen part d'exons, com E7 i E11 que no són exons alternatius i, per tant, en aquest cas, no es pot atenuar en certa manera la introducció d'un codó de terminació. Així doncs, R257 pot constituir un punt calent mutacional que ens explicaria perquè les dues famílies comparteixen la mateixa mutació.

Resten encara gens per descobrir al genoma humà?

La identificació d'un nou gen en el locus RP26, a començaments del 2003, posava de relleu que, dos anys després de la publicació de l'anàlisi inicial de la seqüència del genoma humà (Lander et al. 2001) i, pocs mesos abans de la seva finalització, l'abril del 2003, no havien estat caracteritzats encara tots els gens de la nostra espècie. Ara bé, resten més gens no caracteritzats? quina fracció representen del total de gens existents? quin tipus de gens són i per què no han estat identificats encara?

Evidentment, resten encara gens per identificar, però aquests probablement no modificaran gaire les estimacions actuals del nombre de gens del genoma –situades a l'entorn dels 25000–. Hem de suposar que els gens que resten per identificar no constitueixen, per exemple, una hipotètica classe de gens –nova i nombrosa– amb característiques estructurals específiques que no els permeten ser detectats pels programes de predicció de gens, els quals han estat dissenyats per a detectar determinats patrons comuns, característics dels gens que coneixem. En aquest sentit, la descoberta recent de l'existència dels microRNA –seqüències genòmiques curtes que es transcriuen, però que no codifiquen proteïnes– constitueix un exemple d'una nova classe de “gens” o “elements gènics” que fins ara havien escapat als algorismes predictius, ja que molts d'aquests detecten els senyals d'empalmament característics dels gens multiexònics.

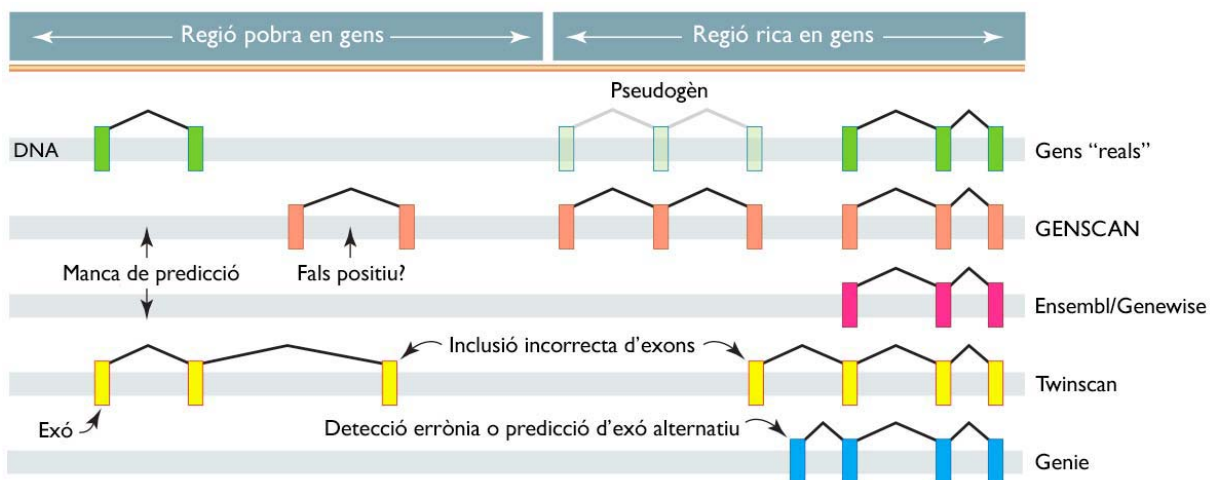


Figura 1 | Errors més freqüents en els quals incorren diversos algorismes de predicció de gens. Aquests van des de la sobrestimació del nombre de gens en considerar com a gens els pseudogens fins a la inclusió incorrecta d'exons, passant per la no predicció de gens que existeixen realment. Adaptada de Ewan Birney/Sanger Institute/Ebi.

Suposem, doncs, que els gens codificants de proteïnes que resten per identificar presenten unes característiques determinades que els fan invisibles als programes de predicció (FIGURA 1). Aquests es fonamenten principalment en dos mètodes computacionals: la detecció de patrons indicatius de l'inici i el final d'un gen (mètodes *ab initio*), i la comparació de seqüència amb gens i proteïnes descrits (mètodes comparatius). Els primers tendeixen a detectar qualsevol seqüència que

s'assembla a un gen, ja sigui un pseudogèn o bé una seqüència no gènica; els segons, només ens permeten detectar gens que s'assemblen a d'altres que ja coneixem. Actualment, els algoritmes combinen diversos d'aquests mètodes i tenen en compte, entre d'altres, els senyals d'empalmament i les dades d'expressió (cDNA parcials, EST de les bases de dades). Tot i això, continuen sent difícils de predir, per exemple, els gens d'un únic exó, els petits, els que no codifiquen proteïnes, els gens solapats i els que codifiquen més d'un producte proteic.

El gen *CERKL*, caracteritzat en aquest treball, és un exemple de gen que encara no havia estat caracteritzat, probablement, perquè no és un gen d'expressió ubíqua ni s'expressa en nivells elevats, i aquest fet dificulta l'ús de les dades d'expressió com a criteri de predicció. En les bases de dades d'EST, vam identificar únicament dues EST que corresponien a *CERKL*, i cap dels mètodes de predicció no el considerava sencer. Per exemple, Twinscan –que prediu els gens mitjançant l'homologia de seqüència entre ratolí i humà– no identificava els sis primers exons del gen ni tampoc E8; la predicció d'Ensembl, en canvi, identificava aquests sis exons, però no ho feia amb els set darrers, i GENSCAN perdia únicament els exons E1, E2, E7, E8, E11 i E12. Aquest algoritme era l'únic que considerava en un mateix model gènic alguns dels exons del cDNA parcial BC020465 (E1 a E6) i alguns dels de l'EST BE781384 (E7 a E13), fet que no és d'extranyar perquè GENSCAN no és restrictiu en el sentit que sovint inclou, en un mateix model gènic, gens propers. A banda que cap dels algoritmes no predia el gen en la seva totalitat, cap d'ells identificava l'exó diferencial que més tard hem identificat entre els exons 4 i 5, i que anomenem E4b. Probablement, els algoritmes no el predien perquè és un exó petit (78 bp), conté en el lloc 3' (donador) d'empalmament *gc* enlloc de *gt* (VEGEU L'APARTAT DE LA DISCUSSIÓ SOBRE EL PROCESSAT ALTERNATIU DE TRANSCRITS), i no forma part de cap de les EST. De manera anàloga, un dels altres exons petits de *CERKL*, l'E8 de 60 bp, no és inclòs a la predicció de Twinscan, i E9 –un exó de tan sols 26 bp– és considerat exclusivament per aquells algoritmes que fan ús de les dades d'expressió.

És curiós, també, que *CERKL* no hagués estat identificat per homologia amb quinases lipídiques i, en concret, amb la ceramida quinasa humana, ja que el percentatge de similitud entre ambdues és prou alt per poder-lo identificar mitjançant una comparació tBLASTn de la proteïna CERK contra el genoma humà. Això, probablement, es deu a que l'anotació de gens al genoma, mitjançant l'homologia de seqüència, està encaminada a la cerca dels ortòlegs de gens caracteritzats en organismes model que, en el cas de *CERKL*, eren inexistents perquè, abans del nostre treball, tampoc havia estat caracteritzat l'ortòleg murí.

En el nostre cas, gràcies a l'estudi en profunditat de la regió cromosòmica associada a una malaltia humana hem pogut identificar un gen fins ara no descrit, basant-nos en informació posicional i en l'escrutini de les bases de dades. Posteriorment, l'escrutini de les bases a la recerca de quinases lipídiques realitzat per diversos grups que treballen en el camp dels esfingolípid ha revelat *CERKL* i d'altres proteïnes, com la quinasa multilipídica (MuLK) caracteritzada recentment (Waggoner et al. 2004). Així doncs, la identificació d'aquest gen remarca la utilitat de la informació derivada de

l'estudi de malalties hereditàries humanes per a caracteritzar no tan sols els gens que les causen, sinó també per a identificar gens humans encara desconeguts.

§

Pel que fa a la funció del gen i a la seva relació amb la patogènesi retinal, l'homologia de seqüència amb la ceramida quinasa ens va fer plantejar la possible connexió de l'alteració del gen *CERKL* amb una desregulació de les vies d'apoptosi mediatades per esfingolípid, i en concret per ceramida, que podria conduir a la mort dels fotoreceptors i a l'establiment d'un procés de degeneració retinal.

Esfingolípid i apoptosi

Durant els darrers 15 anys, s'han anat acumulant dades experimentals que han posat de relleu que els esfingolípid no són mers components estructurals inerts de les membranes cel·lulars. Contràriament a aquesta visió clàssica, s'ha demostrat que el metabolisme dels esfingolípid està constituït per tot un seguit de reaccions molt ben regulades, encaminades a mantenir els nivells precisos d'una sèrie de molècules lipídiques bioactives que intervenen en processos d'importància vital per a la funció cel·lular. Entre aquestes molècules, cal destacar especialment les bases esfingoides: ceramida i esfingosina; i les seves formes fosforil·lades: ceramida-1-fosfat i esfingosina-1-fosfat. Des de 1993, any en què es va demostrar el paper de la ceramida en la inducció de la mort cel·lular programada (Obeid et al. 1993), s'han publicat innumerables estudis que han establert la importància d'aquest i altres esfingolípid en la regulació de la proliferació cel·lular, la resposta a l'estrès, i l'apoptosi.

La ceramida se sintetitza *de novo* a la cara citosòlica del reticle endoplasmàtic a partir de la condensació de l'aminoàcid L-serina amb palmitoïl CoA per acció de l'enzim serina palmitoïl transferasa (FIGURA 2). En passos successius, la 3-ceto-esfinganina formada en aquesta reacció és reduïda i *N*-acilada per una ceramida sintasa, i dóna lloc a deshidroceramida que, al seu torn, és desaturada per acció de la deshidroceramida desaturasa per a formar ceramida (*N*-acilesfingosina). La ceramida també es produeix a partir de la degradació de glicolípid, a través de glicosidases, i d'esfingomielina, mitjançant les esfingomielinases (FIGURA 2).

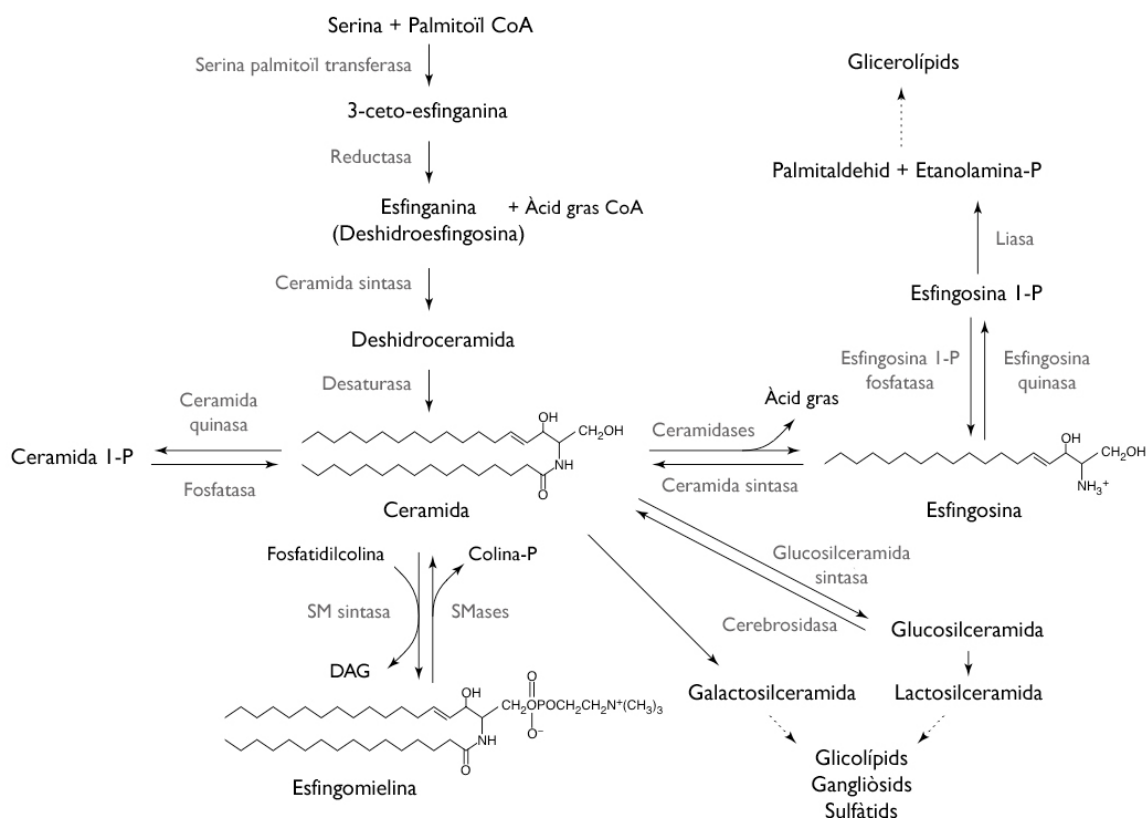


Figura 2 | El metabolisme dels esfingolípids. La síntesi *de novo* de ceramida té lloc al reticle endoplasmàtic amb la condensació de la serina i del palmitoil CoA, seguida de la reducció de la 3-ceto-esfinganina resultant i les reaccions posteriors d'acilació i de desaturació. La ceramida així produïda és incorporada, al Golgi, als esfingolípids complexos esfingomielina (SM) i glucosilceramida (cerebròsid). A partir de la degradació d'aquests compostos per esfingomielinases (SMases) i altres hidrolases que degraden els glicolípids, s'obté també ceramida, que pot ser emprada per la cèl·lula per a la síntesi d'esfingosina, mitjançant ceramidases. Tant la ceramida com l'esfingosina són fosforil·lades per les activitats ceramida quinasa i esfingosina quinasa, i donen lloc a ceramida-1-P i esfingosina-1-P. Abreviatures: DAG, diacilglicerol. Adaptada de (Hannun i Luberto 2000).

A partir de la ceramida, i per acció de les ceramidases, es produeix l'esfingosina, que a l'igual que la ceramida, també s'ha demostrat vastament que té un paper proapoptòtic i antimitogènic. La ceramida i l'esfingosina són fosforil·lades per acció de la ceramida quinasa i l'esfingosina quinasa, respectivament, que generen ceramida-1-fosfat i esfingosina-1-fosfat. El paper de l'esfingosina-1-fosfat en la cèl·lula ha estat molt investigat i s'ha establert que aquesta base esfingòide té funcions antiapoptòtiques i mitogèniques i que exerceix els seus efectes, tant intracel·lularment com extracel·lular, a través d'una família de receptors de membrana acoblats a proteïna G que, en eucariotes superiors, consta de cinc membres (Spiegel et al. 2003). El paper de la ceramida-1-fosfat, en canvi, ha estat molt menys estudiat, perquè fins fa ben poc no s'havia clonat cap gen que codifiqués un enzim amb activitat ceramida quinasa, tot i que s'havia relacionat com a agent mitogènic en fibroblasts (Gómez-Muñoz et al. 1997). D'ençà de la descripció de la proteïna CERK (Sugiura et al. 2002), han aparegut nombrosos estudis que posen de manifest la funció de la ceramida-1-fosfat en diverses vies –des del bloqueig de l'apoptosi en macròfags, fins a la regulació de

processos inflamatoris– i, a l'igual que l'esfingosina-1-fosfat, està començant a emergir com un nou lípid regulador (Pettus et al. 2003, Gómez-Muñoz et al. 2004, Pettus et al. 2004).

Diversos factors d'estrès i apoptosi indueixen un increment de ceramida –tant *de novo*, com per la degradació d'esfingomielina– que precedeix la mort cel·lular. Aquests factors són des de lligands de receptors de mort fins a agents quimioterapèutics, passant per la radiació ultraviolada i gamma, el xoc tèrmic, la manca de factors de creixement o la hipòxia.

Hi ha nombroses proves del paper de la ceramida en la promoció de l'apoptosi que evidencien que l'increment de ceramida que es produeix en una situació d'estrès cel·lular no és conseqüència del procés d'apoptosi, sinó que intervé directament en l'execució del programa de mort cel·lular. Així, s'ha vist que la sobreexpressió en cèl·lules en cultiu de factors antiapoptòtics, com ara Bcl-2 i Bcl-xL, bloqueja la fase efectora de l'apoptosi, però no inhibeix la formació de ceramida i, per tant, l'increment d'aquest esfingolípids no pot ser producte del procés de mort, sinó que l'ha de precedir. També, tant l'exposició a ceramides exògenes, com el tractament amb esfingomielinasa bacteriana o l'expressió heteròloga d'aquest enzim en cultius cel·lulars –que provoca una elevació dels nivells intracel·lulars de ceramida– desencadenen l'apoptosi. La inhibició de determinats enzims del metabolisme dels esfingolípids –ceramidases, glucosilceramida sintasa, esfingosina quinasa– també resulta en un increment de ceramida, que condueix, en molts casos, a la mort cel·lular. Per contra, la inhibició d'enzims de la via de síntesi *de novo*, com la serina palmitoil CoA transferasa o la ceramida sintasa, o la sobreexpressió d'enzims, com la ceramidasa, l'esfingosina quinasa o la glucosilceramida sintasa, es tradueix en una reducció dels nivells de ceramida i en una supressió de l'apoptosi (Levade et al. 2002). En aquesta línia, s'ha determinat que agents que confereixen resistència o que actuen com a protectors contra l'estrès cel·lular, com els èsters de forbol, el diacilglicerol o el factor de creixement de fibroblasts, actuen inhibint enzims que intervenen en la generació de ceramida. També, les cèl·lules humanes derivades de pacients de la malaltia de Niemann-Pick –que tenen una deficiència genètica en esfingomielinasa àcida– o bé les cèl·lules que provenen dels ratolins *knockout* per aquest mateix gen són resistents a diversos agents inductors d'estrès i apoptosi. D'altra banda, s'ha vist que determinats tipus cel·lulars cancerosos resistents a la mort cel·lular induïda per radiació, o bé a drogues inductores de l'apoptosi, tenen defectes en vies de producció de ceramida i que quan, mitjançant fàrmacs, es recupera la síntesi de ceramida, aquestes cèl·lules deixen de ser resistents i se n'indueix l'apoptosi (Levade et al. 2002).

La funció de la ceramida i d'altres esfingolípids en l'estrès i la mort cel·lular està conservada a l'escala evolutiva des de llevats fins a vertebrats. En *Saccharomyces cerevisiae*, s'ha determinat el paper de les bases esfingoides, com la ceramida, en la resposta a l'estrès tèrmic. Recentment, un treball en *Arabidopsis* ha posat de manifest que les ceramides també modulen la mort cel·lular programada en plantes (Liang et al. 2003). En aquest estudi, els autors caracteritzen un mutant de ceramida quinasa anomenat *accelerated cell death 5 (acd5)* que, en ser atacat per patògens, presenta un increment de l'apoptosi i dels signes de la infecció respecte a les plantes silvestres. En *Drosophila* també s'ha posat de manifest que la ceramida indueix l'apoptosi, com veurem més endavant.

Els efectes de l'increment de ceramida en resposta a l'estrès i el seu paper en la inducció i en la regulació de l'apoptosi han estat establerts en nombrosos sistemes i organismes. D'altra banda, actualment es comencen a discernir algunes de les dianes sobre les quals actua la ceramida i les proteïnes efectores que en medien l'acció apoptòtica (FIGURA 3). Entre aquestes cal destacar les fosfatases de serina i treonina activades per ceramida, PP1 i PP2A, que en resposta a ceramida desfosforil·len substrats proteics relacionats amb l'apoptosi com c-Jun, Bcl-2, Akt/PKB, pRb, PKC, i les proteïnes SR (Pettus et al. 2002).

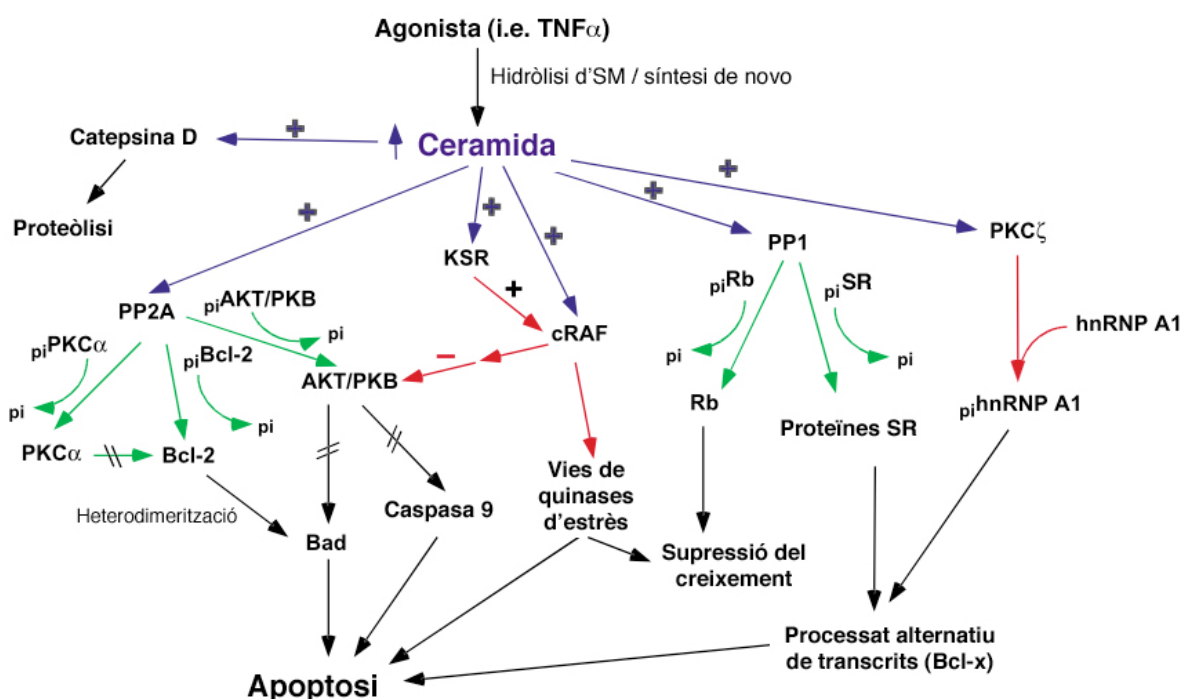


Figura 3 | Vies efectores d'apoptosi regulades per ceramida. Adaptada de (Pettus et al. 2002).

Una altra diana de la ceramida és la proteïna quinasa activada per ceramida (CAPK), també anomenada proteïna quinasa suppressora de Ras (KSR), la qual connecta l'increment de ceramida en resposta a TNF i IL-1 amb les cascades de MAP quinases. També hi ha evidències de la interacció de Raf amb ceramida i de la seva possible activació en la senyalització dependent de ceramida desencadenada per IL-1. De la mateixa manera, s'ha demostrat la interacció de PKC ζ amb la ceramida en resposta a TNF, i que la ceramida augmenta l'activitat de PKC ζ sobre un regulador dels mecanismes de processament alternatiu de transcrits com és hnRNP A1. Així, es comença a discernir que l'increment de ceramida pot regular l'empalmament alternatiu de diversos gens, com veurem més endavant. Finalment, també s'ha establert que la catepsina D, una proteasa lisosomal, és una diana de la ceramida. Aquesta proteïna ha estat implicada en la degradació de proteïnes i en l'apoptosi, procés en el qual sembla que actua per sobre de l'activació de les caspases 3 i 9 i de l'alliberament

del citocrom c, i en resposta a citoquines (TNF , IFN!), agents quimioterapèutics, i a la privació de sèrum (Pettus et al. 2002).

Els enzims del metabolisme de la ceramida mostren localitzacions subcel·lulars diverses, i generalment la diversitat d'efectes atribuïts a aquesta base esfingoide responen en gran mesura a aquesta compartimentació (FIGURES 4 I 5). Per exemple, s'ha vist que en alguns tipus cel·lulars l'increment de la concentració intracel·lular de ceramida no és suficient per a desencadenar el procés apoptòtic i que, probablement, en funció del compartiment cel·lular en el qual es produeix aquest increment –reticle endoplasmàtic-Golgi, membrana plasmàtica, nucli, lisosomes, mitocondris–, els efectes que desencadena són diferents. Això il·lustra la necessitat que tenim d'entendre el metabolisme de la ceramida en el seu context subcel·lular per tal de discernir quins efectes té la ceramida generada en diferents compartiments.

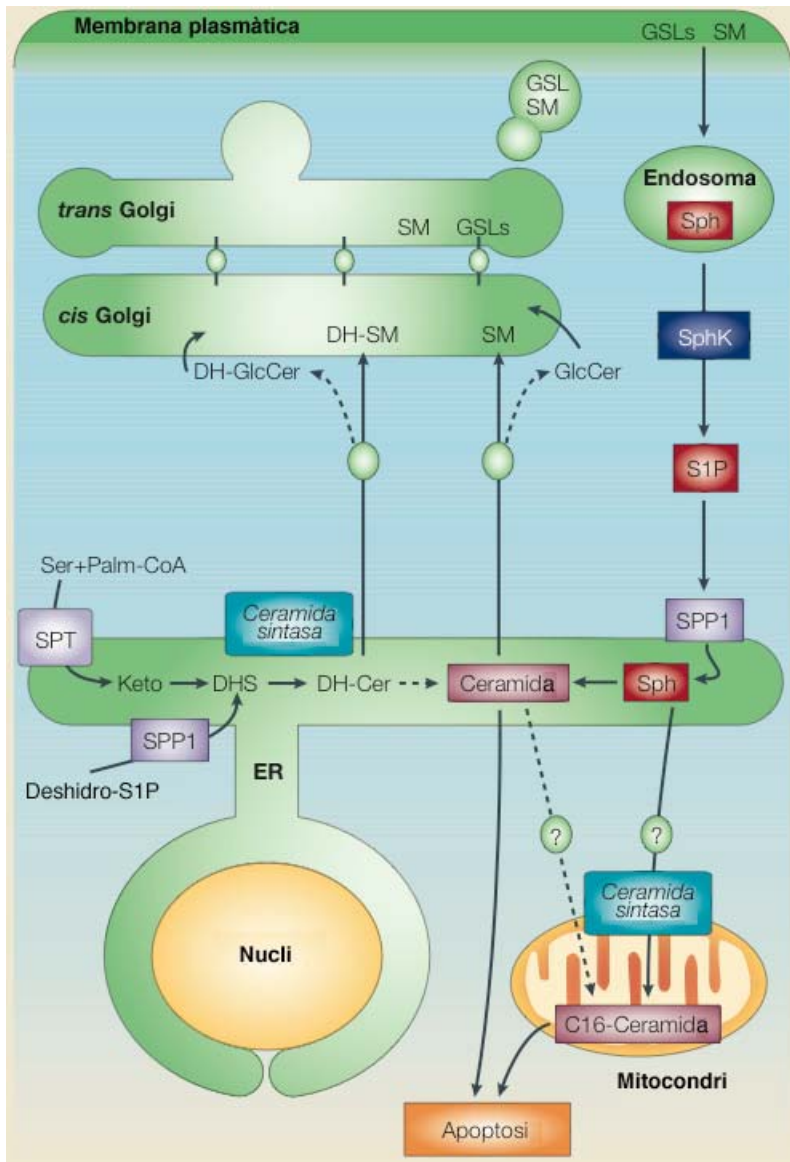


Figura 4 | Localització dels enzims principals del metabolisme de la ceramida. La ceramida és sintetitzada de novo a la cara citosòlica de la membrana del reticle endoplasmàtic. És creu que la síntesi de novo també té lloc al Golgi, les membranes associades als mitocondris i a la membrana mitocondrial (Hannun i Obeid 2002). La deshidroceraida (DH-Cer) i la ceramida produïdes són translocades al Golgi, on seran transformades a deshidroesfingomielina (DH-SM) i esfingomielina (SM) per una esfingomielina sintasa, a la cara luminal, o a deshidroglucosilceramida (DH-GlcCer) i glucosilceramida (GlcCer) per una glucosilceramida sintasa, a la cara citosòlica. Recentment, s'ha descobert un mecanisme de tràfic de ceramida que no depèn de vesícules i que és mediat per una proteïna anomenada CERT, proteïna transportadora de ceramida (Hanada et al. 2003). Al lumen del Golgi, la glucosilceramida esdevé el precursor dels glicoesfingolípid complexos (GSLs). Abreviatures: DHS, deshidroesfingosina. ER, reticle endoplasmàtic. Keto, 3-ceto-esfingianina. Sph, esfingosina. SphK, esfingosina quinasa. SPP1, esfingosina-1-fosfat fosfatasa. SPT, serina palmitoil CoA transferasa. S1P, esfingosina-1-fosfat. Adaptada de (Spiegel et al. 2003).

En aquest context, nivells elevats de ceramida han estat implicats en la patogènesi de diverses malalties neurodegeneratives. Dos dels vuit gens causals de la malaltia de Batten (*CLN1* i *CLN3*) codifiquen proteïnes que modulen els nivells de ceramida en resposta a estímuls apoptòtics. La MALALTIA DE BATTEN, també anomenada LIPOFUSCINOSI CEROÏDE NEURONAL JUVENIL, és una malaltia neurodegenerativa d'herència autosòmica recessiva que es manifesta entre els 5 i 6 anys i que es caracteritza per una ceguesa progressiva, contraccions mioclòniques, i un deteriorament cognitiu i motor que condueix a una mort prematura en la tercera dècada de vida. Es produeix una pèrdua massiva de neurones corticals –així com de cèl·lules granulars i de Purkinje, al cerebel– i de fotoreceptors a la retina que es tradueix en atròfia cerebral i retinitis pigmentària. Tant la mort de les neurones corticals com la dels fotoreceptors és de tipus apoptòtic (Puranam et al. 1999). El gen *CLN1* codifica l'enzim palmitoil tioesterasa [PPT1; OMIM#256730]. La sobreexpressió de *PPT1* en cèl·lules de neuroblastoma les fa resistents a l'apoptosi i, en canvi, el tractament amb oligonucleòtids antisentit les sensibilitza a l'apoptosi induïda per ceramida (Luberto et al. 2002). L'altre gen de la malaltia de Batten relacionat amb les vies d'esfingolípids és *CLN3* [OMIM#204200]. En aquest cas, la sobreexpressió del gen en cèl·lules neuronals evita la inhibició del creixement provocada per la privació de sèrum i la mort cel·lular induïda per diversos agents apoptòtics (Puranam et al. 1999).

La NEUROPATIA SENSORIAL HEREDITÀRIA DE TIPUS 1 [HSAN1; OMIM#162400] –el desordre hereditari més freqüent que afecta les neurones sensorials perifèriques– ha estat atribuïda a mutacions en el gen de la subunitat LCB1 de la serina palmitoil CoA transferasa (*STPLC1*), el primer enzim de la via de síntesi *de novo* de ceramida (Dawkins et al. 2001). HSAN1 és una malaltia autosòmica dominant caracteritzada per la degeneració progressiva del gangli dorsal i de les neurones motores. Es manifesta en la segona o tercera dècada de vida amb una pèrdua de sensibilitat als peus i amb un desgast dels músculs distals i feblesa. Les mutacions en *STPLC1* es creu que provoquen un increment de la síntesi *de novo* de ceramida i que, per tant, fan més sensibles a l'apoptosi les neurones sensorials afectades en la patologia (Dawkins et al. 2001).

La MALALTIA DE FARBER o LIPOGRANULOMATOSI DISSEMINADA [OMIM+228000] és una síndrome molt severa, d'herència autosòmica recessiva, causada per la deficiència de ceramidasa àcida lisosomal. Com a conseqüència d'aquest defecte genètic, es produeix una acumulació de ceramida als teixits, que provoca des dels primers mesos de vida l'aparició de tumefaccions poliarticulars que limiten la mobilitat, irritabilitat, un retard psicomotor greu i la mort de l'infant cap als 2 anys de vida.

L'alteració d'altres enzims del metabolisme dels esfingolípids i, en especial, de les vies de degradació lisosomal dels glicoesfingolípids complexos, afecten de manera severa la funció neuronal i causen diversos desordres coneguts, en conjunt, com a malalties d'acumulació lisosomal o esfingolipidosis. Algunes de les esfingolipidosis més conegudes són la MALALTIA DE NIEMANN-PICK A I B [OMIM#257200 i #607616], en la qual s'acumula esfingomielina com a conseqüència de la deficiència en esfingomielinasa àcida; la MALALTIA DE GAUCHER [OMIM#230800], causada per l'acumulació intracel·lular de glucosilceramida en diversos òrgans i teixits per una manca d'activitat glucocerebrosidasa; i la MALALTIA DE TAY-SACHS [OMIM#272800], en la qual la deficiència en -

hexosaminidasa A provoca l'acumulació del gangliòsid GM2 (Buccoliero et al. 2002, Buccoliero i Futerman 2003).

En totes aquestes malalties, la disrupció de l'equilibri precís que existeix entre les diverses espècies d'esfingolípids compromet enormement la funció neuronal. A la llum de totes aquestes dades, en aquest treball de tesi doctoral, proposem que determinades alteracions del metabolisme dels esfingolípids poden comprometre seriosament la funció retinal i arrossegar els fotoreceptors cap a un procés apoptòtic. En aquest sentit, l'alteració del gen *CERKL*, que codifica una proteïna homòloga a la ceramida quinasa, constitueix la primera observació que connecta el metabolisme dels esfingolípids amb la mort cel·lular apoptòtica observada en la retinitis pigmentària. A continuació, ennumerarem les evidències a favor de la implicació de les vies d'esfingolípids en la patogènesi retinal:

- (1) La natura neuronal dels fotoreceptors suggereix que el metabolisme i les vies regulades per esfingolípids són essencials per a la supervivència d'aquestes cèl·lules, com s'ha determinat en tants altres tipus cel·lulars neuronals, que són especialment sensibles als efectes d'una alteració en aquestes vies.
- (2) El fenotip de la malaltia de Batten presenta, entre les seves manifestacions neurològiques, una mort massiva de neurones fotoreceptores i l'establiment de retinitis pigmentària.
- (3) En una línia cel·lular d'epiteli pigmentari, l'estrès oxidatiu induït per peròxid d'hidrògen o tri-butil-hidroperòxid provoca una apoptosi extensiva i un increment de la producció de ceramida, que podria ser la causa de la mort cel·lular observada (Barak et al. 2001).
- (4) En *Drosophila*, s'ha demostrat l'existència de nivells elevats de ceramida en determinats mutants caracteritzats per l'apoptosi dels fotoreceptors. En aquest mutants, si a través de la sobreexpressió de l'enzim ceramidasa es modulen els nivells de ceramida es produeix un rescat de la degeneració dels fotoreceptors observada (Acharya et al. 2003).

D'una banda, el fenotip retinal de la malaltia de Batten i la connexió de dos dels seus gens causals amb les vies d'esfingolípids posen de manifest la rellevància d'aquestes vies als fotoreceptors. De l'altra, els resultats obtinguts a la línia cel·lular d'epiteli pigmentari, deixen entreveure que l'estrès oxidatiu a la retina modifica els nivells de ceramida. En aquest context, la nostra hipòtesi inicial de treball era que la deficiència en *CERKL*, una ceramida quinasa específica de retina o de funció essencial en aquest teixit, deixaria determinats tipus cel·lulars retinals –com per exemple, les neurones ganglionars i els fotoreceptors– sense un mecanisme que, en situació normal, els permetria contrarestar un increment de ceramida produït en una situació d'estrès. La retina és un teixit exposat directament a la llum, i les seves cèl·lules afronten condicions d'estrès oxidatiu i lumínic. Així doncs, un enzim com *CERKL*, que presumptament fosforil·la ceramida i produeix ceramida-1-fosfat, constituiria un mecanisme protector per a la cèl·lula –la qual podria regular-ne l'expressió segons la necessitat que en tingué–. Aquest mecanisme actuaria de manera anàloga al que succeeix als

experimentos de rescat dels mutants de *Drosophila* amb degeneració de fotoreceptors, en els quals la sobreexpressió de ceramidasa redueix els elevats nivells de ceramida que presenten i contraresta l'apoptosi (Acharya et al. 2003). Per demostrar la hipòtesi del paper de CERKL com a mecanisme cel·lular protector contra l'estrès, és essencial que determinem l'activitat enzimàtica de la proteïna i que analitzem si la seva sobreexpressió en cèl·lules en cultiu les protegeix contra els efectes d'agents inductors d'estrès i apoptosi.

Per contra, podem pensar que la funció de CERKL a la retina és més específica que no pas la regulació a grans trets de la resposta a l'estrès i el rol com a mecanisme cel·lular protector. Potser el producte de la reacció enzimàtica en què intervé aquesta proteïna és un lípid –sigui ceramida-1-fosfat o algun altre lípid similar– que participa en una via relacionada amb la fototransducció o amb algun dels altres processos essencials per a la visió que estan alterats en la retinitis pigmentària. Cal remarcar, que els fotoreceptors són un tipus cel·lular en el qual tant la biogènesi de membranes com la regulació dels processos que hi tenen lloc són primordials per al seu funcionament correcte, i que la majoria de proteïnes de la fototransducció i algunes de les que intervenen en el cicle visual són proteïnes de membrana o associades a membrana. Per tot això, una alteració que afecti la producció d'una determinada espècie lipídica pot comportar conseqüències greus per al funcionament d'aquestes cèl·lules. En aquest sentit, recentment es comença a entreveure el paper de lípids, com ara els fosfatidilinositols, en la regulació de determinats aspectes de la fototransducció en *Drosophila*, que també poden ser aplicats als vertebrats. Sembla que aquests compostos intervenen en la regulació del cicle de translocació de la proteïna arrestina, que depèn de les condicions d'il·luminació i que és necessària per a l'extinció del senyal de la rodopsina fotoactivada (Hardie 2003). Potser CERKL intervé específicament en alguna d'aquestes vies, i la seva alteració no suposa la pèrdua d'un mecanisme per fer front a l'apoptosi, sinó més aviat una alteració concreta d'un procés específic de la retina relacionat amb la fototransducció i que, en darrer terme, acaba conduint a la mort dels fotoreceptors.

És CERKL una ceramida quinasa?

En aquest punt de la discussió, en el qual ja hem apuntat la possible relació entre l'alteració del gen *CERKL* i el paper que pot jugar l'apoptosi regulada per ceramida i altres esfingolípidis en la patogènia retinal, ens hem de plantejar si realment la proteïna CERKL és una ceramida quinasa. A favor de la seva adscripció a aquest grup d'enzims hi ha les dades derivades de la seqüència aminoacídica de la proteïna. Encara que CERKL i CERK presenten una identitat aminoacídica baixa, del 29%, la conservació de posicions assoleix el 50% –calculades amb BLASTP emprant la matriu BLOSUM62– i CERKL conté diversos motius compartits amb enzims de la família de les quinases lipídiques, com ara un domini diacilglicerol quinasa (DAGK) i un lloc d'unió per al donador de fosfat idèntic al de les ceramida quinases (GGDGS) i divergent del de les esfingosina quinases (SGDGL) o del de les diacilglicerol quinases, les altres dues subfamílies de quinases lipídiques. D'altra banda,

l'anàlisi filogenètica de diverses seqüències de quinases lipídiques demostra que aquestes es distribueixen en les tres subfamílies esmentades i que CERKL forma part de la subfamília de ceramida quinases [VEGEU FIGURA 3 EN (TUSON ET AL. 2004)].

Tot i que CERKL comparteix la màxima identitat aminoacídica amb CERK, el gen que la codifica no és una mera còpia duplicada del gen de la ceramida quinasa –no és una CERK2–, i cal considerar aquest fet a l'hora d'analitzar la funció de la proteïna. De l'escrutini de les bases de dades, se'n deriva que CERKL és probablement una invenció dels vertebrats. Actualment, sabem de l'existència d'aquest gen únicament en mamífers (humans i ratolí) i aus (pollastre), i tenim evidències de seqüències homòlogues als genomes dels peixos que ja han estat seqüenciats.

Els invertebrats presenten un únic gen de la subfamília de les ceramida quinases, i probablement a partir d'aquest, mitjançant un fenomen de duplicació gènica a la base del llinatge dels vertebrats, van aparèixer els dos gens que reconeixem actualment. Curiosament, aquestes dues còpies en vertebrats no comparteixen el mateix grau d'identitat aminoacídica que presenta el parell d'esfingosina quinases (SPHK1 i SPHK2). Aquestes darreres presenten una identitat del 48.6% (TAULA 1), que suggereix o bé que la duplicació que va donar lloc a CERK i CERKL és més antiga, o que els gens de la ceramida quinasa i la ceramida quinasa-like han divergit de manera més accelerada. Encara que CERK presenta el mateix grau d'identitat aminoacídica amb CERKL que amb SPHK1 i SPHK2, la presència de determinades seqüències conservades entre CERKL i CERK –divergents en SPHK1/2– indica, inequívocament, que aquestes dues proteïnes provenen d'un gen ancestral comú. En aquest context, l'escenari més plausible seria considerar la incorporació accelerada de canvis a la seqüència de CERKL condicionada, potser, per una especialització funcional de la proteïna.

Taula 1

Percentatge d'identitat/similitud entre ceramida i esfingosina quinases

	CERKL	CERK	SPHK1	SPHK2
CERKL	–			
CERK	30.1 / 71.3	–		
SPHK1	25.3 / 64.5	29.0 / 73.8	–	
SPHK2	25.0 / 66.0	30.3 / 71.9	48.6 / 79.9	–

Els percentatges d'identitat / similitud han estat obtinguts a partir d'un càlcul de similitud local realitzat a partir d'una matriu PAM250 amb el programa Sequence Analysis. Informagen, Inc (v1.3.1)

Les dades experimentals preliminars de què disposem, derivades de la sobreexpressió de CERKL en cèl·lules HEK293 en cultiu i de l'anàlisi de l'activitat quinasa de l'extracte proteic total contra substrats lipídics, posen de manifest que CERKL no fosforil·la ceramida ni tampoc diacilglicerol. Com podem explicar, doncs, la discrepància aparent entre les dades derivades de l'anàlisi de la seqüència de la proteïna i els assajos d'activitat *in vitro*?

Primer de tot, com hem apuntat, podem hipotetitzar que, tot i la identitat en determinats residus i l'homologia de dominis, la divergència de seqüència entre CERKL i CERK reflecteix una especialització funcional real: potser CERK ha conservat una funció més basal com a ceramida quinasa i CERKL ha patit un procés de subfuncionalització i realitza part de les funcions de la ceramida quinasa en un teixit concret (la retina) –en concert amb altres proteïnes– o bé, s'ha especialitzat funcionalment i catalitza la fosforil·lació d'un altre tipus de substrat lipídic. Podem imaginar aquest substrat com una espècie d'esfingolípíd desconeguda o bé, una variant específica de retina d'algun dels esfingolípíds que coneixem. Enzims com la ceramida quinasa tenen una certa especificitat de substrat, com és lògic, per a ceramides fisiològiques de cadena llarga –*D-erythro*-ceramida–, formades per una base esfingoïde amb una cadena de 18 carbonis i un àcid gras de 14 a 26 carbonis, i no pas per als anàlegs de cadena curta, que són igualment fosforil·lats, però amb afinitats més baixes. Ara bé, si es tractés d'això, CERKL probablement hauria de tenir una certa afinitat pel tipus de substrats genèrics emprats als assajos descrits en aquesta tesi, els quals inclouen ceramides provinents del cervell boví.

També podem pensar que la no detecció de l'activitat quinàsica de CERKL respon a algun dels paràmetres de l'assaig experimental. Alguns dels factors que podrien ser crítics en l'assaig són el requeriment iònic i el de pH de l'enzim. L'assaig *in vitro* d'activitat quinasa, que s'exposa al tercer dels treballs inclosos en aquesta tesi, s'ajusta als paràmetres de pH i d'espècies iòniques òptims descrits a la literatura per a mesurar les activitats ceramida i diacilglicerol quinasa. Podria ser, però, que la proteïna CERKL estigués localitzada en un compartiment cel·lular amb un pH diferent del fisiològic, com per exemple als lisosomes o en certs endosomes, i que, per això, no poguéssim determinar-ne l'activitat al pH emprat. Aquesta darrera suposició, però, és discordant amb les dades obtingudes als assajos de localització subcel·lular.

En aquests assajos, s'observa que CERKL presenta una localització dinàmica formada per dos patrons bàsics que coexisteixen en una mateixa preparació; un dels quals, eminentment nuclear, compatible amb el d'altres enzims del metabolisme de la ceramida que també es localitzen al nucli per a modificar la ceramida que allà s'hi produeix i que pot jugar un paper important en l'apoptosi i en la regulació del processat alternatiu de transcrits de determinats gens. L'altre patró bàsic és citosòlic, amb certs acúmuls adjacents a diversos compartiments membranosos en els quals té lloc el metabolisme de la ceramida, com ara el reticle endoplasmàtic, el Golgi o els mitocondris. Aquest segon patró concordaria amb les descripcions de les ceramida quinases com a enzims associats a membrana, i presenta una similitud remarcable amb els patrons que s'obtenen en immunolocalitzar ceramida amb un anticòs específic.

Creiem que la variabilitat observada en la localització subcel·lular de CERKL posa de manifest que aquesta depèn de la situació en què es troba la cèl·lula i que, segons els estímuls que rep, CERKL es concentraria en un determinat compartiment, com, per exemple, el nucli en una situació d'increment de la ceramida nuclear. En absència d'estímuls, en canvi, la distribució de la proteïna seria més difosa. En aquest sentit, s'ha demostrat que diverses proteïnes que intraccionen

amb lípids a través de dominis d'homologia a Plekstrina (PH), i que presenten una distribució citosòlica, són direccionades a la membrana en resposta a estímuls concrets (Irvine 2004), i no seria d'estranyar que la localització de CERKL segueixi un mecanisme similar. Per estudiar aquesta qüestió, seria interessant realitzar assajos de localització subcel·lular havent tractat prèviament les cèl·lules amb agents que incrementin determinats *pools* de ceramida, i veure si aquests modifiquen la localització de la proteïna.

La funció dels enzims del metabolisme dels esfingolípidis està íntimament relacionada amb la seva localització i, probablement, CERKL no n'és una excepció. Tant la localització correcta de la proteïna al compartiment en el qual acompleix la seva funció, com les proteïnes acompanyats o els substrats que allà hi troba són factors crítics. En aquest sentit, podríem pensar que l'extracte proteic total emprat als assajos *in vitro* no conté aquestes hipotètiques proteïnes acompanyants necessàries per a la funció de CERKL i potser seria convenient adaptar els assajos a un sistema el més similar possible a la retina, o bé, realitzar assajos *in vivo* en una línia cel·lular retinal. En darrer terme, la construcció del model animal de la malaltia, a partir de la generació d'una soca de ratolí *knockout* pel gen *Cerkl*, serà de gran utilitat per a valorar *in vivo* tots aquests aspectes i discernir quina és la funció específica de CERKL a la retina. L'escrutini de les interaccions moleculars de CERKL amb d'altres proteïnes, i en concret amb proteïnes retinals, també resoldrà els interrogants plantejats pel fet que l'alteració d'aquest gen causa exclusivament degeneració retinal i no afecta altres teixits.

Processat alternatiu de transcrits i malaltia

Inicialment, vam descriure que el gen *CERKL* presentava una estructura formada per 13 exons distribuïts en un segment genòmic de 118.8 kb entre els gens *ITGA4* i *NEUROD1*. Recentment, l'escrutini en detall per PCR dels transcrits del gen a partir de cDNA total de retina humana ens ha permès caracteritzar un nou exó diferencial situat entre els exons 4 i 5 descrits inicialment –que hem anomenat 4b– i definir el patró d'empalmament alternatiu del gen *CERKL* a la retina humana, en la qual hem identificat diversos transcrits que codifiquen diferents isoformes de la proteïna CERKL. La presència d'alguns d'aquests transcrits diferencials també ha pogut ser demostrada en altres teixits humans tant de l'adult com fetals, però únicament a la retina hem pogut determinar la coexistència dels 4 transcrits principals del gen (isoformes a–c).

En un treball recent, realitzat mitjançant microarrays per a visualitzar el processat en les junctures exó/exó de 10000 gens humans multiexònics i en el qual es van analitzar mostres de 52 teixits i línies cel·lulars diferents, es va posar de manifest que els transcrits d'almenys el 74% dels gens multiexònics són processats alternativament (Johnson et al. 2003). El tipus d'empalmament pel qual s'inclouen (o se salten) un o més exons és una de les formes usuals de processat alternatiu (García-Blanco et al. 2004). Aquest és el tipus de processat alternatiu que hem descrit en el gen *CERKL*, en el qual trobem dos transcrits que divergeixen en la inclusió/exclusió de l'exó diferencial 4b, i quatre transcrits curts, en els quals es produeix l'exclusió dels exons 3,4, 4b i 5 (isoforma c); 4,

4b i 5 (isoforma d); 3,4 i 4b (isoforma e); i 4 i 4b (isoforma f). Aquests dos darrers patrons de processat alternatiu (isoformes e i f) produeixen uns transcrits amb un canvi a la pauta de lectura del gen, que condueix a la inclusió de codons de terminació prematurs.

En un estudi recent, s'ha demostrat que el 35% dels transcrits alternatius analitzats contenen codons de terminació prematurs (Lewis et al. 2003). La inclusió o exclusió errònia d'exons als transcrits –com en el cas de les isoformes e i f de *CERKL*–, que es produeix probablement per la flexibilitat del mateix mecanisme que permet l'empalmament alternatiu, i que condueix a canvis a la pauta de lectura i a codons de terminació prematurs, sembla que podria estar connectada amb el procés de degradació d'mRNA induïda per codons de terminació (*non-sense-mediated decay*, NMD). Així, l'NMD actuarà com a mecanisme d'eliminació dels transcrits erronis que es produeixen en el procés d'empalmament alternatiu (García-Blanco et al. 2004). L'NMD elimina els transcrits que contenen codons de terminació situats més de 50 nucleòtids 5' enllà de l'última unió exó/exó. En aquest sentit, pot semblar que les cèl·lules produeixen un nombre desmesurat de transcrits alternatius irrellevants, però alguns autors han suggerit que la producció d'aquests transcrits, que contenen codons de terminació prematurs i que són eliminats per l'NMD, constitueix un mecanisme de regulació addicional dels nivells d'expressió proteica (Lewis et al. 2003).

El processat alternatiu de transcrits del gen *CERKL* se circumscriu en el grup d'exons que van d'E3 a E5. D'aquests, E4b i E5 presenten *gc* en el lloc 3' (donador) d'empalmament, en comptes del *gt* canònic. El parell donador-acceptor no canònic *gc-ag* constitueix el 0.69% dels introns dels mamífers, respecte al parell canònic *gt-ag* que representa el 99.24% (Burset et al. 2000). Donada la baixa freqüència del parell *gc-ag*, és remarcable que el gen *CERKL* en contingui dos, i que aquests coincideixin amb els exons alternatius. Els llocs d'empalmament no canònics són freqüents en gens com els de les immunoglobulines, els quals presenten fenòmens de reordenament gènic. És temptador establir un paral·lelisme entre aquest fet i l'empalmament a nivell de transcrit, i hipotetitzar que, en el gen *CERKL*, la presència d'aquests senyals intervindria en la regulació de l'empalmament alternatiu dels exons E4b i E5. En aquest sentit, la presència de *gc* podria permetre una unió més laxa dels factors d'empalmament, necessària per a l'alternativitat en la inclusió o exclusió d'aquests exons en el transcrit madur. La seqüència consens que flanqueja a 3' el lloc donador *gc* comparteix els llocs conservats amb la que flanqueja el *gt*, però està molt més conservada en el primer cas (Burset 2002). Aquesta major conservació probablement permet la unió del mateix tipus de maquinària d'empalmament en els dos casos (Burset 2002), i el lleuger desparellament produït per la presència del *gc* introdueix l'especificitat necessària per a que es produeixi l'empalmament alternatiu.

El significat funcional dels transcrits diferencials que no contenen codons de terminació prematurs és divers segons el gen que considerem. En alguns casos, com en el de la proteïna Bcl-x, les isoformes generades a partir dels transcrits diferencials tenen efectes oposats –Bcl-x(L) és un factor antiapoptòtic i Bcl-x(S) pot tenir efectes apoptòtics (Boise et al. 1993)–. En d'altres, una de les isoformes alternatives és única d'un teixit en concret en el qual desenvoluparà una funció específica (Valero et al. 2001). En d'altres casos, com el que ens ocupa, la natura funcional de cadascuna de les

isoformes és encara desconeguda i necessitarà de l'adscripció funcional prèvia de CERKL abans de determinar l'especificitat que confereixen a aquesta funció cadascuna de les isoformes de la proteïna.

Tot i això, des d'un punt de vista funcional, és atractiu poder especular sobre les isoformes alternatives de CERKL, produïdes per l'empalmament alternatiu del grup d'exons E3-E5. Per què són necessàries diverses isoformes de la proteïna? D'una banda, hem de suposar que les dues isoformes llargues, que contenen el domini diacilglicerol quinasa (DAGK), són les isoformes que tenen activitat catalítica. La inclusió de l'exó 4b, en la isoforma més llarga de la proteïna (isoforma b, 558 aminoàcids), pot modular d'alguna manera la funció catalítica de CERKL, probablement modificant (incrementant, o bé, disminuint) l'afinitat de la proteïna per al donador de fosfat, necessari per a l'activitat quinàsica, ja que aquest exó diferencial introdueix 26 aminoàcids immediatament abans del lloc d'unió per al donador de fosfat (GGDGS). D'altra banda, és remarcable que l'empalmament alternatiu que dona lloc a les isoformes curtes, c i d, elimini quasi perfectament el domini DAGK. Aquest fet suggereix un paper regulador, en comptes de catalític, per a les isoformes curtes de CERKL, ja que, probablement, no tenen activitat quinàsica. Les isoformes c i d presenten, però, el presumpte domini d'homologia a Plekstrina (PH), un domini d'interacció amb determinades espècies de fosfatidilinositols de membrana. Podem especular que, a través de la interacció amb lípids com aquests, les isoformes curtes modularien la unió de les isoformes llargues a determinats dominis de la membrana, impedit-n'hi la unió. Els dos tipus d'isoformes tindrien capacitat d'unir-se a aquests dominis de la membrana, però la unió de les isoformes curtes, que no tendrien activitat catalítica, atenuaria els efectes de les isoformes llargues. Aquest mecanisme seria anàleg al que succeeix en la regulació del senyal de determinats receptors de membrana, en els quals l'existència de "receptors solubles" actua com a mecanisme modulador de la unió del lligand (Kawano i Kypta 2003). Tornant a CERKL, el balanç en la síntesi d'un i altre tipus d'isoformes podria estar regulat per estímuls i/o determinades condicions cel·lulars. Finalment, la unió de la proteïna a determinats dominis de la membrana podria respondre també a aquests estímuls, com passa en altres casos, en els quals intervé l'ancoratge de proteïnes a la membrana a través de dominis tipus PH.

L'exó E5 –en el qual, els pacients de les famílies P2 i E1 presenten la mutació R257X– és present únicament en les isoformes llargues. Les curtes no l'inclouen i, per tant, aquestes isoformes escapen els efectes de la mutació. Aquest fet l'hem de situar en el context que la retina és l'únic teixit en el qual hem trobat les 4 isoformes. En d'altres teixits, detectem per RT-PCR la presència d'alguna de les isoformes llargues. El fet que en cap d'aquests teixits la mutació R257X causi patologia, ens pot fer pensar que la presència de les isoformes llargues en determinats teixits extraretinals és irrellevant. Potser, tot i detectar-les, no s'expressen en nivells adients per a la seva funció, o potser és necessària la presència de les quatre isoformes per a la funció de CERKL. Alternativament, la retina podria ser l'únic teixit que conté altres proteïnes que intervenen en la funció de CERKL, o l'únic on hi ha el substrat d'aquest enzim. Per contra, podríem suposar que, als teixits extraretinals, s'expressa alguna altra proteïna, per exemple CERK, que pot complementar els efectes de la mutació, i que no ho fa a la retina. A banda de caracteritzar la funció de CERKL, seria interessant estudiar si CERK s'expressa a la

retina. En el cas que no s'expressés, constituïria un element per entendre perquè la mutació de *CERKL* afecta exclusivament la retina. En el cas que s'expressés, caldria determinar quins papers juguen ambdues a la biologia retinal i si participen en el mateix procés.

També és suggerent pensar en una regulació de l'empalmament alternatiu del gen exclusiva de la retina, en la qual es produiria un balanç entre les formes llargues i curtes que dependria de determinats estímuls o factors. Per exemple, podríem pensar, de manera similar al que succeeix en el cas de la proteïna Bcl-x, que determinades formes d'empalmament del gen (les llargues) són antiapoptòtiques –a la llum d'una funció de *CERKL* com a ceramida quinasa-like– i que la seva producció depèn d'estímuls d'estrès cel·lular que inicieixen d'alguna manera sobre la maquinària d'emplament alternatiu per a decantar la balança cap a unes isoformes en concret. En aquest sentit, podríem pensar en l'efecte de l'estrès lumínic al qual estan sotmesos els fotoreceptors i en mecanismes per fer-hi front, en els qual podríen estar-hi implicades determinades isoformes de la proteïna *CERKL*. Per valorar aquesta qüestió, seria interessant modelitzar aquestes condicions d'estrès en el ratolí i analitzar si es produeix un increment de l'expressió del gen, així com, específicament, d'alguna de les isoformes.

La inclusió d'exons com E4b o E5 específicament als transcrits retinals –per exemple, per acció de l'estrès lumínic– és una hipòtesi atractiva que lliga amb la patogènia de la malaltia, que se circumscriu exclusivament a la retina. En altres gens causals retinitis, que també s'expressen en teixits extraretinals, ha estat documentada la presència d'exons específics als transcrits retinals. Aquest és el cas del gen *RPGR*, que codifica una proteïna reguladora de GTPasa i que és responsable de la major part dels casos de retinitis pigmentària lligada al sexe. *RPGR* presenta un transcrit diferencial a la retina, amb un exó 15 més llarg (ORF15), en el qual es concentren la major part de les mutacions causals de retinitis. De manera anàloga, podríem pensar en l'E5 de *CERKL* com una mena d'exó diferencial de retina, que alhora és un punt calent mutacional. Aquesta darrera qüestió necessitarà de la descoberta de noves famílies lligades a RP26 i de la caracterització de noves mutacions en el gen *CERKL* per tal de discernir si aquestes es concentren en l'E5 i en exons alternatius propers, o bé, es distribueixen uniformement per tot el gen. En el cas en què una mutació de terminació, com la identificada en aquest treball, se situés fora dels exons alternatius i afectés les 4 isoformes, caldria esperar efectes fenotípics en altres teixits en els quals només s'expressen les formes curtes i, potser llavors, el fenotip clínic seria de retinitis sindròmica en comptes de retinitis aïllada. Per il·lustrar aquest concepte, el gen *RPGR* torna a ser útil, ja que determinades mutacions del gen s'associen a formes de retinitis pigmentària que inclouen l'afectació d'altres teixits, i en canvi, d'altres mutacions cursen exclusivament a la retina.

La descoberta de transcrits diferencials en un nombre creixent de gens remarca la importància del seu coneixement quan considerem aquests gens com a candidats. En l'estudi de malalties monogèniques, s'analitza usualment la regió codificant dels gens i, després de seqüenciar-ne tots els exons, si no hem identificat cap canvi presumptament patogènic, el gen en qüestió és descartat com a responsable de la malaltia. Ara bé, es podria donar el cas –i aquí torna a ser útil

l'exemple del gen *RPCR*— que no haguéssim tingut en compte un, o diversos, exons diferencials, i que haguéssim descartat el gen sense haver-los analitzat. En d'altres casos, si els gens que analitzem en una regió candidata no han estat caracteritzats en profunditat, podem trobar-nos amb què la seva estructura genètica no és definitiva i que resten per identificar exons no diferencials, com va succeir en aquest treball amb els gens *UBE2E3* i *SSFA2*.

L'alteració dels processos d'empalmament alternatiu s'ha vist que té un paper important en la patogènia molecular de diverses malalties hereditàries humanes. Del total de mutacions humanes indexades a la base de dades *Human Gene Mutation Database*, gairebé un 10% afecten llocs d'empalmament dels gens. Tot i això, aquestes no inclouen les mutacions localitzades en altres seqüències en *cis* reguladores (estimuladores o silenciadores) de l'empalmament alternatiu, ni tampoc les mutacions en gens que codifiquen factors reguladors en *trans*. En aquest sentit, cal destacar les mutacions en els gens de tres factors de processat de premRNA (*PRPF8*, *PRPF31*, i *HPRP3*) que causen retinitis pigmentària autosòmica dominant en diverses famílies. Les mutacions d'aquests gens d'expressió ubíqua, que únicament causen degeneració retinal, ens poden fer hipotetitzar que el funcionament correcte dels mecanismes de processat alternatiu de transcrits és d'importància vital a la retina, i que l'alteració d'aquest factors afavoriria el processat aberrant de gens "retinals", la funció dels qual és necessària per a la viabilitat dels fotoreceptors. En aquest sentit, gens implicats en distròfies retinals i que, com *CERKL*, presenten un patró d'empalmament alternatiu podrien ser dianes d'aquests factors de processat de premRNA.

Actualment, es comencen a entreveure nivells addicionals de regulació de l'empalmament alternatiu de diversos gens que posen de relleu la complexitat cada cop més evident d'aquests processos. En concret, i com apuntàvem anteriorment, el processat alternatiu de transcrits del gen *Bcl-x* produeix dues formes alternatives que donen lloc a dues proteïnes amb funcions contràries en el procés d'inducció de l'apoptosi: la forma *Bcl-x(s)*, proapoptòtica, i la forma *Bcl-x(L)*, antiapoptòtica. De manera anàloga el processat alternatiu de transcrits del gen de la caspasa 9, també genera una forma proapoptòtica i una d'antiapoptòtica (caspasa 9b). El balanç entre les formes alternatives dels dos gens respon a ceramida. Així, la ceramida decanta la balança cap a la producció de les formes proapoptòtiques, *Bcl-x(s)* i caspasa 9 (Chalfant et al. 2002). Recentement, s'han descobert dos elements en *cis* que regulen l'empalmament alternatiu de *Bcl-x* i que responen a ceramida (Massiello et al. 2004). Així doncs, la relació entre el processat alternatiu de transcrits i la inducció i regulació de l'apoptosi per ceramida es postula com una futura diana d'intervenció farmacològica per a fer front a malalties caracteritzades per l'apoptosi de determinades poblacions cel·lulars.

La retinitis pigmentària a l'era genòmica

Com hem vist a la introducció, resten quatre loci de retinitis pigmentària autosòmica recessiva i tres loci de lligada a l'X, en els quals encara no s'ha identificat el gen responsable. La seqüenciació

del genoma humà obre la porta a la caracterització de les alteracions subjacents en aquests casos i en moltes altres famílies afectades de retinitis pigmentària que no cosegüen amb cap dels loci que coneixem actualment. Després de gairebé dues dècades en les quals s'han descobert molts dels gens causals de la retinitis pigmentària, l'horitzó que s'ha anat perfilant és molt més complex del que es podia suposar inicialment: no s'ha trobat cap gen major de la malaltia –que expliqui una gran part dels casos– i el que han anat apareixent són gens cadascun dels quals causa la malaltia en un conjunt més o menys reduït de famílies. El gen *CERKL*, identificat en aquest treball de tesi, encaixa perfectament en aquest escenari perquè, de moment, s'ha caracteritzat com a gen responsable de la malaltia en dues famílies, i això representa una prevalença, en població espanyola, equivalent a la dels altres gens RP descrits.

Així doncs, la seqüència completa del genoma humà és una valuosa font d'informació tant per a la cerca dels gens dels loci de retinitis orfes com per a la dels gens causals en aquest 50% de famílies que alguns autors apunten que encara resten per caracteritzar. Ara bé, la seqüència del genoma *per se* no és d'utilitat si no és en combinació amb una sèrie de desenvolupaments tècnics que, en el present i en el futur immediat, esdevindran eines essencials de la recerca en retinitis pigmentària a l'era genòmica.

D'entrada, els estudis de lligament genètic que ens han de conduir a la definició de nous loci i a la identificació de nous gens RP ja comencen a fer ús de la gran quantitat de marcadors indexats en les bases de dades d'SNP, en combinació amb la tecnologia més moderna que permet l'anàlisi d'un gran nombre de marcadors de manera automatitzada. En aquest sentit, cal remarcar que el darrer gen causal d'un tipus sever de distròfia retinal infantil, *RDH12*, ha estat identificat a partir d'un mapatge d'homozigositat genòmic realitzat mitjançant un xip que permet l'anàlisi de 10000 SNP distribuïts arreu del genoma (Affymetrix GeneChip Human Mapping 10K Array version Xba131), i que va conduir els autors del treball a la definició d'una regió crítica de 2.86 Mb a 14q23.3–q24.1 en tan sols 3 dies de feina (Janecke et al. 2004). El disseny de xips específics per analitzar la cosegregació amb gens de retinitis ja caracteritzats obrirà la porta, en un futur immediat, a l'automatització del diagnòstic i de la recerca.

Un cop definida una regió del genoma candidata, la seqüència completa de la regió i les bases de dades de gens coneguts, mRNA caracteritzats i transcrits parcials, així com els algorismes de predicció de gens, ens permeten disposar del llistat dels gens continguts en aquesta regió. En aquest punt, podem analitzar si algun dels gens de la regió definida per lligament ha estat proposat com a gen candidat de distròfia retinal en algun dels nombrosos treballs que han aparegut els darrers temps. Entre aquests estudis, encaminats a la caracterització de nous gens candidats per a patologies retinals, n'hi ha d'exclusivament experimentals, que fan ús de tècniques com l'anàlisi seriada de l'expressió gènica, SAGE (Sharon et al. 2002), o d'altres que combinen estratègies *in silico* i *in vitro* (Lord-Grignon et al. 2004) i que, generalment, tenen com a objectiu la identificació de gens que s'expressen de manera diferencial a la retina. En un d'aquests darrers estudis s'han localitzat 27 EST de ratolí que s'expressen molt a la retina i a l'epiteli pigmentari i que corresponen a gens no

caracteritzats (Lord-Grignon et al. 2004). Dos dels gens mapen en loci de retinitis ja caracteritzats (RP22 i RP28), i el primer d'aquests codifica una proteïna que té homologia amb el gen *CLN8*, un dels gens responsables de la malaltia de Batten, en la qual, com hem vist, es produeix la mort dels fotoreceptors. Per tant, als propers anys, de la conjunció d'aproximacions genòmiques com l'anterior i de l'estudi en profunditat de cadascun dels gens identificats, podrem veure com s'identifiquen nous gens responsables de distròfies retinals.

El gran nombre de gens candidats que aquest tipus d'estudis i d'aproximacions genòmiques generen necessita de mètodes d'anàlisi de la funció gènica que puguin ser d'aplicació a un organisme model i que ens permetin establir la rellevància funcional de cadascun d'aquests gens a la retina. Recentment, ha estat dissenyat un mètode ràpid d'electroporació que permet l'anàlisi del guany de funció d'un gen –amb vectors d'expressió– i de la pèrdua de funció –mitjançant la interferència d'RNA– tant *in vivo* com *in vitro* a la retina del ratolí, amb el qual han pogut reproduir els efectes observats als ratolins transgènics o *knockout* generats per tècniques més costoses (Matsuda i Cepko 2004).

En un altre estudi recent, s'ha generat un transcriptoma no redundat de la retina (*retinoma*) compost per 13037 transcrits, mitjançant les dades derivades dels nombrosos estudis de transcriptòmica retinal (Schulz et al. 2004). En aquest treball, s'ha analitzat la distribució dels gens *retinals* al llarg dels cromosomes i s'ha determinat que no és aleatòria. Els gens *retinals* s'agrupen en determinades localitzacions cromosòmiques que, probablement, reflecteixen una regulació comuna *in cis* de l'expressió. D'ara endavant, treballs com aquest poden facilitar la caracterització de nous gens candidats i la definició de regions del genoma riques en gens *retinals* en les quals podrien estar situats els gens responsables de la retinitis pigmentària i d'altres distròfies retinals que encara no han estat identificats.

Totes aquestes noves aproximacions enriqueixen encara més el còrpus metodològic existent per a l'estudi de la retina dels mamífers, la qual continua sent un excel·lent sistema model per tal de poder estudiar la biologia neuronal en la salut i en la malaltia.