# Resultats

All experimentalists know you have to do an experiment four times.
The first one is a complete mess and shows only a hint that it might have worked.
The second one is better but still messy. Then you do it the third time for the book.
This is when you forget to add a reagent, or mix up the tubes or the centrifuge leaks.
That is why there is always a fourth time.

**Sydney Brenner**

# CAPÍTOL 1

## Caracterització funcional d'un nou gen del locus RP26: el gen *ORMDL1* i la família gènica *ORMDL*

Publicació a la qual ha donat lloc aquest treball:

ORMDL proteins are a conserved new family of endoplasmic reticulum membrane proteins. *Genome Biology* 3 (6):research0027.1–0027.16 (2002).

L'objectiu principal d'aquest treball de tesi doctoral ha estat la cerca del gen responsable de retinitis pigmentària autosòmica recessiva en el locus RP26. En iniciar-se aquesta tesi, l'any 2000, tant les eines de què disposàvem com les dimensions de la regió del cromosoma 2, en la qual estava situat el locus RP26, feien inabordable una cerca exhaustiva al llarg de tot l'interval de cosegregació. Les dades derivades del projecte de seqüenciació del genoma humà eren encara escadusseres i es concentraven en altres autosomes, com ara el 22 i el 21, més petits i d'interès mèdic evident.

El tipus d'estratègies que podíem aplicar a la cerca del gen del locus RP26 se centraven en la cerca i anàlisi de gens candidats funcionals. Abans que s'iniciés aquest treball de tesi doctoral, al nostre grup, es va dur a terme una cerca *in silico* d'EST (*Expressed Sequence Tags*) de retina –mapades a la regió del cromosoma 2 que estàvem estudiant– per tal d'identificar gens del locus que s'expressessin a la retina i que, per tant, fossin bons candidats. Una de les EST identificades –l'EST de retina H86933–, traduïda conceptualment, mostrava un alt grau d'identitat aminoacídica amb dues pautes de lectura oberta (ORF) de llevat, que corresponien a dos gens no caracteritzats funcionalment. A partir d'una sonda derivada de l'EST, es va realitzar el triatge d'una genoteca de retina, de la qual es va poder aïllar un cDNA de 1092 bp, que contenia sencera la regió codificant d'un nou gen humà. Aquest nou gen del locus RP26 presentava una pauta de lectura oberta de 462 bp i codificava una proteïna de 153 aminoàcids que no contenia dominis funcionals coneguts i de la qual, per tant, no en podíem predir la funció ni el procés cel·lular en el qual intervenia. La cerca d'homologies en les bases de dades de seqüència havia posar de manifest l'existència de dues seqüències en llevat similars a aquest nou gen humà, una de les quals havia estat designada amb el nom *ORM1*. A partir d'aquesta denominació, i seguint les directrius del Comitè de nomenclatura de la Organització del Genoma humà (HUGO), vam denominar aquest nou gen humà *ORMDL* [*ORM1 (Saccharomyces cerevisiae)-like gene*]. La seqüenciació dels exons del gen *ORMDL*, en la família P2, no va revelar cap canvi patogènic en la regió codificant del gen, que va ser descartat com a gen *RP26*.

Cerques posteriors en les bases de dades, ens van permetre determinar que *ORMDL* formava part d'una família gènica molt conservada en eucariotes, des de llevat fins a mamífers. Diversos dels llinatges considerats contenien còpies duplicades d'aquests gens. En vertebrats, vam caracteritzar tres gens: el ja descrit –que va passar a anomenar-se *ORMDL1*– i dues còpies addicionals, *ORMDL2* i *ORMDL3*, que en humans es troben als cromosomes 12 i 17, respectivament, en regions cromosòmiques producte, probablement, de les grans duplicacions segmentals (o bé de tot el genoma) que van tenir lloc en la base del llinatge dels vertebrats. Tot i que *ORMDL1* havia estat descartat com a gen *RP26*, la no adscripció de funció a cap dels membres d'aquesta família gènica i el seu alt grau de conservació al llarg de l'escala evolutiva eren elements atractius que ens van portar a abordar la caracterització funcional de la família de gens *ORMDL*.

En aquest sentit, en aquest primer capítol de resultats s'exposen les dades obtingudes en la caracterització dels tres gens *ORMDL* en diversos llinatges: en humans, l'estudi del patró d'expressió i la localització subcel·lular de les proteïnes codificades; a *Drosophila*, la caracterització del gen

*DORMDL* i la seva expressió a l'embrió i als discos imaginals; i en llevat, la generació dels *knockout* senzills dels gen *ORM1* i *ORM2* i del doble mutant, i la caracterització fenotípica d'aquests tres mutants. També s'estudia la relació dels diversos membres d'aquesta família mitjançant una anàlisi comparativa i filogenètica de les seqüències aminoacídiques de les proteïnes ORMDL en diversos llinatges eucariòtics, incloses les plantes. Finalment, a partir de les dades experimentals obtingudes es proposa una funció pressumptiva per a aquesta família proteica. Tots aquests resultats van donar lloc a una publicació, i és en aquest format que s'exposen a continuació. La meva contribució a aquest treball queda reflectida en les FIGURES 2, 3, 5A, 6, 7A-F, 8, i en les TAULES 2 I 3.

Research

# ORMDL proteins are a conserved new family of endoplasmic reticulum membrane proteins

Lars Hjelmqvist*†‡ Miquel Tuson*‡ Gemma Marfany*, Enric Herrero§, Susana Balcells* and Roser Gonzàlez-Duarte*

Addresses: *Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain. §Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina, Universitat de Lleida, 25198 Lleida, Spain. †Current address: Karolinska Institutet, MBB, Chemistry I, S-171 77 Stockholm, Sweden. ‡Both authors contributed equally to this work.

Correspondence: Roser Gonzàlez-Duarte. E-mail: roser@bio.ub.es

Published: 14 May 2002

*Genome* ***Biology*** 2002, **3(6)**:research0027.1–0027.16

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/6/research/0027


(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 10 January 2002
Revised: 12 March 2002
Accepted: 10 April 2002

## Abstract

**Background:** Annotations of completely sequenced genomes reveal that nearly half of the genes identified are of unknown function, and that some belong to uncharacterized gene families. To help resolve such issues, information can be obtained from the comparative analysis of homologous genes in model organisms.

**Results:** While characterizing genes from the retinitis pigmentosa locus *RP26* at 2q31-q33, we have identified a new gene, *ORMDL1*, that belongs to a novel gene family comprising three genes in humans (*ORMDL1*, *ORMDL2* and *ORMDL3*), and homologs in yeast, microsporidia, plants, *Drosophila*, urochordates and vertebrates. The human genes are expressed ubiquitously in adult and fetal tissues. The *Drosophila ORMDL* homolog is also expressed throughout embryonic and larval stages, particularly in ectodermally derived tissues. The *ORMDL* genes encode transmembrane proteins anchored in the endoplasmic reticulum (ER). Double knockout of the two *Saccharomyces cerevisiae* homologs leads to decreased growth rate and greater sensitivity to tunicamycin and dithiothreitol. Yeast mutants can be rescued by human ORMDL homologs.

**Conclusions:** From protein sequence comparisons we have defined a novel gene family, not previously recognized because of the absence of a characterized functional signature. The sequence conservation of this family from yeast to vertebrates, the maintenance of duplicate copies in different lineages, the ubiquitous pattern of expression in human and *Drosophila*, the partial functional redundancy of the yeast homologs and phenotypic rescue by the human homologs, strongly support functional conservation. Subcellular localization and the response of yeast mutants to specific agents point to the involvement of ORMDL in protein folding in the ER.

## Background

The human genome project has generated raw information on an increasing number of novel genes and gene families whose function is still unknown. Positional cloning and large-scale genome analysis allow preliminary functional assignment of human genes on the basis of linkage to genetic diseases and reported information from model organisms. Although the available computational tools may fail to provide clear functional clues, they are still of great value in defining structural domains, pinpointing intra- and inter-specific sequence homologies and establishing new gene families. In the human genome, a mutational approach to

characterizing genes functionally is limited to patients that carry well characterized disease alleles. On the other hand, the availability of the mouse genome sequence is providing new tools for systematic functional characterization. This approach has already been used in yeast by the European Functional Analysis Network (EUROFAN) and has provided functional insights on evolutionarily conserved genes.

We previously reported linkage of autosomal recessive retinitis pigmentosa (*RP26*) to chromosome 2q31-33, between markers D2S148 and D2S117 [1]. Subsequent analyses narrowed the locus to the D2S350-D2S161 interval. This chromosomal region, (approximately 8 million base pairs (Mbp)) is characterized by a low percentage of G+C - basically made of L1+L2 isochores - and by sparse gene content [2]. As no obvious positional candidates were found among known genes, an expressed sequence tag (EST) database search was undertaken to identify sequences expressed in the retina. One of the identified ESTs was homologous with two yeast open reading frames (ORFs), two predicted genes in *Arabidopsis* and one in *Drosophila* and other human genomic and EST sequences. After full cDNA analyses and characterization of the corresponding genomic regions, a functional approach was undertaken. We report here a new evolutionarily conserved gene family, named *ORMDL,* for *ORM1* (*Saccharomyces cerevisiae*)-like genes, following the HUGO Gene Nomenclature Committee guidelines, and present what could be called a 'vertical' genomics approach in several model organisms. This comprises subcellular localization of the encoded proteins, expression analyses on human tissues and *Drosophila* embryos, and single and double yeast knockouts.

## Results

### Characterization of the full-length human *ORMDL1* cDNA

A human retinal cDNA library was screened using a 647 base pair (bp) probe containing the WI-18706 STS (located at the *RP26* locus, see Materials and methods). A total of 13 positive clones were isolated, subcloned in pBluescript II KS(+), and sequenced (Figure 1). Eight of the clones contained an apparently complete ORF, and the other five were truncated. The 5′ and 3′ ends of the messages were verified by rapid amplification of cDNA ends (RACE) using placental RNA as template. In the 5′ experiment, two extended products were detected with the same 5′ end but a differentially spliced 110 bp non-coding exon. The longer RACE product started 175 bp upstream of a putative initiation codon and this 5′ untranslated region (5′-UTR) contained two in-frame stop codons. The shorter RACE product did not contain an in-frame stop codon. In the 3′ experiment, a single extension product was detected which contained a polyadenylation signal (ATTAAA) situated 24 nucleotides 5′ of the poly(A) tail. Some of the cDNA clones had an extended 3′-UTR which could be the result of the use of different polyadenylation signals further downstream. The full-length cDNA (1,092 bp) contained an ORF consisting of 462 bp, from nucleotides 176 to 637. The deduced protein chain consisted of 153 amino acids with an estimated molecular mass of 17.4 kDa.

### Characterization of *ORMDL1* homologs cDNAs

When searching the nucleotide databases with the full-length human *ORMDL1* cDNA, human homologous EST sequences were identified which belonged to two separate UniGene clusters (Hs.13144 and Hs.293711). Corresponding IMAGE cDNA clones were obtained and sequenced. The deduced ORFs (denoted *ORMDL2* and *ORMDL3*) had the same size as *ORMDL1*. Comparison of the proteins showed between 80% and 84% positional identities (Table 1), and 116 out of 153 amino-acid residues were conserved between the three sequences. Moreover, in 26 of the 37 remaining positions the substitutions are conservative.

No homologous sequences were identified in *Drosophila* EST databases. Screening of an adult *Drosophila melanogaster* cDNA library using the human *ORMDL1* cDNA as a heterologous probe was not successful either. We then designed a homologous probe based on the *Drosophila* genomic high-throughput sequence. Five positive clones were isolated and sequenced. Although none of the clones was full-length, one covered more than 80% of the ORF. The complete ORF could then be deduced by overlapping this sequence to the genomic data. The conceptual translated sequence was one amino acid longer at the amino terminus than the human forms and shared between 48% and 50% residue identities. Subsequently, in the completed *Drosophila* Genome Project, this sequence has appeared annotated as *predicted gene CG14577* [3].

### *ORMDL1* belongs to an evolutionarily conserved gene family

The *ORMDL1* ORF was used for database sequence comparisons using the National Center for Biotechnology Information (NCBI) BLAST server [4,5] (tBLASTN against nucleotide databases, and BLASTP, PSI-BLAST and PHI-BLAST against the non-redundant protein database). Homologous protein sequences were found in yeast, microsporidia (*Encephalitozoon cuniculi*, an opportunistic pathogen in AIDS), plants (including *Arabidopsis*), invertebrates (*Drosophila*), urochordates (*Ciona intestinalis*) and vertebrates. The search for distant relatives using PSI-BLAST and PHI-BLAST programs and for short, nearly exact, matches (BLASTP) did not yield any additional homologs. In vertebrates, three different genes were distinguished, corresponding to *ORMDL1*, *ORMDL2* and *ORMDL3*. *S. cerevisiae* and *A. thaliana* showed two copies each, while a single copy was found in *C. intestinalis, D. melanogaster*, *Saccharomyces monacensis, Schizosaccharomyces pombe* and *E. cuniculi*. Table 2 lists most of the genes, with their corresponding annotations in the public databases, if available, and summarizes other relevant features.

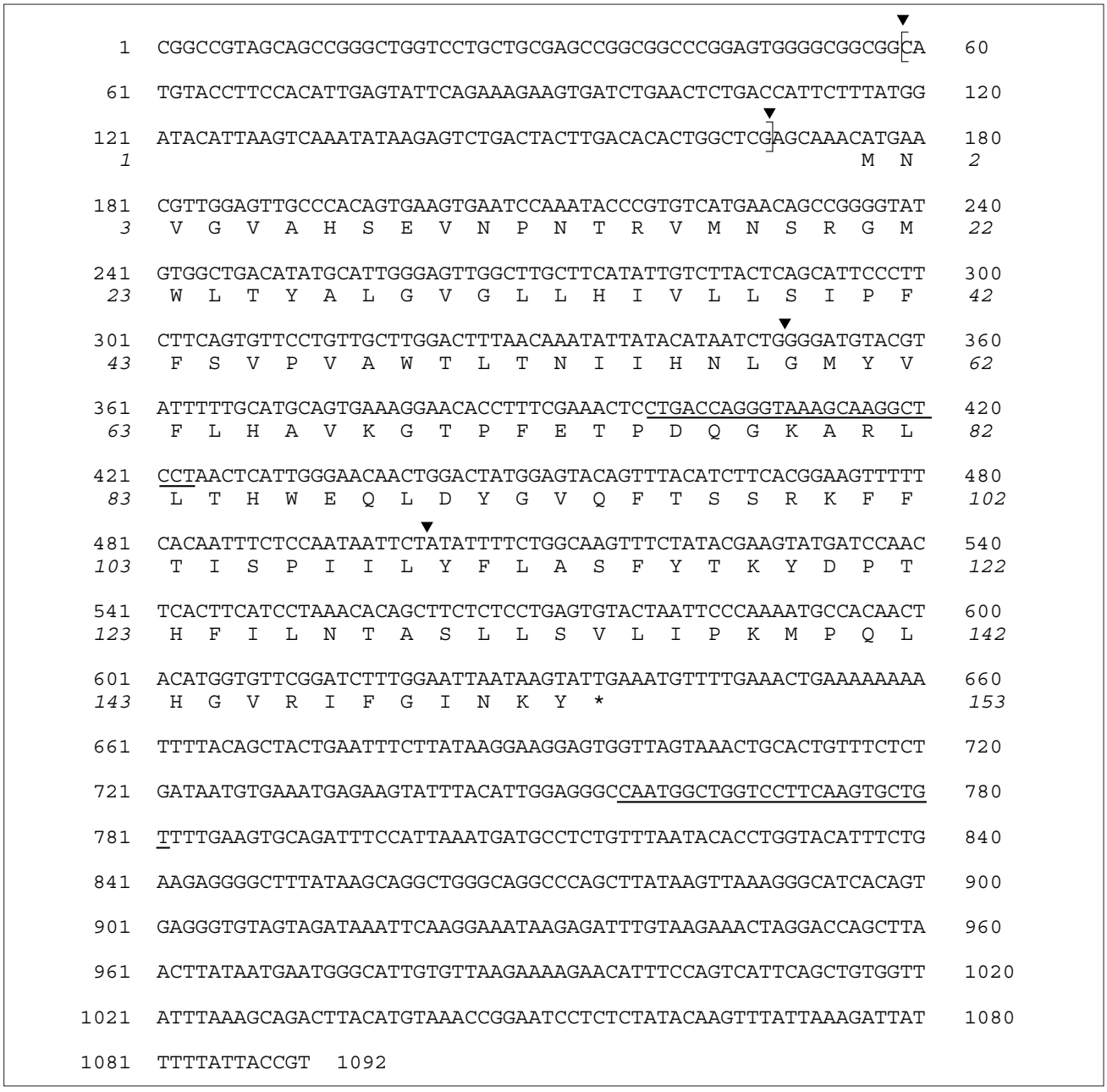



**Figure 1**
Nucleotide sequence of the *ORMDL1* cDNA. The translation is shown below. Intron positions are marked with black triangles. The exon shown between square brackets in the 5′-UTR is alternatively spliced. Underlines mark the positions of the primers used for the RACE experiments.

When all the protein sequences were aligned using CLUSTALW, several conserved domains became evident (Figure 2). In addition, all yeast proteins were longer, at both the amino- and carboxy-terminal ends. Pairwise comparisons revealed a very high level of identity (Table 1). Searches for characterized functional domains did not reveal any significant homology. However, four putative transmembrane domains were detected when considering the protein alignment of all known members of the family. Similar results were obtained when analyzing the ORMDL1 sequence on its own (Figure 2).

A phylogenetic tree using all the available protein sequences was drawn with CLUSTALW and the bootstrap values were also calculated (Figure 3). According to this tree, each human *ORMDL* gene has an extremely conserved

**Table 1**

**Percentage identity between members of the ORMDL family**

| | HsapORMDL1 | MmusORMDL1 | HsapORMDL2 | MmusORMDL2 | HsapORMDL3 | MmusORMDL3 | DmelORMDL | AthaORMDLa | AthaORMDLb | ScerORM1 | SmonORM1 | ScerORM2 | SpomORM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HsapORMDL1 | - | | | | | | | | | | | | |
| MmusORMDL1 | 99 | - | | | | | | | | | | | |
| HsapORMDL2 | 83 | 82 | - | | | | | | | | | | |
| MmusORMDL2 | 83 | 82 | 97 | - | | | | | | | | | |
| HsapORMDL3 | 84 | 83 | 80 | 82 | - | | | | | | | | |
| MmusORMDL3 | 84 | 84 | 81 | 83 | 96 | - | | | | | | | |
| DmelORMDL | 48 | 48 | 50 | 50 | 50 | 50 | - | | | | | | |
| AthaORMDLa | 40 | 41 | 39 | 39 | 41 | 41 | 35 | - | | | | | |
| AthaORMDLb | 39 | 39 | 38 | 38 | 39 | 39 | 37 | 81 | - | | | | |
| ScerORM1 | 32 | 32 | 32 | 32 | 34 | 34 | 32 | 31 | 32 | - | | | |
| SmonORM1 | 33 | 33 | 34 | 33 | 36 | 36 | 33 | 32 | 32 | 92 | - | | |
| ScerORM2 | 31 | 31 | 33 | 34 | 34 | 34 | 31 | 29 | 30 | 68 | 68 | - | |
| SpomORM | 39 | 39 | 39 | 40 | 41 | 41 | 35 | 31 | 34 | 48 | 49 | 46 | - |

The calculations are based on the alignment shown in Figure 2.

counterpart in all vertebrates analyzed. Vertebrate sequences cluster together in a separate branch. Similarly, yeast, microsporidian and plant ORMDLs group in well differentiated lineages.

### Genomic structure of the *ORMDL* sequences

The full-length human *ORMDL1* cDNA clone was compared to non-redundant DNA databases and the 5′ end was found to be homologous to the promoter region of the *PMS1* gene [6]. The P1 genomic clone 1670, containing this region, was analyzed by restriction mapping and Southern blot with the *ORMDL1* cDNA as a probe. Positive fragments were subcloned and sequenced. The nucleotide sequence contained 4.8 kilobases (kb) of the upstream region, all of the five exons, and the four introns (Figure 4). The third intron was not fully sequenced and the size was determined by restriction mapping after amplification by polymerase chain reaction (PCR). All introns were flanked by the consensus donor and acceptor splice sites. No putative TATA box was found on the upstream genomic sequence, although several Sp1-binding sites were detected using the TESS server [7]. Using the EMBL CpG Islands service [8] a presumptive 0.9 kb CpG island containing 0.4 kb of the upstream region, the first exon of the 5′-UTR, and parts of the first intron was detected.

Database searches against releases of the public Human Genome Project [4] confirmed the genomic *ORMDL1* structure. Furthermore, the genomic structures of the coding region of the two human paralogs *ORMDL2* and *ORMDL3* and the *D. melanogaster* homolog were deduced after alignment of the characterized human and *Drosophila* cDNAs with genomic sequence databases. For the *A. thaliana ORMDL* homolog, tBLASTN searches were carried out with the human ORMDL proteins against the *Arabidopsis* genome. The data obtained were contrasted with gene predictions from the *Arabidopsis* Genome Initiative (actually, one of these predictions has already been confirmed; GenBank accession no. AF360237). In all these genes a genomic structure of three coding exons is totally conserved, including the positions of the exon-intron boundaries (Figure 4). In contrast, the yeast homologs were each encoded in one single continuous ORF. Interestingly, the introns of human *ORMDL2* and *ORMDL3* genes were found to be much smaller than those of *ORMDL1*. The human *ORMDL* homologs mapped to different chromosomes from *ORMDL1* (*ORMDL2* to 12q13.2 and *ORMDL3* to 17q21.1). No further human homologs were detected in the public databanks [4] or in the Celera database [9], except for two putative processed pseudogenes. One of these showed 86% positional identity to *ORMDL1* and mapped to chromosome 10p14, and the other showed 84% identity to *ORMDL2* and mapped to 8q22.1. Both *ΨORMDL1* and *ΨORMDL2* lacked introns and appeared to have acquired several indels (insertions/deletions). In addition, *ΨORMDL2* lacked the first coding exon of *ORMDL2*.

**Table 2**

**Summary of experimental and *in silico* data on the *ORMDL* gene family members**

| Organism | Gene name | cDNA (bp) | Protein (amino acids) | Annotation (from RefSeq, GenBank, TAIR, FlyBase, SGD, YPD, PombePD) | Chromosomal location | Structural domains* | Expression pattern† | Subcellular localization† |
|---|---|---|---|---|---|---|---|---|
| *H. sapiens* | ORMDL1 | 1,092† | 153 | LOC51240 [39] | 2q32.2 | 4 TM | Ubiquitous in adult and fetal tissues | Endoplasmic reticulum |
| | *ORMDL2* | 934† | 153 | HSPC160 [39] | 12q13.2 | 3 TM | Ubiquitous in adult and fetal tissues | Endoplasmic reticulum |
| | *ORMDL3* | 869† | 153 | None | 17q21.1 | 4 TM | Ubiquitous in adult and fetal tissues | Endoplasmic reticulum |
| | *ΨORMDL1* | (pseudogene) | None | 10p14 | - | - | - | |
| | *ΨORMDL2* | (pseudogene) | None | 8q22.1 | - | - | - | |
| *M. musculus* | *ORMDL1* | 1,815 | 153 | None | 1‡ | 4 TM | NT | NT |
| | *ORMDL2* | 1,053 | 153 | 0610012C09 gene [12] | 10‡ | 4 TM | NT | NT |
| | *ORMDL3* | 2,024 | 153 | 2810011N17 gene [12] | 11‡ | 4 TM | NT | NT |
| *T. rubripes* | *ORMDL1* | - | 153 | None | | 4 TM | NT | NT |
| | *ORMDL2* | - | 153 | None | | 4 TM | NT | NT |
| | *ORMDL3* | - | 153 | None | | 4 TM | NT | NT |
| *C. intestinalis* | *ORMDL* | - | 153 | None | | 4 TM | NT | NT |
| *D. melanogaster* | *ORMDL* | 388†, partial cds | 154 | CG14577 [3] | 3L 78E6 | 3 TM | Ubiquitous at early embryonic stages and in ectodermal derived tissues at later stages | NT |
| *A. thaliana* | *ORMDLa* | 795 | 157 | F6F3.4; At1g01230§ | 1 | 2 TM | NT | NT |
| | *ORMDLb* | none | 154 | MJC20.10; At5g42000¶ | 5 | 4 TM | NT | NT |
| *S. cerevisiae* | *ORM1*¥ | none | 222 | ORM1 [40]; G4089; YGR038w | VII | 4 TM | NT | NT |
| | *ORM2*# | none | 216 | YLR350w [41]; L8300.1 | XII | 4 TM | NT | NT |
| *S. monacensis* | *ORM1* | None | 222 | ORM1 [40] | | 4 TM | NT | NT |
| *S. pombe* | *ORM* | 1167, partial cds | 186 | SPBC119.09c [42] | 2 | 3 TM | NT | NT |
| *E. cuniculi* | *ORMDL* | - | 147 | ECU11_1150 [43] | XI | 4 TM | NT | NT |

*Potential transmembrane segments were predicted using the HMMTOP 2.0 program [26]. †Experimental data obtained in the present study. ‡Mapping of mouse *ORMDL* genes was inferred from syntenic regions between human and mouse chromosomes. §GenBank accession no. AF360237. ¶GenBank accession no. NC_003076. ¥The null mutant constructed in the present study is viable; the double null mutant with ORM2 is viable but shows an impaired growth rate†; the *ORM1* mRNA is more abundant in *MAT*$\alpha$ cells than in *MAT***a** cells, and at 39°C than at 30°C [44]; *ORM1* is coregulated with other 117 genes under 26 cell-damaging conditions [45] and is induced during sporulation [46]; Orm1 protein is produced in mid-log cells [47]. #The null mutant constructed in the present study is viable; the double null mutant with YGR038w (*ORM1*) is viable but shows an impaired growth rate†; Orm2 interacts with Slt2p in a systematic two-hybrid assay [21]; *ORM2* is coregulated with other 121 genes under 26 cell-damaging conditions [45] and is induced by a 30 min and 90 min treatment with 1 M NaCl [48]. TM, transmembrane segments; NT, not tested.

### Expression analysis of human *ORMDL* genes

To study the expression pattern of human *ORMDL* genes, RT-PCR and northern analyses were performed. RT-PCR expression analysis on sixteen adult and fetal tissue samples showed a ubiquitous expression pattern for all three human genes (Figure 5a). Northern analysis using full-length as well as 3′-UTR *ORMDL1* cDNA probes identified three transcripts of 1.4, 2.5 and 3.3 kb in most human adult tissues (Figure 5b). The expression of *ORMDL1* was moderately high in pancreas, placenta and brain but low in skeletal muscle and lung.

### Subcellular localization of human ORMDL proteins: expression of ORMDL-EGFP fusion proteins in COS-7 cells

To determine the subcellular localization of the ORMDL proteins, the cDNAs corresponding to the human *ORMDL* genes

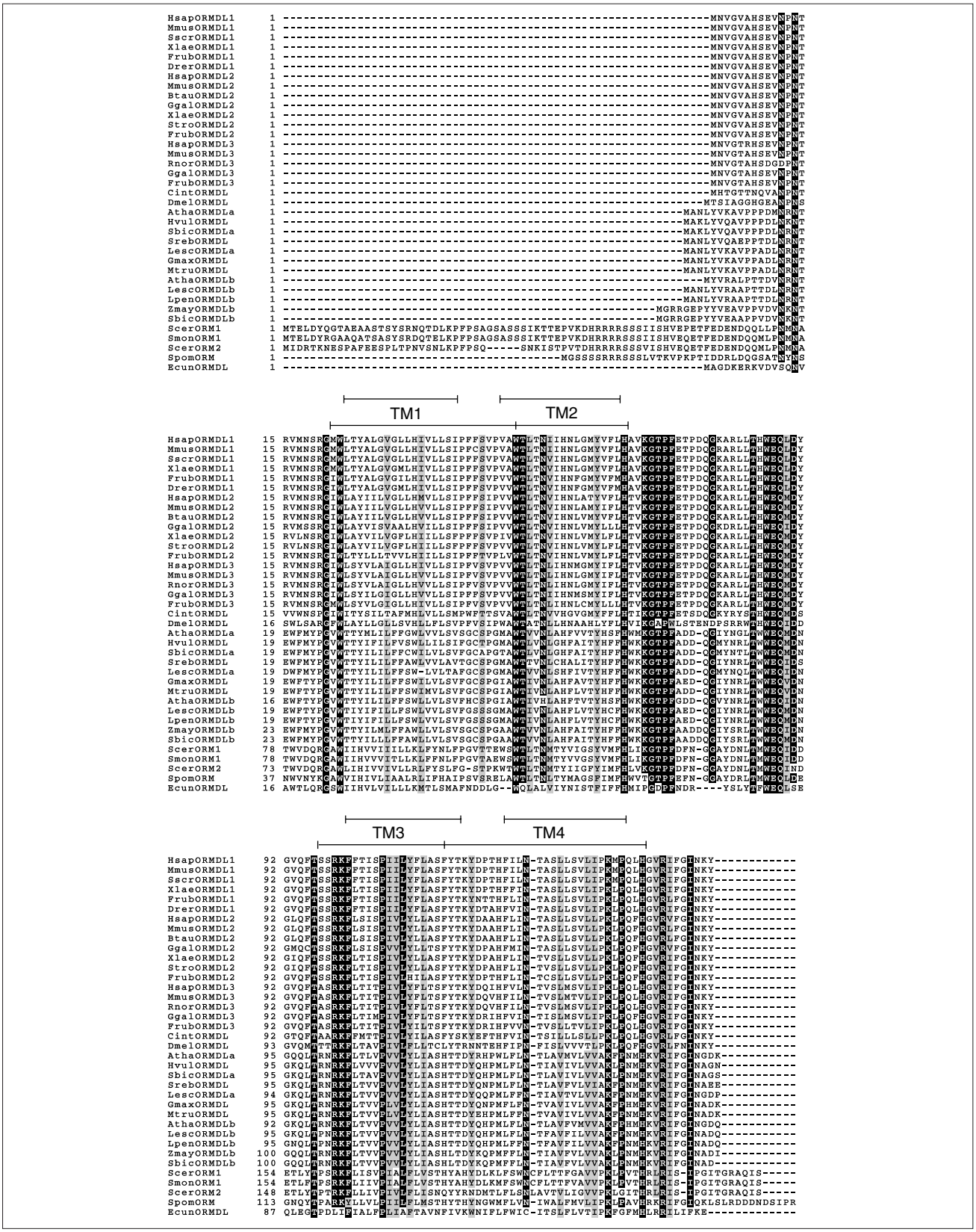



**Figure 2** *(see legend on the next page)*

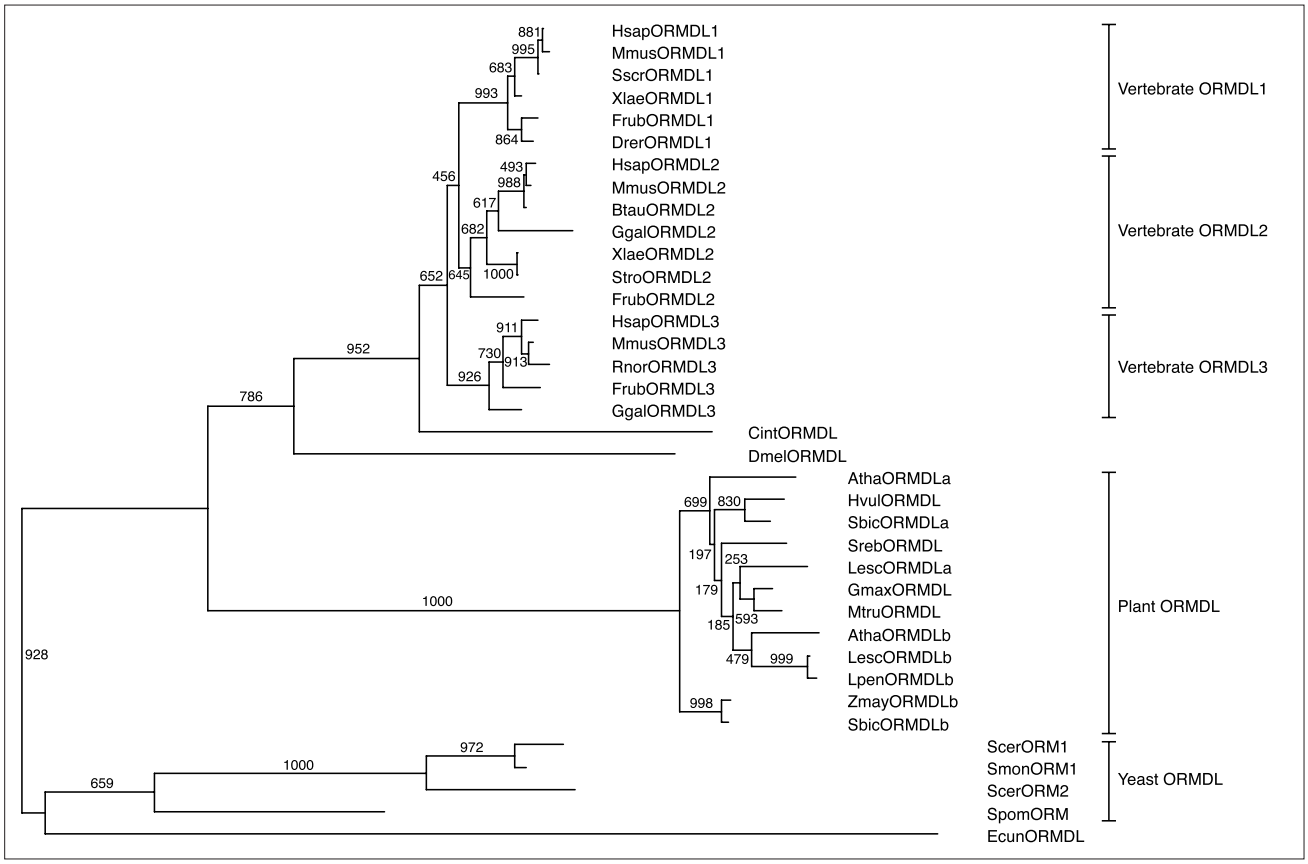



**Figure 3**
Relationships between the ORMDL amino-acid sequences shown as an unrooted phylogenetic tree. The tree was obtained with the program CLUSTALW [23], with distances corrected for multiple substitutions, and positions with gaps excluded. Numbers show results from bootstrap analysis [24]. The branches are labeled in the same manner as for the alignment (Figure 2).

were cloned into pEGFP-derived vectors and transiently expressed in transfected COS-7 cells. Twenty-four hours after transfection, fusion proteins of enhanced green fluorescent protein and ORMDL (EGFP-ORMDL) appeared predominantly in the perinuclear endoplasmic reticulum (ER) and, to a lesser extent, throughout the extended ER network (Figure 6a). Confocal scanning microscopy using an antibody against protein-disulfide isomerase (PDI), an ER-marker protein, showed an overlapping signal in the endoplasmic reticulum (Figure 6a-c). In contrast, double-label fluorescence with biotin-labeled concanavalin A (which shows affinity for terminal α-D-mannosyl and α-D-glucosyl residues, and thus binds to the plasma membrane in non-permeabilized cells), detected with Texas Red-conjugated streptavidin, showed no overlapping signal with EGFP-ORMDL at the plasmatic membrane (Figure 6d-f). The subcellular patterns observed for EGFP-ORMDL1, EGFP-ORMDL2 and EGFP-ORMDL3 fusion proteins were similar, in either fixed (Figure 6g-i, respectively) or *in vivo* cells (Figure 6j,k), co-localizing with an *in vivo* ER marker. Similar results were obtained from COS-7 cells transfected with either pEGFP-N2- or pEGFP-C2-derived constructs. In contrast, EGFP alone was uniformly

**Figure 2** *(see figure on the previous page)*
Alignment of deduced ORMDL amino-acid sequences. Highly conserved positions (≥ 95%) are shown against a black background, whereas those with conservative exchanges are shown against a grey background. Alignment was performed using the CLUSTALW program [23]. Potential transmembrane segments (TM1 to TM4) are marked with bars above the alignment: Upper bars according to HMMTOP [25] and lower bars following TMAP [26]. Species abbreviations are as follows: Hsap, human; Mmus, mouse; Rnor, rat; Sscr, pig; Btau, cow; Ggal, chicken; Xlae, *Xenopus laevis*; Stro, *Silurana tropicalis*; Frub, *Takifugu rubripes* (pufferfish); Drer, *Danio rerio* (zebrafish); Cint, *Ciona intestinalis*; Dmel, *Drosophila melanogaster*; Atha, *Arabidopsis thaliana*; Hvul, *Hordeum vulgare* (barley); Sbic, *Sorghum vulgare*; Sreb, *Stevia rebaudiana*; Lesc, *Lycopersicon esculentum* (tomato); Gmax, *Glycine max* (soybean); Mtru, *Medicago truncatula*; Lpen, *Lycopersicon pennellii*; Zmay, *Zea mays* (maize); Scer, *Saccharomyces cerevisiae*; Smon, *Saccharomyces monacensis*; Spom, *Schizosaccharomyces pombe*; Ecun, *Encephalitozoon cuniculi*.

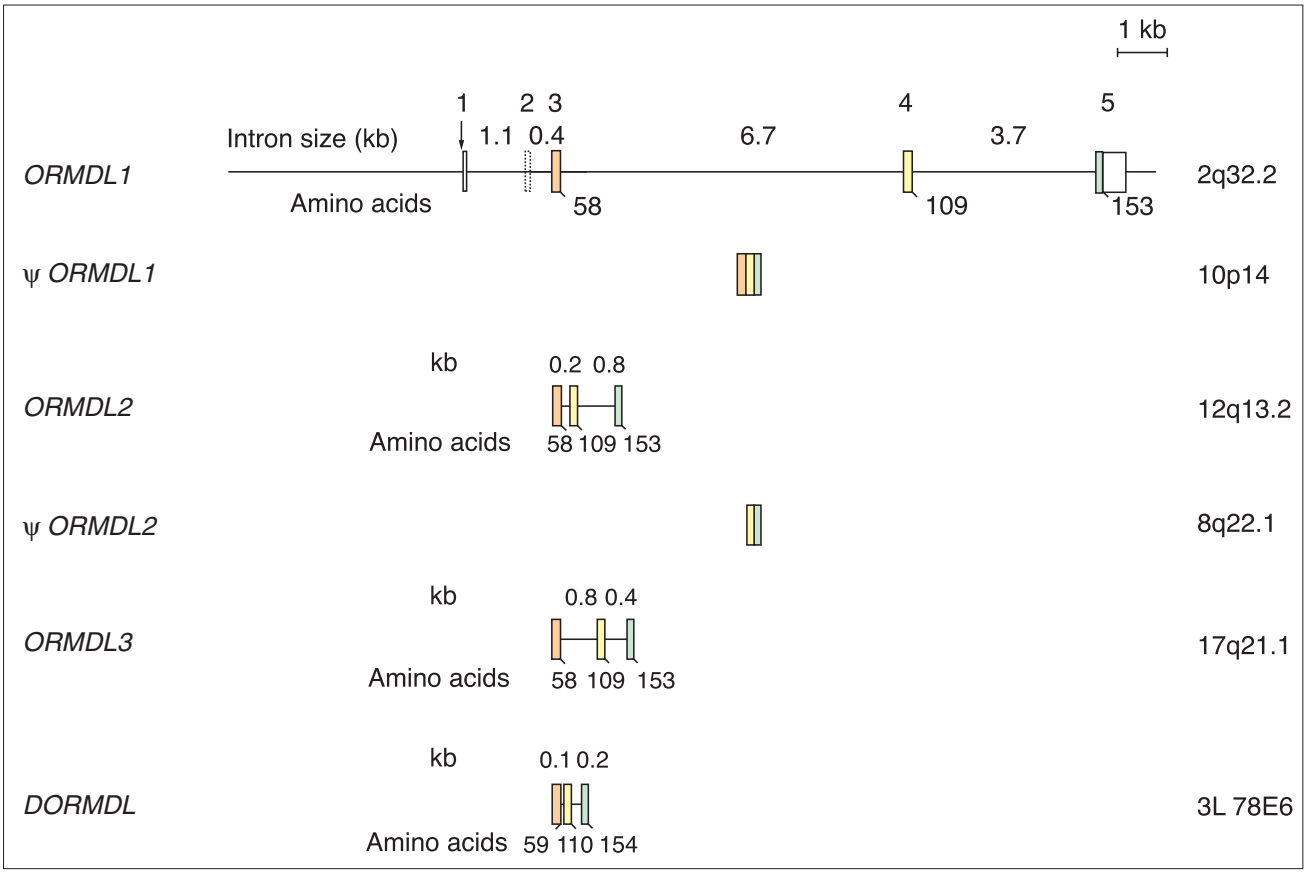



**Figure 4**
Gene organization of human *ORMDL1, ORMDL2, ORMDL3, ΨORMDL1*, *ΨORMDL2,* and *Drosophila ORMDL*. Exon numbers of *ORMDL1* are shown above the bars. Coding exons are shown in color and their sizes (in nucleotides) are: exon 3, 181 (orange); exon 4, 152 (yellow); exon 5, 133 (green). The intron sizes in kilobases are shown in-between the bars. The numbers below the bars denote the amino-acid positions for the exon-intron boundaries, and for the end of the ORF. The dotted appearance of the bar representing exon 2 of *ORMDL1* indicates that it is alternatively spliced. The structure of the two human pseudogenes is also shown. The chromosomal location of each gene or pseudogene is indicated to the right.

distributed throughout the cell, including the nucleus (Figure 6l), in all experiments. Overall, these data show that the ORMDL proteins locate at the ER membrane in agreement with their predicted transmembrane topology.

### *In situ* hybridization in *Drosophila* embryos and larval imaginal discs

We analyzed the expression pattern of the *Drosophila ORMDL* homolog (*DORMDL*) during different developmental stages. To this end, a 391 bp antisense riboprobe containing most of the coding sequence was used for *in situ* hybridization on wild-type Canton S *D. melanogaster* 0 h to 24 h embryos (stages 1 to 17). *DORMDL* was ubiquitously and homogeneously expressed in the syncytial blastoderm and during the cellularization stage (Figure 7a,b). In stages 11-12, the germ-band layer (which later gives rise to all somatic cell lines, including ectoderm, mesoderm and endoderm) gave a positive hybridization signal (Figure 7c,d). In later stages (stage 14), the signal was mainly detected in the ectodermal tissues (Figure 7e). The *DORMDL* riboprobe was also used to analyze the expression pattern in imaginal discs in third-instar larvae. Uniform and homogeneous expression was observed in eye-antenna, leg, wing, (Figure 7g-i) and haltere (data not shown) imaginal discs, which was consistent with the former ectodermal expression detected in late-stage embryos.

### Yeast single and double knockouts: analysis of deletion strains

The *S. cerevisiae* genome contained two members of this family: *ORM1* (ORF YGR038w) at chromosome VII and *ORM2* (ORF YLR350w) at chromosome XII. To provide functional clues for the human homologs, we generated and analyzed single and double knockouts. We used the W303-1A and W303-1B yeast wild-type strains (see genotypes in Materials and methods) and the *kanMX4* cassette for gene disruption [10]. Single gene targetings were assessed by designing specific PCR reactions. The double knockout mutant was

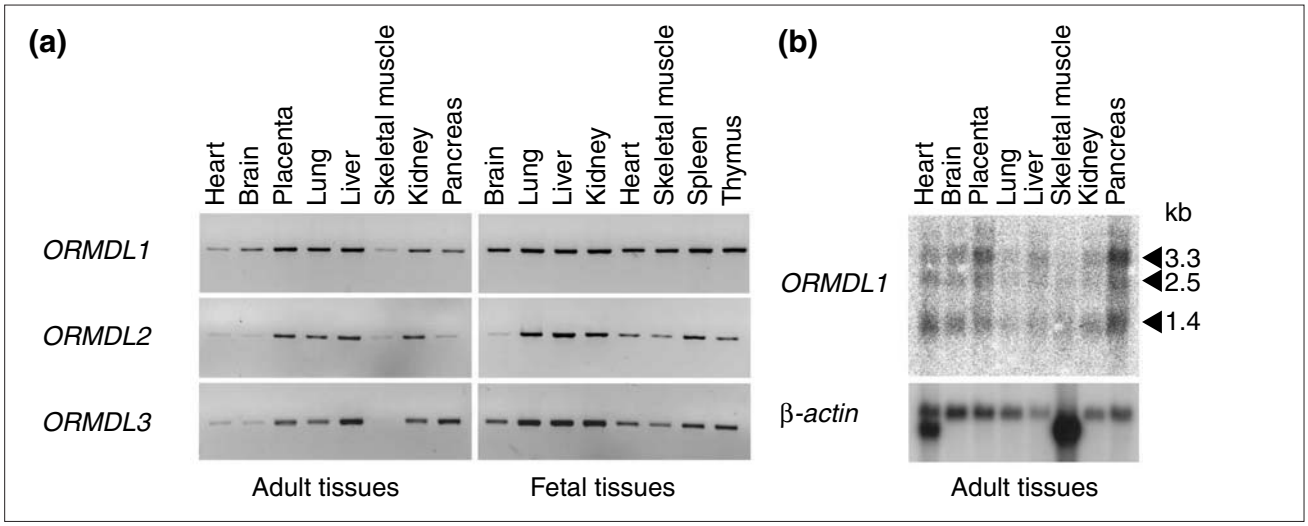



**Figure 5**
RT-PCR analyses of expression patterns of *ORMDL1*, *ORMDL2*, and *ORMDL3*. **(a)** Pattern of expression in a normalized panel of adult and fetal human tissues. **(b)** Northern blot hybridization using a *ORMDL1* probe (containing the whole coding region) against the mRNA of human adult tissues. The size of the three detected transcripts is shown. The same pattern was obtained using a probe containing the *ORMDL1* 3′-UTR (data not shown).

obtained by mating the two single knockout strains, subsequent sporulation, tetrad dissection and analysis of single spores. Verification of *ORM1* and *ORM2* gene integrity and/or deletion for the clones derived from each spore was carried out by independent PCRs with specific primers (Table 3; Figure 8). Overall, 42 independent clones were analyzed, from which 23.8% were wild-type for *ORM1* and *ORM2* (and thus did not grow on G418-supplemented medium), 28.57% single *ORM1* deletants, 28.57% single *ORM2* deletants, 14.28% double knockouts and 4.76% diploids. The two single and double knockouts were viable and no morphological differences were detected between them and the wild-type cells. However, growth rates on rich media (YPD) differed: the double mutant grew much slower than the wild-type and single-knockout strains (data not shown).

A battery of tests was carried out to characterize the yeast mutants on the basis of their sensitivity or resistance to toxic compounds. The YPAD medium was supplemented with one of the following: dithiothreitol (DTT), tunicamycin, $HgCl_2$, cycloheximide, $CaCl_2$, KCl, caffeine, EGTA or SDS. In all cases the double-knockout strain was more sensitive to the toxic agent. In contrast, the growth of the single knockouts did not differ from that of the wild-type strain. This result was particularly evident in the plates supplemented with DTT, $HgCl_2$ or tunicamycin (Figure 8b,c). We also carried out complementation analyses with *ORMDL3*, the human counterpart with the greatest sequence similarity to yeast *ORM1* and *ORM2*. To this end, double-knockout yeast cells were transformed with either multicopy or centromeric expression vectors bearing *ORMDL3* under the control of a constitutive promoter, and dropout-plated in media supplemented with DTT, $HgCl_2$ or tunicamycin. In all cases, some degree of phenotypic rescue was clearly observed, further arguing in favor of functional conservation in the *ORMDL* family (Figure 8c).

## Discussion

We report here the characterization of a novel gene family highly conserved in eukaryotes from yeast to mammals. Most of the species studied have more than one member of this gene family: *S. cerevisiae* and *A. thaliana* each contain two *ORMDL* genes, whereas three copies are present in vertebrates. On the other hand, *Encephalitozoon, Drosophila* and *Ciona* appear to contain a single *ORMDL* gene and no homologous sequence was found in *Caenorhabditis elegans*.

While this work was being carried out, many of the family members became incorporated in the genomic databases of the corresponding organisms, some only as predicted genes. However, even after a significant number of sequences had been annotated, the family had not been formally defined. This omission most probably reflects the fact that present annotations of gene families rely on previously characterized functional signatures or domains. In their absence, protein sequence comparisons remain as the only valuable tool.

A phylogenetic tree was constructed to study the evolutionary relationships between the *ORMDL* members (Figure 3). Several gene duplication events have taken place independently in different lineages, and have given rise to multiple copies in most of the species analyzed. The three human paralogs (*ORMDL1*, *ORMDL2* and *ORMDL3*) showed amino-acid identities of around 80%, whereas pairwise comparisons

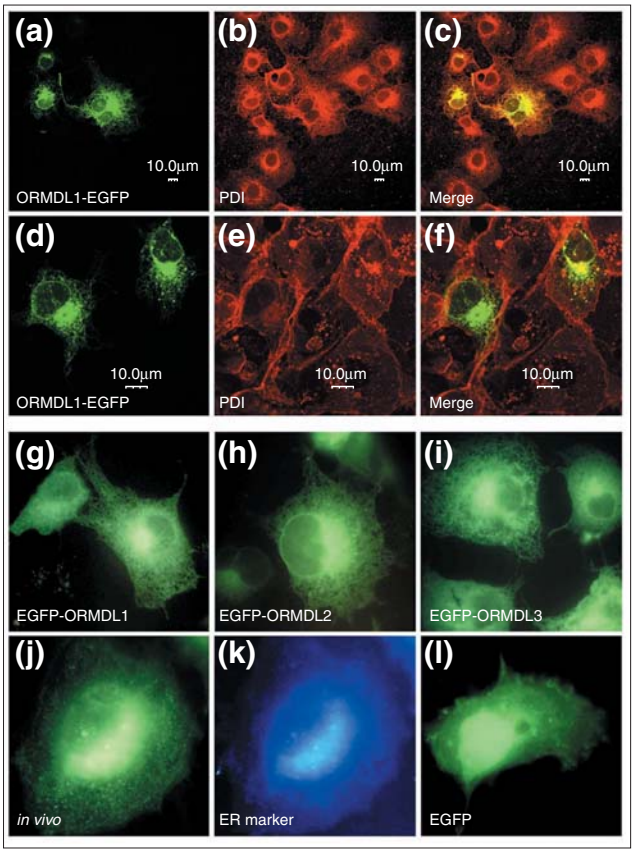



**Figure 6**
ORMDL fused to EGFP at the amino or carboxyl terminus localizes in the perinuclear ER and throughout the ER network. **(a-c)** Confocal scanning microscopy of COS-7 cells transfected with ORMDL1-EGFP (a) co-localize with protein-disulfide isomerase (PDI) (b), an ER-marker protein, showing an overlapping signal in the endoplasmic reticulum (c). **(d-f)** Double-label fluorescence with biotin-labeled concanavalin A showed no overlapping signal with ORMDL1-EGFP at the plasma membrane. **(g-l)** COS-7 cells transfected with EGFP-ORMDL1, EGFP-ORMDL2 and EGFP-ORMDL3 fusion proteins were similar either in fixed (g, h, and i, respectively) and *in vivo* (j) cells, co-localizing with an *in vivo* ER marker (k). EGFP alone was uniformly distributed throughout the cell, including the nucleus (l).

revealed more than 95% identity between human and mouse orthologs (Table 1). Recently, genomic data derived from the complete genome sequence of the pufferfish *Takifugu rubripes* [11], cDNA annotation data in *Mus musculus* [12], and EST data collected from several vertebrate EST projects [4] support the evidence that the presence of three *ORMDL* genes - as characterized in humans - is common to all vertebrates. In addition, a single-copy *ORMDL* has been identified in the pre-vertebrate urochordate *C. intestinalis*. Overall, these data support the evidence that vertebrate *ORMDL* genes are true orthologs, probably arising by the postulated large-scale duplications, either genome doublings or extensive chromosomal duplications, in the vertebrate ancestry ([13] and references therein). In the human genome, the three *ORMDL* genes map within the large paralogous regions defined at 2q, 12q and 17q, which also include segments of chromosomes 7 and 3 [13,14]. These regions contain the HOX clusters as well as two to four members of other families: *ORMDL1* and HOXD at 2q31-32; *ORMDL2* and HOXC at 12q13, and *ORMDL3* and HOXB at 17q21. Moreover, the HOX clusters in mammals are located within chromosomal regions that have maintained extensive chromosomal synteny throughout the vertebrates [15]. No further *ORMDL* sequences were found next to HOXA at 7p15 but, in this context, loss of a fourth copy is a feasible assumption.

In invertebrates, one *ORMDL* gene has been characterized in *Drosophila*. Although no sequence with amino-acid identities above 30% has been found in *Caenorhabditis elegans*, it is possible that a homolog will be identified as the worm genome assembly is completed. Alternatively, it could be argued that *ORMDL* may have been lost in the phylogenetic lineage that leads to the nematode, as has happened with some genes involved in developmental pathways, which are broadly represented in animal phylogeny but missing in *C. elegans* [16,17].

In plants, different lineages show duplicated copies of *ORMDL* genes: the Solanaceae *Lycopersicon esculentum* (tomato) and *Solanum tuberosum* (potato; data not shown), the grasses *Sorghum bicolor* (sorghum) and *Zea mays* (maize; data not shown), and the brassica *Arabidopsis*. The two gene sequences found in *A. thaliana* and yeast are consistent either with the postulated genome duplications or large-scale segmental duplications in the *Arabidopsis* and *Saccharomyces* lineages [18-20]. Interestingly, the yeast *ORM1* and *ORM2* are located within duplicated chromosomal blocks in chromosomes VII and XII, respectively, which include 15 other paralogous gene pairs [19].

Protein alignment showed extended amino-terminal segments of around 20-60 amino acid residues for the yeast ORM proteins in comparison to the plant and animal forms. The conserved residues were mostly clustered in the middle segment of the protein chain. Transmembrane (TM) segment prediction yielded from two to four TM domains, (Table 2 and Figure 2). When using aligned sequences, four hydrophobic domains were apparent, and in most programs were recognized as putative TM sequences. Interestingly, the TM domains are conserved in their relative position but are much less conserved at the residue level than their flanking segments. In fact, the most conserved amino-acid segments are flanked by the two central hydrophobic domains, and probably indicate the location of an important functional domain. Whether this region is relevant to the reported association of yeast ORM2 with SLT2, a serine/threonine protein kinase of the MAP kinase family, remains to be proved [21]. Unfortunately, the reported interaction, from a comprehensive yeast proteome-wide two-hybrid analysis, has not been substantiated with

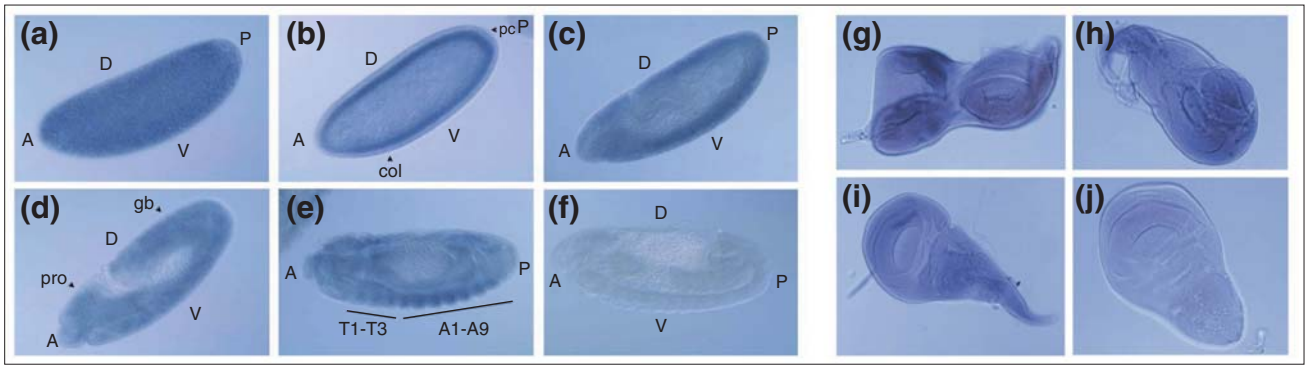



**Figure 7**
Whole-mount *in situ* hybridization to detect *DORMDL* expression in *Drosophila* embryos and imaginal discs. *DORMDL* is ubiquitously expressed in the ectodermal tissues. **(a)** Syncytial blastoderm (stage 3/4, approximately 2 h); **(b)** cellular blastoderm focused in the plane where polar cells are visible (stage 5, 3 h-3 h 15 min); **(c)** lateral view of an stage 11 embryo with the fully stretched germ band; **(d)** lateral view of an stage 11-12 embryo when the germ band begins to recede; **(e)** lateral view of a stage 14 embryo; **(f)** sense probe (stage 13). A, anterior; P, posterior; D, dorsal; V, ventral; pc, polar cells; cel, cellularization; pro, procephalon; gb, germ band; T1-T3, thoracic segments; A1-A9, abdominal segments. Timings of developmental stages are approximate (development at 25°C). **(g)** Eye-antenna imaginal disc; **(h)** leg imaginal disc; **(i)** wing imaginal disc; **(j)** negative control - wing imaginal disc hybridized with the sense probe.

further experimental data. In accordance with the TM prediction, subcellular localization using GFP fusion proteins showed ER anchorage of the human ORMDL proteins. Although co-localization with PDI was used to verify the ER-staining pattern, one cannot discard the possibility that overexpression could influence protein folding and trafficking. Further experiments to characterize protein topology are underway. Searches for signal peptides, nuclear-localization signals, protease-cleavage sites and conserved protein domains were also carried out but did not yield any significant similarities.

Human *ORMDL* homologs were ubiquitously expressed in adult and fetal samples with some minor tissue-specific

**Table 3**

**Primers used for the generation and verification of yeast single knockouts**

Oligonucleotides for the synthesis of *kanMX4* cassettes

| Primer | Sequence 5′ to 3′ |
|---|---|
| YGR038w F | AGG AAC TGC TGA GGC GGC TTC TAC CTC GTA TAG TCG AAA TCA AAC **CGT ACG CTG CAG GTC GAC** |
| YGR038w R | TGT TCA ACT AAT TTG GGC GCG ACC TGT GAT ACC TGG GAT AGA AAT **ATC GAT GAA TTC GAG CTC G** |
| YLR350w F | ATC ATG ATT GAC CGC ACT AAA AAC GAA TCT CCA GCT TTT **GCG TAC GCT GCA GGT CGA C** |
| YLR350w R | CTA ACT AAT TTG AGC ACG GCC CGT AAT ACC AGG GAT GGA T**AT CGA TGA ATT CGA GCT CG** |

Oligonucleotides for gene integrity and gene deletion verification

| | |
|---|---|
| Verification of *ORM1* gene integrity | |
| A1 (F) | GTT TAG GAA CGG ATT ATA GA |
| A2 (R) | CTG GTC ATT TTC ATC TTC AA |
| Verification of *ORM1* substitution | |
| A1 (F) | GTT TAG GAA CGG ATT ATA GA |
| KAN5 (R) | GTT CGG ATG TGA TGT GAG |
| Verification of *ORM2* gene integrity | |
| YLR350w up400 (F) | GCG TAT TTT GAT TGC TCA TA |
| YLR350w ver (R) | TGG TTT CAG GTT AGA CAC AT |
| Verification of *ORM2* substitution | |
| YLR350w up400 (F) | GCG TAT TTT GAT TGC TCA TA |
| KAN5 (R) | GTT CGG ATG TGA TGT GAG |

Oligonucleotides in bold are 18-19 nucleotide homology regions to pFA6a MCS sequences.

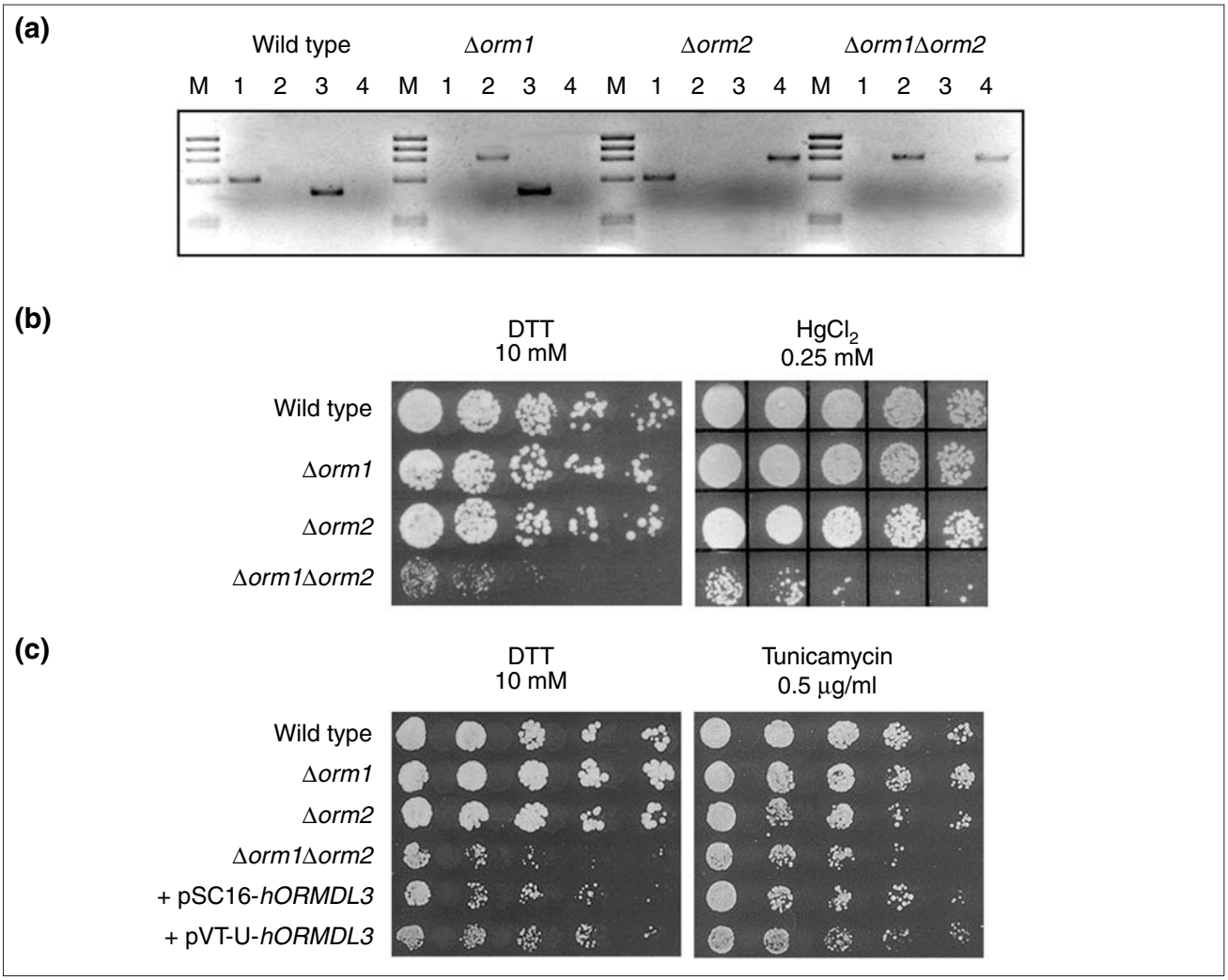



**Figure 8**
Qualitative phenotypic analysis of yeast knockout strains and functional complementation of human *ORMDL3*. **(a)** Genotype PCR analyses of spores derived from tetrad dissection. The deduced genotype for each clone is indicated above. Four different PCR reactions were performed to assess the following: lane 1, *ORM1* integrity; lane 2, *ORM1* deletion; lane 3, *ORM2* integrity; lane 4, *ORM2* deletion. **(b)** Dropout plate showing the growth of haploid wild-type, single-knockout and double-knockout strains. **(c)** Growth of wild-type, single-knockout and double-knockout strains as well as functional complementation of double knockouts transformed with human *ORMDL3*. pSC16-*hORMDL3* and pVT-U-*hORMDL3* denote, respectively, yeast centromeric and non-centromeric plasmid constructs used in the complementation assay. Five sequential dilutions (1:5 each) were plated, left to right, in YPAD medium supplemented with the toxic agents at the indicated concentrations.

differences. There was no evidence of alternative splicing affecting the coding region. However, the northern-blot analysis of *ORMDL1* revealed three transcripts of different size, which may be explained by the use of either different promoters or polyadenylation signals, and/or by alternative splicing affecting only the 5′- or 3′-UTRs. In fact, an alternative exon in the 5′-UTR of *ORMDL1* is reported here.

Functional characterization of gene homologs in model organisms is an invaluable reference for investigating human genes with unknown function. In this respect, *Drosophila* is particularly informative when mutant strains for a specific gene are available. Unfortunately, neither P-element insertions next to *DORMDL* nor information on gene expression were available. Our data from *in situ* hybridization analysis in different developmental stages and imaginal discs showed ubiquitous expression at early embryonic stages, although at late stages the signal was stronger in the ectodermally derived tissues. Overall, these data support a basic cellular function for this gene family.

The search for additional functional clues prompted us to construct and analyze single and double yeast knockouts. Partial functional redundancy of the two yeast genes was

deduced, as single mutants were viable and showed similar responses to the wild-type strain in sensitivity/resistance assays to toxic compounds. Deletion of both genes was not lethal, although yeast-cell growth and ability to respond to chemical aggression were partially affected in double knockouts. It is well known that DTT, a reducing agent that impairs disulfide bond formation, or tunicamycin, a glycosylation inhibitor that blocks the first step in the biosynthesis of *N*-linked oligosaccharides, induce the accumulation of unfolded proteins in the ER lumen. The cellular response to this type of stress (the unfolded protein response (UPR) pathway) eventually leads to the activation of several ER-resident protein chaperons and chaperon cofactors [22]. Given the ER-membrane localization of the ORMDL proteins and the higher sensitivity of the double knockouts to both tunicamycin and DTT, it is reasonable to postulate that these proteins participate in correct protein folding and/or trafficking in the ER, or are involved in the cellular UPR. Interestingly, phenotypic rescue of the double-knockout mutants was achieved by functional complementation with a human homolog, thus clearly supporting functional as well as structural evolutionary conservation of this gene family.

In summary, this study defines a novel eukaryotic gene family whose function is not yet clear. Several questions should be addressed in the future. They mainly concern protein topology, the cellular function of the family members and their association, if any, to human disease.

## Materials and methods

### Cloning of human *ORMDL1* cDNAs

A PCR-derived probe of 647 bp was obtained from an IMAGE clone (220362, Research Genetics) using a primer pair (5′-GGGAACAACTGGACTATG-3′ and 5′-ACTTGTATAGAGAGGATTCC-3′) designed after the retinal EST H86933. This EST contained the STS WI-18706, which mapped to the *RP26* locus. The probe was labeled with [α-$^{32}$P]dCTP by random priming (Roche Molecular Biochemicals) and was used to screen $10^6$ recombinant phages from a commercial λgt10 retinal cDNA library (Clontech). Hybridization was carried out using standard protocols under highly stringent conditions [23]. Positive λ phages were isolated through several rounds of rescreening. Purified phage DNA was *Eco*RI-digested and inserts were subcloned into pBluescript II KS(+) (Stratagene). Sequencing was carried out using a ThermoSequenase II sequencing kit (Amersham Biosciences) and an ABI377 automatic sequencer (Applied Biosystems).

### Rapid amplification of *ORMDL1* cDNA ends

The 5′ and 3′ transcript ends were characterized using the SMART RACE cDNA amplification kit (Clontech) according to the manufacturer's instructions. Total RNA from human placenta was used for first-strand cDNA synthesis using SuperScript II RNase H$^-$ reverse transcriptase (Invitrogen Life Technologies). The primers used were: 5′-AGGAGCCTTGCTTTACCCTGGTCAG-3′ for the 5′-RACE and 5′-CAATGGCTGGTCCTTCAAGTGCTGT-3′ for the 3′-RACE. The RACE products were subcloned using the SureClone ligation kit (Amersham Biosciences) and sequenced.

### Characterization of human *ORMDL1* homologs

BLAST searches [4,5] against the EST section of the GenBank/EMBL databases identified sequences homologous to *ORMDL1*, which belonged to the two UniGene clusters Hs.13144 and Hs.293711. The IMAGE clones 212500, 221470, 1407036, 1407667, and 2364281 were obtained from the UK HGMP Resource Centre and sequenced.

### Computer sequence analysis

A search for homologous protein sequences was carried out with tBLASTN, BLASTP, PSI-BLAST and PHI-BLAST programs at the NCBI BLAST server [4,5]. Translated sequences were aligned using the program CLUSTALW [24], and visualized using the program Boxshade (created by Kay Hofmann and Michael D. Baron). Phylogenetic trees were constructed with the program CLUSTALW, excluding gap positions and correcting for multiple substitutions. Bootstrap analysis [25] was used for confidence evaluation. Potential transmembrane segments were predicted using the HMMTOP, TMAP and TMPRED programs [26-28]. Conserved protein domains and signal peptide searches were performed with reverse position-specific BLAST at the NCBI Conserved Domain Database, including Pfam and Smart databases [4,5,29-31]. Analysis of putative binding sites for transcription factors was carried out using the TESS server [7]. CpG islands were detected using the EMBL CpG Islands service [8]. The human *ORMDL* sequences were mapped using the BLAT program against the Human Genome Working Draft (December 22, 2001 assembly) at UCSC [32,33].

Accession numbers of sequences used are as follows. This work: translations from AF395704 (HsapORMDL1), AF395705 (HsapORMDL2), and AF395708 (HsapORMDL3). From data banks: AAH02146 (MmusORMDL2), BAB26397 (MmusORMDL3), AAF51758 (DmelORMDL), AAK25947 (AthaORMDLa), BAB08433 (AthaORMDLb), CAA97026 (ScerORM1), AAB67252 (ScerORM2), CAA17924 (SpomORM), and translations from BE625123 (MmusORMDL1), AC098491 (RnorORMDL3), BI182456 (SscrORMDL1), BE588447 (BtauORMDL2), BM486897 (GgalORMDL2), BM489594 (GgalORMDL3), BJ037340 (XlaeORMDL1), BJ067072 (XlaeORMDL2), AL633210 (StroORMDL2), fugu clone T008329 (FrubORMDL1), fugu clone T005201 (FrubORMDL2), fugu clone T008602 (FrubORMDL3), AL591593 (DrerORMDL1), AL669202+AV885287 (CintORMDL), BE421632 (HvulORMDL), AW747032 (SbicORMDLa), AW678224 (SbicORMDLb), BG522023 (SrebORMDL), AI772256 (LescORMDLa), AI899675 (LescORMDLb), BM093316 (GmaxORMDL), BE202511 (MtruORMDL), BG140432 (LpenORMDLb), BE238653 (ZmayORMDLb), Y08688 (SmonORM1), and AL590450 (EcunORMDL).

Sequence data for this article were deposited with the GenBank data library under accession nos. AF395701 (*ORMDL1* exons 1, 2, and 3); AF395702 (*ORMDL1* exons 4 and 5); AF395704 (*ORMDL1* mRNA); AF395705, AF395706, AF395707 (*ORMDL2* mRNA); AF395708 (*ORMDL3* mRNA); AF395703 (*DORMDL* mRNA).

### Characterization of the genomic region of *ORMDL1*

The human P1 genomic clone 1670, mapped to 2q31-2q33 [6], was obtained from Incyte Genomics. Plasmid DNA was obtained according to the supplier's instructions. Restriction fragments containing the *ORMDL1* exons were detected by Southern blot analysis using a full-length *ORMDL1* cDNA probe. Positive fragments were subcloned into pBluescript II KS(+), and sequenced. The size of intron 3 was determined by restriction mapping after PCR-amplification.

### RT-PCR and northern blot analysis

For RT-PCR expression analysis on 16 samples from adult and fetal tissues (Multiple Tissue cDNA Panels, Clontech), a pair of primers was designed for each human *ORMDL* gene to amplify the complete ORF, as follows: ORMDL1 forward: 5′-CATGAACGTTGGAGTTGCC-3′, ORMDL1 reverse: 5′-GTAAAATTTTTTTTCAGTTTCAAA-3′, ORMDL2 forward: GAAGAATGAATGTGGGGGTG-3′, ORMDL2 reverse: 5′-CATGGAGCTGTCCCAAAAC-3′, ORMDL3 forward: 5′-GCAGGATGAATGTGGGCACA-3′, ORMDL3 reverse: 5′-GGCAGGGGAAGGGGCTGCA-3′.

PCR amplifications were carried out in 25 μl reactions containing 2.5 μl template cDNA, 0.2 mM dNTPs, 5 μM each primer, 20 mM Tris-HCl, 50 mM KCl, 3 mM $MgCl_2$, and 1 U Taq Platinum (Invitrogen Life Technologies). After an initial 3 min step at 94°C, reactions were subjected to 35 cycles of a denaturation step of 30 sec at 94°C, and an annealing-extension step of 45 sec at either 59°C (*ORMDL1*), 62°C (*ORMDL2*), or 65°C (*ORMDL3*). A final extension step was performed at 72°C for 5 min.

For northern blot analysis, radiolabeled full-length and 3′-UTR *ORMDL1* cDNA probes were hybridized against a commercial human Multiple Tissue Northern blot (Clontech) using standard highly stringent conditions. A human actin probe was used as a control for RNA loading.

### Cell culture, transient expression of *ORMDL1*, *ORMDL2* and *ORMDL3* fused to EGFP, immunolocalization, and confocal laser scanning microscopy

COS-7 (African green monkey kidney) cells were routinely grown in DMEM containing 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml of penicillin and 100 μg/ml of streptomycin (Invitrogen Life Technologies). *ORMDL1*, *ORMDL2* and *ORMDL3* cDNAs were cloned in-frame either upstream or downstream of the gene for GFP using the pEGFP-N2 and pEGFP-C2 vectors (Clontech), respectively. All constructs were verified by sequencing. Transient transfections on COS-7 cells grown on coverslips were performed using the FuGENE reagent (Roche Molecular Biochemicals) following the manufacturer's protocol (the ratio was 1.5 μg DNA to 7.5 μl FuGENE). Empty pEGFP-N2 and pEGFP-C2 vectors were used as controls. Twenty-four hours after transfection, cells were rinsed with 100 mM PBS and fixed in 3% paraformaldehyde and 2% sucrose in 0.1 M phosphate buffer at 4°C for 45 min, then washed and permeabilized (if required) with 0.1% Triton X-100/20 mM glycine/10 mM PBS for 10 min. Following permeabilization, cells were rinsed and blocked in 0.5% BSA/20 mM glycine/10 mM PBS. Cells were incubated either with a specific anti-PDI primary antibody (permeabilized cells) or with biotin-labeled concanavalin A (Sigma) (non-permeabilized cells) at 37°C for 1 h. Upon washing, cells were incubated either with Rhodamine Red™-conjugated anti-rabbit secondary antibody (Jackson ImmunoResearch Laboratories) (permeabilized cells) or with Texas Red-streptavidin (Amersham Biosciences) (non-permeabilized cells) at 37°C for 45 min. *In vivo* co-localization with the ER marker (ER-Tracker Blue-White DPX, from Molecular Probes) was performed 24 h after transfection by rinsing coverslips twice with PBS followed by 30 min incubation at 37°C in a 4 μM ER-Tracker solution. All preparations were embedded in Vectashield mounting medium (Vector Laboratories). Fluorescence was analyzed with a Zeiss Axiophot epifluorescence microscope. Confocal laser scanning microscopy was performed with an Olympus IX-70 inverted laser scanning microscope.

### Cloning of *D. melanogaster* cDNA

A 529-bp probe was amplified by PCR from genomic *D. melanogaster* DNA. The primers used (5′-TCAGTGCCCTTCGTTAGC-3′ and 5′-AGTGTTCCGTGTTGTTTC-3′) were designed after the high-throughput genomic sequences (GenBank) that were identified in a BLAST search using the human *ORMDL1* cDNA sequence. The radiolabeled probe was used to screen 8 x $10^5$ recombinant phages from an adult *D. melanogaster* cDNA library (Stratagene) constructed in Lambda ZAP II. Hybridization, washes, and autoradiography were carried out in standard conditions. After several rescreening rounds, pBluescript SK(+) plasmids containing the cDNA insert were excised from the isolated phages and sequenced.

### *In situ* hybridization on *Drosophila* embryos and larval imaginal discs

Digoxigenin (DIG)-labeled antisense and sense riboprobes covering 84% of the *Drosophila ORMDL* ORF were synthesized using the DIG RNA labeling kit (Roche Molecular Biochemicals) according to the manufacturer's instructions. *Drosophila* embryos at different developmental stages were collected from a 24-h lay, dechorionated, fixed in 2% formaldehyde and 0.5 M EGTA in PBS, prehybridized for 1 h, and hybridized overnight at 55°C. After removing the hybridization solution and washing, embryos were incubated

with 1/2000 anti-DIG antibody conjugated to alkaline phosphatase (Roche Molecular Biochemicals), preabsorbed against fixed embryos. The signal was detected with NBT/BCIP (Roche Molecular Biochemicals). Embryos were mounted on Permount, and examined and photographed under a Zeiss Axiophot microscope.

Wild-type *D. melanogaster* Canton S third-instar larvae were dissected in PBS and fixed in 4% paraformaldehyde in PBS for 20 min. After rinsing in PBS, a second round of fixation was performed in 4% paraformaldehyde in PBS with addition of 0.1% Triton X-100 and 0.1% deoxycholate. After further rinsing with PBS, the imaginal discs were dehydrated through a graded ethanol series. Further protocol steps before the hybridization were carried out as described [34] except that proteinase K digestion was omitted. After 1 h prehybridization at 60°C, the riboprobe was added hybridized overnight at 65°C. Thereafter, the discs were washed for 20 min each in a series of decreasing concentrations of hybridization solution in PBT at 65°C. Two final washes were done with PBT at room temperature. After 1 h of blocking with 2% BSA in PBT, the discs were incubated with preabsorbed anti-DIG, washed and detected as described above. Dissected discs were then mounted in glycerol, examined and photographed under a Zeiss Axiophot microscope.

### Yeast strains and culture media

*S. cerevisiae* W303-1A (*MAT***a**; *ura3-1*; *ade2-1*; *leu2-3,112*; *his3-11,15*; *trp1-Δ2*; *can1-100*) and W303-1B (*MATα*; *ura3-1*; *ade2-1*; *leu2-3,112*; *his3-11,15*; *trp1-Δ2*; *can1-100*) strains were grown on yeast extract peptone dextrose (YPD) media. G418-resistant strains were grown on YPD plates containing 200 mg/l of G418 sulphate (geneticin, Invitrogen Life Technologies).

### Generation of yeast single knockouts and selection

To disrupt YGR038w and YLR350w ORFs, *kanMX4* substitution cassettes were PCR-amplified on plasmid pFA6a-kanMX4 [10] using two sets of primers with 18-19 nucleotide homology to the pFA6a MCS plus 45 nucleotide homology to the flanking regions of either YGR038w (*ORM1*) or YLR350w (*ORM2*) ORFs (Table 3). One microgram of each gel-purified PCR product was used to transform yeast W303-1A and W303-1B strains by the lithium acetate method [35]. Transformed yeast cells were recovered by G418 selection. Homologous integration of the *kanMX4* cassette was verified by colony PCR. Two sets of primers were used for each gene to test for gene disruption and gene integrity, respectively (Table 3). Primer design and PCR conditions essentially followed the EUROFAN guidelines [36].

### Generation of yeast double knockouts

The W303-1A *Δorm1* (*MAT***a**) and W303-1B *Δorm2* (*MATα*) deletant strains were crossed to obtain the double heterozygous diploid strain. The reciprocal cross was also performed. The double heterozygous diploids were plated on sporulation medium (1% potassium acetate, 0.1% yeast extract, 0.05% dextrose, 2% agar) to induce meiosis [37]. Tetrad dissection and analysis were carried out as described [37]. Clones from spores were re-streaked on YPD plates containing 200 mg/l G418, and G418-resistant transformants were verified by colony PCR to assess single and double knockouts.

### Qualitative phenotypic analysis of knockout strains

To test growth differences and sensitivity/resistance of the mutant strains against toxic compounds, single and double knockouts as well as wild-type clones were grown on liquid yeast extract peptone adenine dextrose (YPAD) medium for 14 h at 30°C, diluted to $OD_{600}$ 0.1 and grown to $OD_{600}$ 0.6-0.8. They were then sequentially diluted (5 dilutions, 1:5 each) with YPAD into the wells of microtiter plates and dropout-plated onto the corresponding test media (0.25 mM and 0.4 mM $HgCl_2$; 10 mM and 15 mM DTT; 0.5 μg/ml and 1 μg/ml tunicamycin; 0.7 M $CaCl_2$; 0.1% caffeine; 0.18 mg/ml and decreasing concentrations of cycloheximide; 3 mM EGTA; 1.5 M KCl; 0.01% SDS, all in YPAD). To test heat sensitivity, cultures were incubated in a water bath at 52°C for 35 min, sequentially diluted and plated on YPAD. We analyzed five independent clones for each mutant strain and three for each wild type, respectively.

### Functional complementation assay with human *ORMDL3* in yeast double knockouts

A yeast centromeric plasmid, pSC16 [38] (kindly provided by S. Cervantes), and a non-centromeric plasmid (pVT-U) containing the *ADH1* promoter and termination expression cassettes were used to clone human *ORMDL3* ORF. Each construct (100 ng) was used to transform yeast W303 *Δorm1 Δorm2* strain by the lithium acetate method. Transformed yeast cells were recovered by selection on minimal synthetic dropout (SD) medium without leucine (centromeric plasmid) or without uracil (non-centromeric plasmid). Transformants were grown on liquid YPAD, sequentially diluted and dropout-plated onto the corresponding test media (0.25 mM and 0.4 mM $HgCl_2$; 10 mM and 15 mM DTT; 0.5 μg/ml and 1 μg/ml tunicamycin, all in YPAD).

## Acknowledgements

We thank Florenci Serras and Sergio González-Crespo for support on *Drosophila in situ* hybridizations, Ricardo Casaroli for advice on immunolocalization, and Maite Rodríguez-Manzaneque for technical advice on spore dissection. We are grateful to Charles Brenner for critically reviewing the manuscript and suggesting the use of tunicamycin and DTT tests in the yeast mutants. We also thank Robin Rycroft for revising the English. The Serveis Científico-Tècnics (Universitat de Barcelona) provided DNA sequencing and confocal laser scanning microscopy facilities. This study was funded by MCyT PM99-0168, CIRIT 1999SGR 00027, and Fundaluce to R.G.-D., and Magn. Bergvall Foundation to L.H. L.H. was in receipt of a postdoctoral fellowship from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) and a travel grant from the Swedish Society of Medicine. M.T. is in receipt of a predoctoral fellowship from the CIRIT (Generalitat de Catalunya).

## References


1. Bayés M, Goldaracena B, Martínez-Mir A, Iragui-Madoz MI, Solans T, Chivelet P, Bussaglia E, Ramos-Arroyo MA, Baiget M, Vilageliu L, *et al.*: **A new autosomal recessive retinitis pigmentosa locus maps on chromosome 2q31-q33.** *J Med Genet* 1998, **35:**141-145.
2. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
3. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, *et al.*: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287:**2185-2195.
4. **NCBI BLAST Server** [http://www.ncbi.nlm.nih.gov/BLAST/]
5. Altschul, SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: A new generation of protein database search programmes.** *Nucleic Acids Res* 1997, **25:**3389-3402.
6. Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, *et al.*: **Mutations of two *PMS* homologues in hereditary nonpolyposis colon cancer.** *Nature* 1994, **371:**75-80.
7. **Transcription Element Search System (TESS) Server** [http://www.cbil.upenn.edu/tess/index.html]
8. **EMBL CpG Islands Service** [http://www.ebi.ac.uk/cpg/]
9. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.*: **The sequence of the human genome.** *Science* 2001, **291:**1304-1351.
10. Wach A, Brachat A, Pöhlmann R, Philippsen P: **New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*.** *Yeast* 1994, **10:**1793-1808.
11. **JGI Fugu Genome Project** [http://www.jgi.doe.gov/fugu/index.html]
12. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium: **Functional annotation of a full-length mouse cDNA collection.** *Nature* 2001, **409:**685-690.
13. Sidow A: **Gen(om)e duplications in the evolution of early vertebrates.** *Curr Opin Genet Dev* 1996, **6:**715-722.
14. Lundin LG: **Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse.** *Genomics* 1993, **16:**1-19.
15. Popovici C, Leveugle M, Birnbaum D, Coulier F: **Homeobox gene clusters and the human paralogy map.** *FEBS Lett* 2001, **491:**237-242.
16. Holland PW, Garcia-Fernandez J, Williams NA, Sidow A: **Gene duplications and the origins of vertebrate development.** *Development* 1994, **Suppl:** 125-133.
17. Ruvkun G, Hobert O: **The taxonomy of developmental control in *Caenorhabditis elegans*.** *Science* 1998, **282:**2033-2041.
18. Vision TJ, Brown DG, Tanksley SD: **The origins of genomic duplications in *Arabidopsis*.** *Science* 2000, **290:**2114-2117.
19. Seoighe C, Wolfe KH: **Updated map of duplicated regions in the yeast genome.** *Gene* 1999, **238:**253-261.
20. Llorente B, Malpertuy A, Neuveglise C, de Montigny J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, *et al.*: **Genomic exploration of the Hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*.** *FEBS Lett* 2000, **487:**101-112.
21. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y**: A comprehensive two-hybrid analysis to explore the yeast protein interactome.** *Proc Natl Acad Sci USA* 2001, **98:**4569-4574.
22. Patil C, Walter P: **Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals.** *Curr Opin Cell Biol* 2001, **13:**349-356.
23. Sambrook J, Russel D: *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 2000.
24. Thompson JD, Higgins DG, Gibson TJ: **CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
25. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39:**783-791.
26. Tusnády GE, Simon I: **The HMMTOP transmembrane topology prediction server.** *Bioinformatics* 2001, **17:**849-850.
27. Persson B, Argos P: **Prediction of transmembrane segments in proteins utilising multiple sequence alignments.** *J Mol Biol* 1994, **237:**182-192.
28. **TMpred Server** [http://www.ch.embnet.org/software/TMPRED_form.html]
29. **NCBI Conserved Domain Database** [http://www.ncbi.nlm.nih.gov/Structure/cdd /cdd.shtml]
30. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL: **The Pfam Protein Families Database.** *Nucleic Acids Res* 2000, **28:**263-266.
31. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: Identification of signalling domains.** *Proc Natl Acad Sci USA* 1998, **95:**5857-5864.
32. **UCSC Human Genome Working Draft** [http://genome.cse.ucsc.edu/]
33. Kent WJ, Haussler D: **Assembly of the working draft of the human genome with GigAssembler.** *Genome Res* 2001, **11:**1541-1548.
34. Lehner CF, O'Farrell PH: ***Drosophila cdc2* homologs: a functional homolog is coexpressed with a cognate variant.** *EMBO J* 1990, **9:**3573-3581.
35. Gietz RD, Woods RA: **High efficiency transformation with lithium acetate.** In *Molecular Genetics of Yeast: A Practical Approach.* Edited by Johnston JR. Oxford: IRL Press; 1994: 121-134.
36. **Yeast EUROFAN Projects Web Page** [http://www.mips.biochem.mpg.de/proj /eurofan/]
37. Kaiser C, Michaelis S, Mitchell A: *Methods in Yeast Genetics.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 1994.
38. Cervantes S, Gonzàlez-Duarte R, Marfany G: **Homodimerization of presenilin N-terminal fragments is affected by mutations linked to Alzheimer's disease.** *FEBS Lett* 2001, **505:**81-86.
39. Zhang QH, Ye M, Wu XY, Ren SX, Zhao M, Zhao CJ, Fu G, Shen Y, Fan HY, Lu G, *et al.*: **Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34$^+$ hematopoietic stem/progenitor cells.** *Genome Res* 2000, **10:**1546-1560.
40. Børsting C, Hummel R, Schultz ER, Rose TM, Pedersen MB, Knudsen J Kristiansen K: ***Saccharomyces carlsbergensis* contains two functional genes encoding the acyl-CoA binding protein, one similar to the ACB1 gene from *S. cerevisiae* and one identical to the ACB1 gene from *S. monacensis*.** *Yeast* 1997, **13:**1409-1421.
41. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, *et al.*: **Life with 6000 genes.** *Science* 1996, **274:**546-567.
42. Yoshioka S, Kato K, Nakai K, Okayama H, Nojima H: **Identification of open reading frames in *Schizosaccharomyces pombe* cDNAs.** *DNA Res* 1997, **4:**363-369.
43. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, *et al.*: **Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*.** *Nature* 2001, **414:**450-453.
44. Roth FP, Hughes JD, Estep PW, Church GM: **Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation.** *Nat Biotechnol* 1998, **16:**939-945.
45. Jelinsky SA, Estep P, Church GM, Samson LD: **Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes.** *Mol Cell Biol* 2000, **20:**8157-8167.
46. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I: **The transcriptional program of sporulation in budding yeast.** *Science* 1998, **282:** 699-705.
47. Washburn MP, Wolters D, Yates JR 3rd**: Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nat Biotechnol* 2001, **19:**242-247.
48. Yale J, Bohnert HJ: **Transcript expression in *Saccharomyces cerevisiae* at high salinity.** *J Biol Chem* 2001, **276:**15996-16007.

# CAPÍTOL 2

## Cerca i identificació del gen responsable de retinitis pigmentària autosòmica recessiva del locus RP26: el gen *CERKL*

Publicació a la qual ha donat lloc aquest treball:

Mutation of *CERKL*, a novel human ceramide kinase gene, causes autosomal recessive retinitis pigmentosa (RP26). *American Journal of Human Genetics* 74: 128–138 (2004).

En aquest segon capítol de resultats, s'exposa el conjunt d'aproximacions que, al llarg d'un any, vam dur a terme per a la cerca del gen del locus RP26 i que van conduir a la caracterització del gen *CERKL*. A la primera part del capítol, tractarem en detall tota l'anàlisi de gens candidats, la determinació precisa de les recombinacions que defineixen els marges del locus, l'anàlisi en profunditat de marcadors al llarg de tot l'interval de cosegregació, el mapatge d'homozigositat, la caracterització d'una regió candidata dins del locus i la identificació final del gen *CERKL*. La meva intenció en escriure aquesta primera part ha estat la de reflectir tota la feina prèvia a la caracterització d'aquest nou gen de retinitis pigmentària. La segona part del capítol la compon la publicació a la qual va donar lloc el treball de cerca del gen *RP26*, i que inclou també la caracterització inicial del gen.

§

A començaments de l'any 2002, quan vam iniciar el projecte de cerca del gen *RP26*, les dades parcials de seqüència dipositades a la base de dades pública (GenBank/EMBL/DDBJ) encara eren fragmentàries i continuaven fent difícil una exploració en profunditat del locus RP26, com la que duríem a terme més endavant. Per començar, no disposàvem d'informació suficient sobre la localització precisa en el mapa físic dels marcadors polimòrfics de tipus microsatèl·lit que havien estat utilitzats en l'anàlisi de lligament realitzada en la família P2 (FIGURA 1), no existia la quantitat d'informació que tenim actualment sobre marcadors polimòrfics d'un únic nucleòtid (SNP), i la regió del mapa físic ocupada pel locus RP26 contenia nombrosos buits de seqüència i contigs de clons mal encavalcats. Tot això, conjuntament amb les dimensions del locus –17.4 Mb de seqüència, gairebé un 7% del total del cromosoma 2– feien pràcticament inabordable la cerca del gen *RP26*, en un interval que podia contenir més d'un centenar de gens, sense més informació que la derivada de l'anàlisi de lligament realitzada en la família P2 (FIGURA 1).

A l'inici del projecte, i amb dades completament parcials de la seqüència d'aquesta regió a les mans, ens vam plantejar dues vies paral·leles per a la identificació del gen *RP26*: (*a*) l'anàlisi de marcadors microsatèl·lits addicionals, per tal de reduir l'interval de cosegregació del locus i determinar si, dins el locus RP26, existia realment una regió d'homozigositat, en la qual s'hauria de trobar el gen *RP26*; i (*b*) la caracterització en profunditat d'una pressumpta regió crítica, determinada pel solapament de l'interval de lligament, establert en la família P2 (FIGURA 1), i la cosegregació amb marcadors del cromosoma 2, documentada en una família xinesa per un grup d'investigadors de la Universitat de Fudan, a Xangai (resultats no publicats, FIGURA 2).

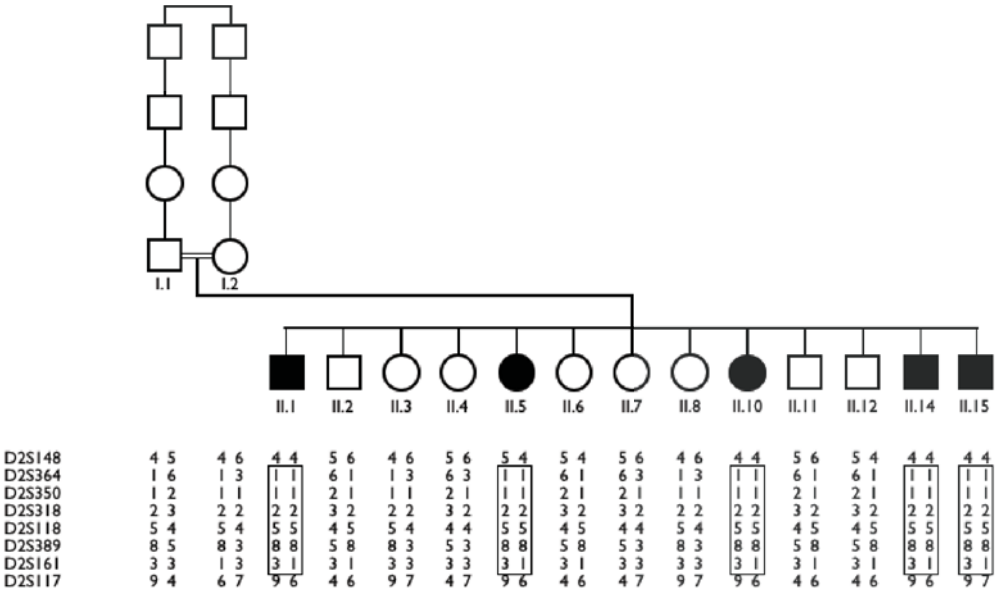
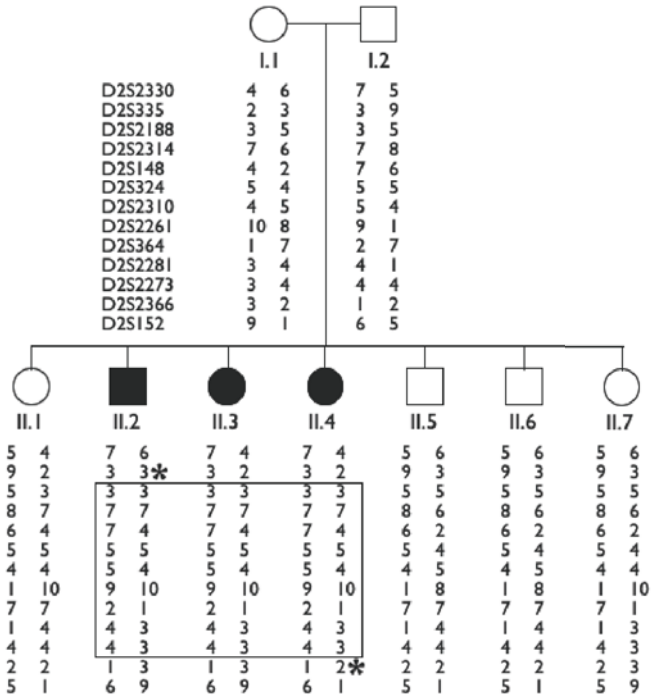



**Figura 1** | Anàlisi d'haplotips en la família P2. Els requadres delimiten l'interval de cosegregació entre D2S148 i D2S117.

## 2.1. Anàlisi de gens candidats a la regió crítica del locus RP26

L'any 2000, el grup del professor Jianhua Chai, de la Universitat de Fudan, Xangai, Xina, s'havia posat en contacte amb el nostre grup perquè disposaven d'una família xinesa amb retinitis pigmentària autosòmica recessiva que cosegregava amb marcadors de 2q31-q32, en l'interval comprès entre els marcadors D2S335 i D2S2366 (FIGURA 2).




| Pedigrí xinès afectat de retinitis pigmentària en el qual es mostra l'anàlisi d'haplotips i la cosegregació amb marcadors del cromosoma 2 realitzada al laboratori del professor Jianhua Chai (dades no publicades). El requadre indica els marcadors que cosegreguen amb la malaltia, i els asteriscos els marcadors recombinants que defineixen els marges de l'interval de cosegregació.

El solapament dels intervals de cosegregació d'aquesta família xinesa i de la família P2 –si consideràvem la hipòtesi que la malaltia en les dues famílies era causada per l'alteració del mateix gen– definia una regió crítica comuna sobre el mapa físic de 6 Mb, entre els marcadors D2S148 i D2S2366. Com veurem al capítol següent, l'anàlisi de nous marcadors microsatèl·lits va permetre reduir aquesta regió crítica a un interval d'aproximadament 3 Mb, entre els marcadors D2S2261 i D2S2366 (FIGURA 3). Basant-nos en l'ordenament del genoma realitzat a partir de les dades del desembre del 2001 (*Human Draft Genome Assembly* UCSC hg10), aquesta regió contenia inicialment vuit gens descrits a la base de dades RefSeq (FIGURA 3). Aquests gens eren *UBE2E3* (enzim conjugador d'ubiquitina E2E3), *PRO1316* (proteïna predita HQ1316), *ITGA4* (precursor de la subunitat β de la integrina 4), *NEUROD1* (factor de diferenciació neurogènica 1), *SSFA2* (antigen específic d'esperma 2), *PDE1A* (fosfodiesterasa 1A), *FRZB* (proteïna relacionada amb *frizzled*), i *NCKAP1* (proteïna associada a NCK 1).

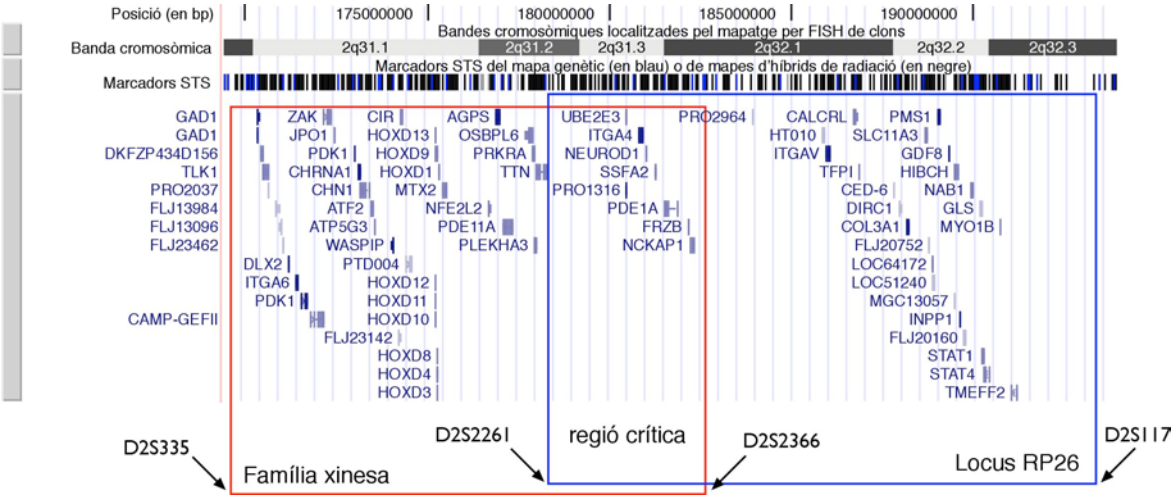



**Figura 3** | Esquema en el qual es mostren, sobre el mapa físic (ordenament de desembre de 2001), els intervals de cosegregació de la família xinesa (requadre vermell) i del locus RP26 (requadre blau), al moment d'iniciar aquest estudi. La regió de solapament entre els dos intervals va determinar una regió crítica en la qual hi havia 8 gens (indexats a la base RefSeq). Els símbols dels gens es mostren en blau i les fletxes indiquen els microsatèl·lits que flanquegen cadascun dels intervals. Mapa físic adaptat d'UCSC Genome Bioinformatics (genome.ucsc.edu).

Abans d'analitzar la presència de mutacions en cadascun d'aquests gens, mitjançant la seqüenciació directa de la regió codificant de tots ells, vam voler prioritzar la seqüenciació d'uns respecte a d'altres, basant-nos en criteris funcionals i dades sobre la seva expressió a la retina. L'anàlisi de l'expressió a la retina dels vuit gens de la regió crítica va ser realitzada per PCR sobre una genoteca de cDNA de retina humana (FIGURA 4). D'aquests 8 gens, únicament vam poder descartar clarament l'expressió d'*ITGA4*. *NEUROD1* –un factor de diferenciació neurogènica que s'expressa tant al pàncrees com a la retina en desenvolupament– també va ser inclòs al llistat de gens candidats basant-nos en les dades de la literatura, tot i que mostrava un senyal molt poc intens després de l'amplificació.

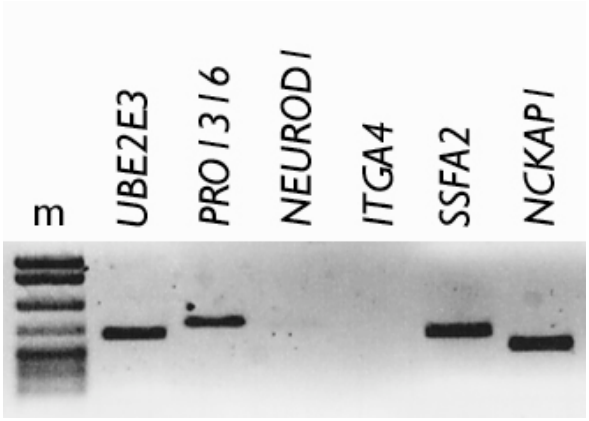



**Figura 4** | Anàlisi de l'expressió a la retina de 6 dels 8 gens de la regió crítica. No es mostren els resultats dels gens *PDE1A* i *FRZB*, que també s'expressen a la retina. Dels gens analitzats, únicament no vam poder detectar l'expressió d'*ITGA4*. *NEUROD1* s'expressa a la retina, tot i que en aquesta anàlisi el nivell d'amplificació és baix.

Inicialment, doncs, vam considerar 7 gens, dels quals en vam obtenir les seqüencies dels cDNA a les bases de dades, incloent els exons diferencials. Vam comparar cadascun d'aquests cDNA amb les dades de seqüència genòmica disponibles d'aquesta regió del cromosoma 2, mitjançant els algoritmes BLAST i BLAT (Altschul et al. 1990, Kent 2002) i, d'aquesta manera, vam poder obtenir l'estructura genòmica de tots set; encara que algun d'ells presentava discontinuïtats de seqüència. A partir de l'estructura genòmica de cada gen, podíem dissenyar oligonucleòtids específics que flanquegessin cada exó per tal d'amplificar-los i poder-los analitzar per seqüenciació. Com hem apuntat anteriorment, l'existència de discontinuïtats de seqüència en aquesta regió dificultava la localització, en la seqüència genòmica, d'alguns segments dels cDNA d'aquests gens (FIGURA 5). Aquest va ser el cas del primer exó del gen *FRZB* i dels 7 darrers exons del gen *NCKAP1*; per a l'obtenció de la seqüència genòmica completa dels quals, vam haver de recórrer a la base de dades de l'empresa Celera Genomics, i a les actualitzacions de la seqüència pública.

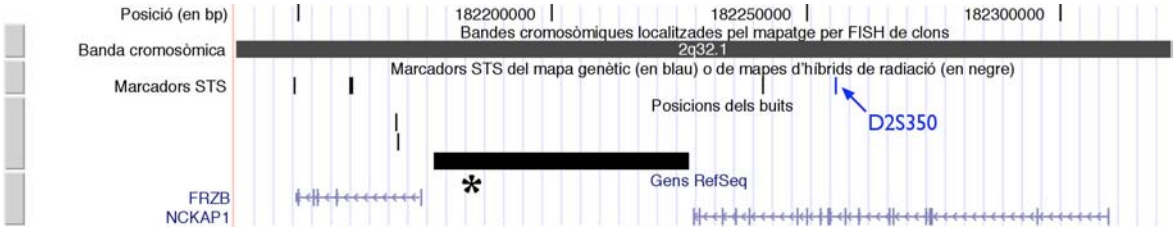



**Figura 5** | Regió genòmica entre els gens *FRZB* i *NCKAP1*. Es mostra la discontinuïtat de seqüència (caixa negra) que s'extenia al llarg de les regions 5' del gen *FRZB* i 3' del gen *NCKAP1* i que impedia situar sobre la seqüència genòmica el primer exó de *FRZB* i els darrers 7 exons d'*NCKAP1*. A l'esquema, s'indiquen amb línies horitzontals, en blau clar, les posicions dels gens, sobre les quals els exons (amb línies verticals) i el sentit de transcripció (amb fletxes). Mapa físic, que correspon a l'ordenament de desembre 2001, adaptat d'UCSC Genome Bioinformatics.

En paral·lel a la seqüenciació de 7 dels 8 gens de la regió crítica descrits en RefSeq, vam iniciar la cerca d'mRNA, poc caracteritzats, a les bases de dades de GenBank i de Celera Genomics. D'aquesta manera, vam localitzar 5 nous mRNA a la regió crítica (FIGURA 6): AF494535 –que denominem PP1I–, similar a la subunitat 1A reguladora (inhibidora) de la fosfatasa 1; AK027147 –que denominem PDIP5–, lleugerament similar a un putatiu precursor de la proteïna disulfur isomerasa P5

(més endavant va ser anotat com a *ERdj5*, *DNAJC10*); AF411516, relacionat amb una fosfatasa d'especificitat dual (més endavant va ser anotat com a *SKRP1, DUSP19*); AK021479 (MP44, *LOC129401*), similar a la proteïna mitòtica 44 de *Xenopus*; i, a la base de Celera, KER8II, queratina 8 tipus II citoesquelètica, que a la base pública mai ha estat localitzat a la regió. D'aquests cinc nous mRNA, únicament 3 s'expressaven a la retina –PDIP5, *SKRP1* i MP44– i van ser incorporats al llistat de gens candidats de la regió crítica.

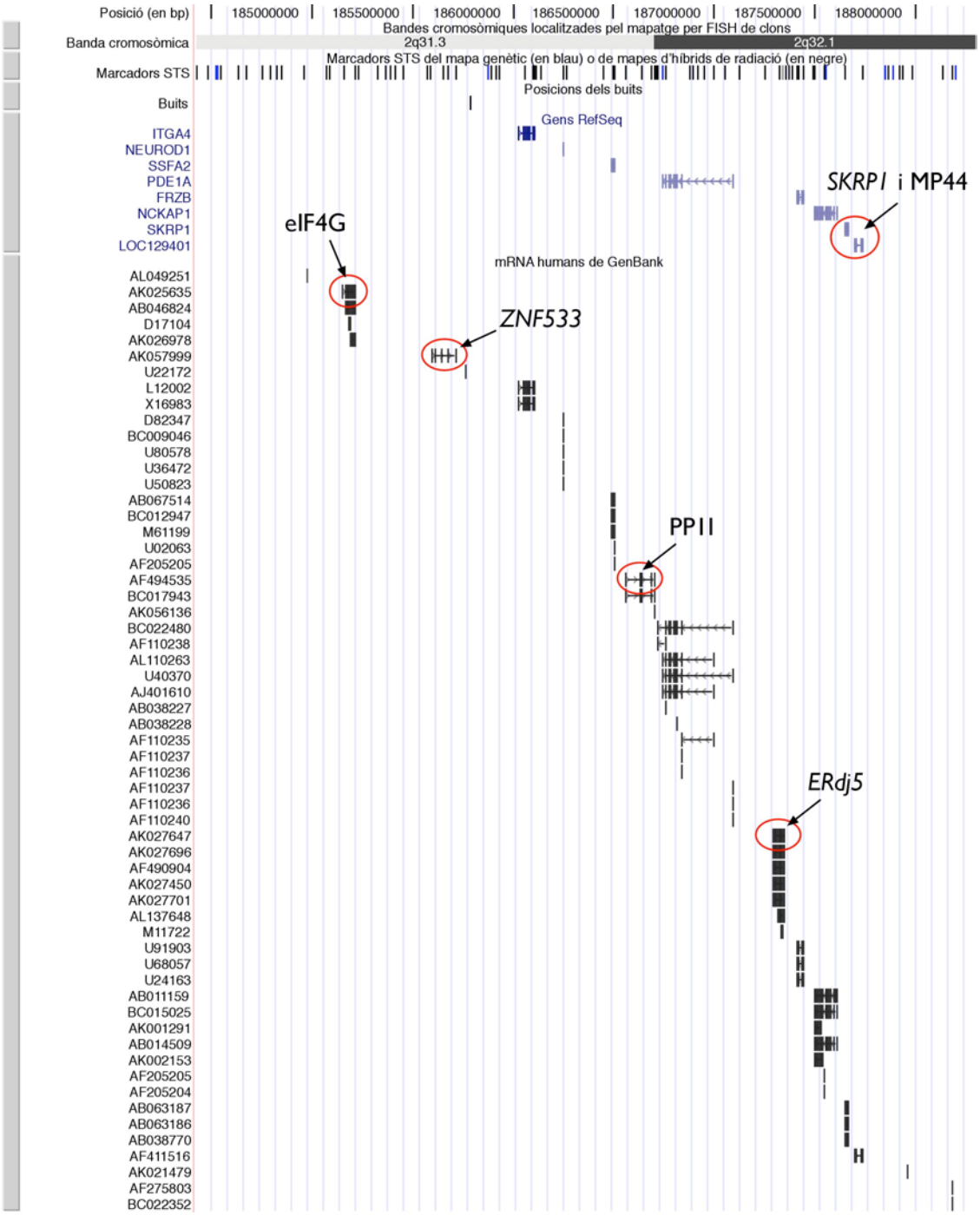



**Figura 6** | Regió crítica, entre D2S2261 i D2S2366, a l'ordenament d'abril de 2002. Es mostren els gens indexats en RefSeq (en blau) i els mRNA de GenBank (en negre). S'hi han indicat amb cercles vermells els gens i mRNA localitzats per primer cop en aquest ordenament (eIF4G, ZNF533, PP11, *ERdj5*, *SKRP1* i MP44). Adaptada d'UCSC Genome Bioinformatics.

A finals de la primavera del 2002, es va fer pública la nova actualització de l'ordenament del genoma humà, basada en les dades de seqüència de l'abril del 2002 (*Human Draft Genome Assembly* UCSC hg11). A la regió crítica del locus RP26 (FIGURA 6), van ser incorporades noves entrades d'mRNA de GenBank: AK025635, que vam anomenar eIF4G perquè codifica un putatiu factor iniciador d'allargament; i AK057999, que codifica una proteïna amb dits de zinc, que actualment ha estat anotat com a *ZNF533*. En aquest ordenament, dos dels gens analitzats –*UBE2E3* i *PRO1316*– van ser resituats fora de la regió.

En aquest punt del projecte, a finals de l'estiu de 2002, havíem seqüenciat tots els exons codificants de 12 gens de la regió crítica –*ZNF533*, eIF4G, *UBE2E3*, *PRO1316*, *NEUROD1*, *SSFA2*, *PDE1A*, PDIP5 (*ERdj5*, *DNAJC10*), *FRZB*, *NCKAP1*, *SKRP1*, MP44– sense trobar cap variant patogènica.

§

## 2.2. Refinament de l'interval de cosegregació del locus RP26

En paral·lel a la seqüenciació de gens candidats en la regió crítica del locus RP26, ens vam proposar d'analitzar nous marcadors microsatèl·lits situats en els marges del locus per tal de reduir l'interval de cosegregació. Els marges de l'interval de cosegregació havien estat definits per l'existència de dues recombinacions: en el marge centromèric, el pacient II.5 de la família P2 era recombinant pel marcador D2S148 (cromosoma patern); i en el marge telomèric, el pacient II.15 era recombinant pel marcador D2S117 (cromosoma matern). Com hem vist anteriorment, l'interval de cosegregació, entre els marcadors D2S148 i D2S117, abarcava 11 cM del mapa genètic. Actualment, aquesta distància genètica correspon a una regió de 17.4 Mb sobre el mapa físic (*Human Draft Genome Assembly* UCSC hg16; NCBI Build 34).

En el marge centromèric del locus, vam analizar els marcadors microsatèl·lits D2S324, D2S2978 i D2S2261, i els marcadors de tipus SNP TSC0036660 i TSC1026876. El pacient II.5 era recombinant pels tres marcadors microsatèl·lits (FIGURA 7). Dels dos SNP analitzats, únicament TSC1026876 era informatiu per a analitzar la recombinació paterna, perquè el pare era heterozigot per aquest marcador. El pacient II.5 presentava un patró no recombinant per a TSC1026876. Així, el mapatge de la recombinació en el cromosoma patern, la situa en la regió entre el microsatèl·lit D2S2261 (recombinant) i l'SNP TSC1026876 (no recombinant).

En el marge telomèric del locus, vam analitzar els marcadors microsatèl·lits D2S273, D2S315 i D2S280 i els marcadors SNP TSC0040380 i rs3738882. El patró del microsatèl·lit D2S273 en el pacient II.15 era recombinant (FIGURA 8). Tant el microsatèl·lit D2S315, com l'SNP TSC0040380, eren no informatius per a estudiar la recombinació en el cromosoma matern, perquè la mare era homozigota pels dos marcadors. En canvi, l'SNP rs3738882 si que era informatiu per a analitzar la

recombinació en el cromosoma matern heredat pel pacient II.15 i, en aquest cas, el patró era no recombinant. D'aquesta manera, podem afirmar que la recombinació en el cromosoma matern d'aquest individu està localitzada entre els marcadors rs3738882 (no recombinant) i D2S273 (recombinant).

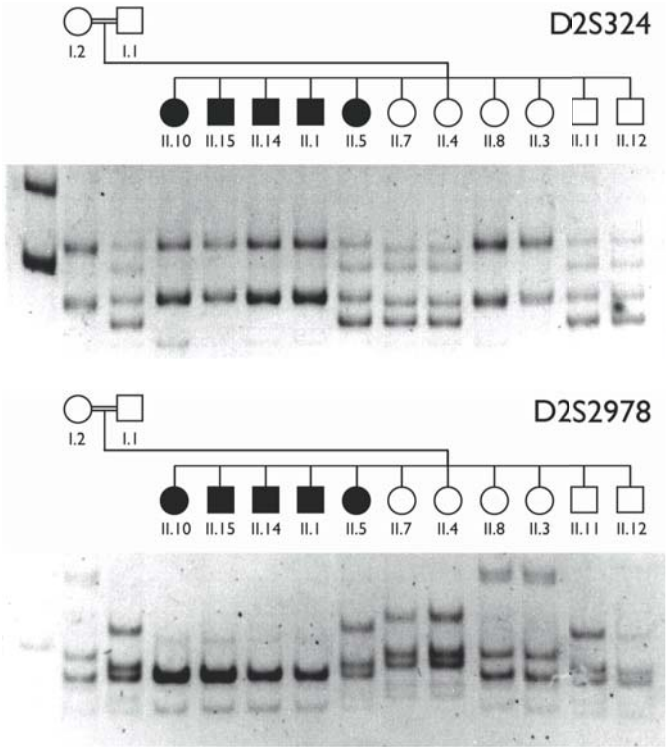



**Figura 7** | Patró al·lèlic dels microsatèl·lits D2S324 i D2S2978 en la família P2, visualitzat en un gel de poliacrilamida. Aquests dos marcadors es troben en el marge centromèric del locus RP26 (vegeu-ne la localització a la TAULA 1 DE RESULTATS). El pacient II.5 presenta una recombinació en el cromosoma patern, com es posa de manifest en analitzar aquests microsatèl·lits. L'anàlisi dels microsatèl·lits D2S324 i D2S2978, conjuntament amb D2S2261, va permetre acotar el marge centromèric –definit per D2S148– en un punt situat entre D2S2261 i l'SNP TSC1026876. En la figura, els fills de la família han estat agrupats segons el seu genotip i s'ha preservat la numeració original, que respon a l'edat.

Amb aquestes noves dades, vam poder acotar més finament el marge centromèric del locus 3.3 Mb endins i el marge telomèric 2 Mb endins. Així, vam reduir l'interval de cosegregació inicial de 17.4 Mb, entre els marcadors D2S148 i D2S117, a un interval de 12 Mb, entre els marcadors D2S2261 i D2S273 (FIGURA 9).

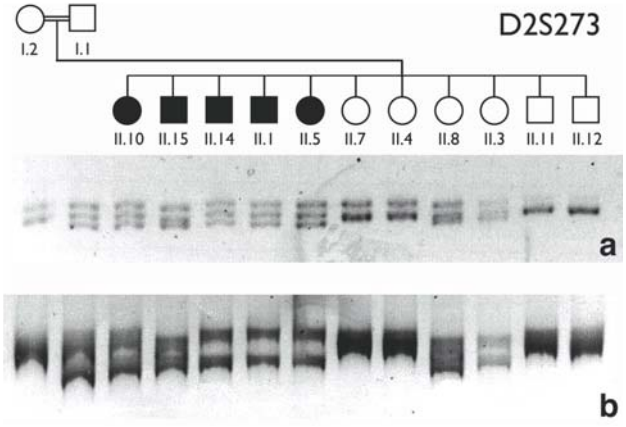



**Figura 8** | Patró al·lèlic del microsatèl·lit D2S273 a la família P2, visualitzat en un gel d'agarosa al 4% (a) i en un gel de poliacrilamida (b). D2S273 està situat en el marge telomèric del locus RP26 (vegeu-ne la localització a la TAULA 1). El pacient II.15 presenta un patró al·lelic que evidencia una recombinació en el cromosoma matern. A partir de l'anàlisi d'aquest microsatèl·lit i de marcadors SNP vam poder córrer el marge telomèric del locus 2 Mb endins.

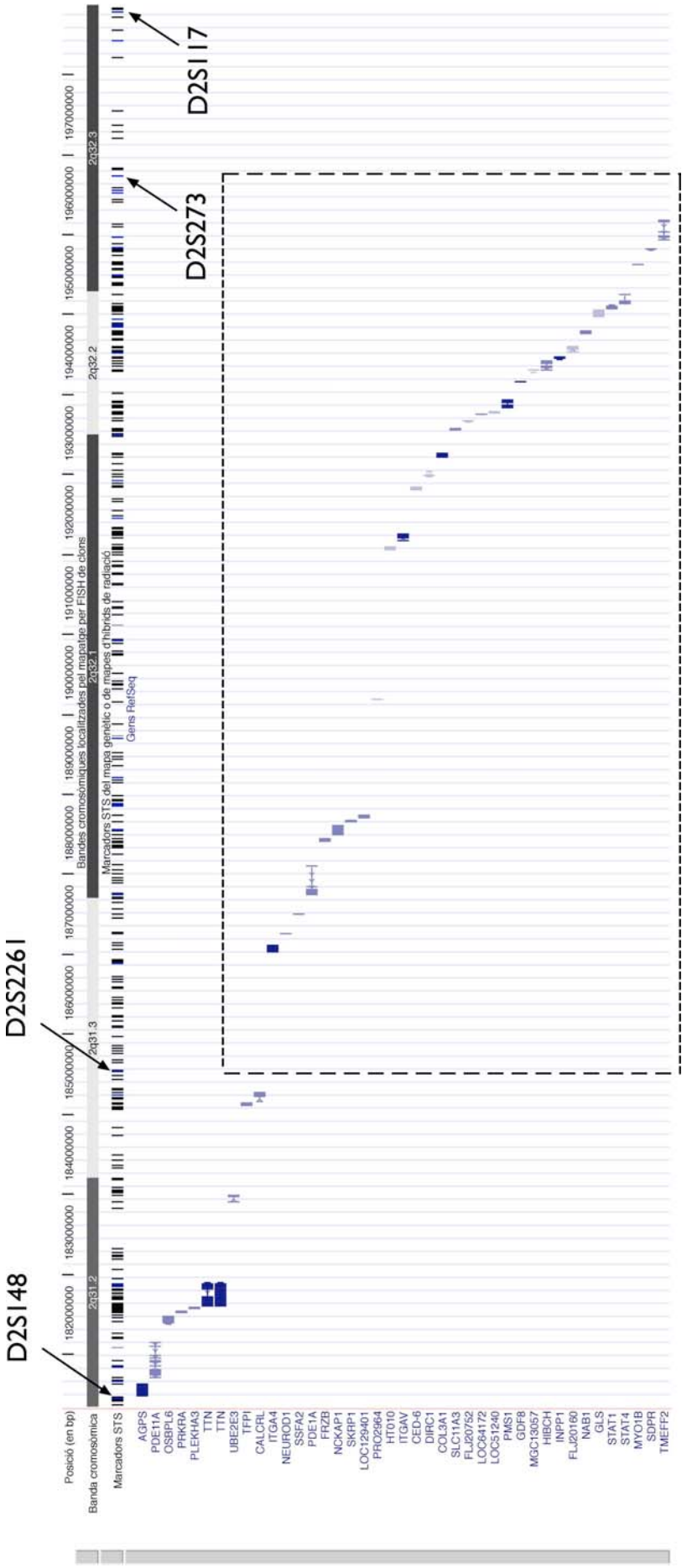



**Figura 9** | Mapa físic del locus RP26 a l'ordenament de l'abril del 2002. S'hi indiquen els marges originals –marcadors D2S148 i D2S117– i els marges nous, un cop refinat l'interval de cosegregació –marcadors D2S2261 i D2S273–. El requadre delimita la posició actual del locus. Adaptada d'UCSC Genome Bioinformatics.

## 2.3. Anàlisi de gens candidats en el locus RP26

Conjuntament a l'anàlisi mutacional per seqüenciació dels gens de la regió crítica –seleccionats en base a criteris purament posicionals–, vam analitzar altres gens candidats atenent a criteris funcionals, alguns dels quals es trobaven fora de la regió crítica, en d'altres zones del locus RP26. Aquest grup de candidats funcionals estava format pels gens *FRZB* i *NEUROD1*, inclosos al llistat de candidats de la regió crítica, i pels gens *ITGAV* i *GLS*, com a candidats externs. La consideració de candidats externs a la regió crítica estava justificada per diversos elements, entre els quals: (*a*) pel fet que la informació posicional derivada de la família xinesa es basava en una anàlisi de cosegregació i no en una anàlisi de lligament i, per tant, la cosegregació, de significació estadística menor, podia ser espúria; i (*b*) en cas que la cosegregació fos real, no necessàriament les dues famílies havien de tenir alterat el mateix gen i, en aquest cas, com en l'anterior, la regió crítica tampoc seria tal.

### 2.3.1. *FRZB*

El gen *FRZB* (*SFRP3*) codifica una proteïna secretada relacionada amb el receptor frizzled de *Drosophila*. En vertebrats, s'han trobat diversos gens d'aquesta família (*secreted frizzled-related proteins, SFRP*), que codifiquen, tots ells, proteïnes secretades que competeixen amb els receptors frizzled per la unió dels lligands Wnt, i per tant modulen el primer pas d'aquestes vies de senyalització. Inicialment, el gen *FRZB* ens va cridar l'atenció com a candidat funcional perquè la via de senyalització en què intervé –la via de Wnt/frizzled– havia estat implicada en la determinació de la polaritat en diversos sistemes, entre els quals, la retina de *Drosophila*. A banda d'això, hi havia evidències que dos dels gens d'aquesta família en vertebrats s'expressaven en teixits retinals: el gen *SFRP2*, específicament en la capa nuclear interna de la retina, i el gen *SFRP5*, a l'epiteli pigmentari de la retina a banda d'expressar-se també al pàncrees (Chang et al. 1999). A més, també s'havia observat que els nivells d'mRNA de diversos gens d'aquesta família, inclòs *FRZB* (*SFRP1*, *SFRP2*, *SFRP3* i *SFRP5*), estaven alterats en retines afectades de retinitis pigmentària i que, per tant, modificacions en la via Wnt/frizzled podien participar en la patogènesi d'aquesta malaltia (Jones et al. 2000a, b). Tot aquest conjunt de dades va fer que consideréssim *FRZB* com a candidat funcional. Més tard, s'ha determinat que mutacions en el gen *FZD4* (*frizzled homolog 4*), que codifica un receptor homòleg a la proteïna frizzled de *Drosophila*, alteren la vascularització perifèrica de la retina i causen una forma de vitreoretinopatia exudativa familiar (Robitaille et al. 2002). Recentment, s'ha determinat que la proteïna SFRP1 participa en la diferenciació de la retina del pollet, on promou la generació de cèl·lules ganglionars i de cons (Esteve et al. 2003).

El gen *FRZB* té 6 exons, dels quals inicialment vam poder analitzar-ne cinc, perquè a l'ordenament del genoma de desembre del 2001 (*Human Draft Genome Assembly* UCSC hg10), tota la regió 5' del gen, que contenia el primer exó, estava situada en un buit de seqüència genòmica; el mateix buit que abarcava també els 7 darrers exons del gen *NCKAP1* (FIGURES 5 I 10). Com a

conseqüència d'això, no podíem dissenyar oligonucleòtids que ens permetessin amplificar tota la regió codificant d'aquest primer exó i les seqüències d'empalmament flanquejants. Finalment, gràcies a les dades de la base privada Celera i a l'actualització de les dades de la base pública, vam poder seqüenciar el gen completament, però no vam identificar cap canvi patogènic, ni a la regió codificant ni a les seqüències donadores i acceptores d'empalmament.

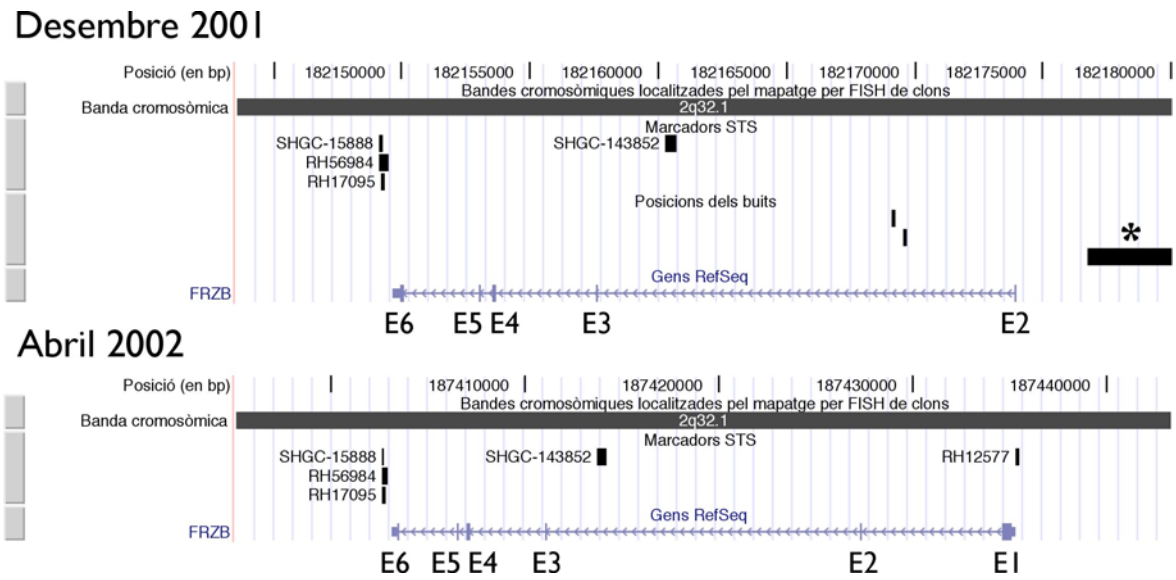



**Figura 10** | Estructura genòmica del gen *FRZB* situada als ordenaments del genoma humà del desembre del 2001 i de l'abril del 2002. A l'ordenament del desembre del 2001, el primer exó (E1) del gen *FRZB* no es podia col·locar a la seqüència genòmica per l'existència d'una discontinuïtat (marcada amb un asterisc) que abarcava també els darrers set exons del gen *NCKAP1*, situat a 3' de *FRZB* (FIGURA 5). A l'actualització de l'abril del 2002, aquesta discontinuïtat de seqüència va ser corregida. Mapes obtinguts d'UCSC Genome Bioinformatics.

### 2.3.2. *NEUROD1*

El gen *NEUROD1* codifica un factor de transcripció de la família de factors de transcripció amb un domini d'unió al DNA de tipus hèlix-llaç-hèlix. Aquest tipus de factors de transcripció estan implicats en fenòmens de determinació i diferenciació cel·lular en diversos teixits, entre els quals la retina (Acharya et al. 1997, Morrow et al. 1999). En concret, BETA2/NEUROD1 va ser caracteritzat com un factor necessari per a l'expressió del gen de la insulina i, per una altra banda, com un factor que intervenia en la determinació de destí cel·lular de tipus neuronal a partir de progenitors epidèrmics; d'aquí, la denominació dual d'aquesta proteïna. L'expressió ectòpica de *NeuroD1* al neuroepiteli retinal del pollet produeix una retina amb tres capes de cèl·lules fotoreceptores, en comptes de dues (Yan i Wang 1998). Els individus d'una de les soques de ratolí *knockout* per a *BETA2/NeuroD1* desenvolupen una diabetis severa i moren poc després de néixer, a causa d'una reducció dramàtica del nombre de cèl·lules beta i per la conseqüent incapacitat de generar illots pancreàtics madurs (Naya et al. 1997). En humans, mutacions en heterozigosi en aquest gen han estat associades al desenvolupament de diabetis mellitus de tipus 2 (Malecki et al. 1999).

Treballs duts a terme en cultius de cèl·lules retinals mostraven que la manca de BETA2/NEUROD1 provocava la mort dels bastons, incrementava la població de neurones bipolars i de cèl·lules glials, i retardava la diferenciació de les cèl·lules amacrines (Morrow et al. 1999). Recentment, ha estat estudiat un model de ratolí *knockout* per *BETA2/NeuroD1* que sobreviu fins a edat adulta i que, per tant, permet analitzar l'efecte de la delecció d'aquest gen sobre la funció retinal. El ratolí homozigot per la deleció mostra una reducció del 50% del nombre de bastons als 2-3 mesos d'edat, i una absència total de resposta electroretinogràfica més enllà dels 9 mesos. Als 18 mesos d'edat, les retines d'aquests ratolins presenten una manca total de fotoreceptors (Pennesi et al. 2003a). En aquest treball, també es demostra un increment de l'apoptosi a la retina d'aquest model, que té un pic al tercer dia postnatal i es manté al llarg de l'edat adulta del ratolí. Així, BETA2/NeuroD1 actua com un factor de supervivència i manteniment, a banda de la seva funció en la diferenciació terminal dels fotoreceptors. En aquest model, la pèrdua de *BETA2/NeuroD1* produeix un procés de degeneració, tant de bastons com de cons, anàleg a la degeneració retinal observada als pacients de retinitis pigmentària i, basant-nos en aquestes evidències, vam considerar aquest gen com a candidat funcional.

El gen *NEUROD1* presenta 2 exons. L'exó 1 conté part de la regió 5'UTR i l'exó 2 conté tota la regió codificant del gen. La seqüenciació d'aquest gen en els membres de la família P2 no ens va permetre detectar cap variant patogènica associada a la retinitis pigmentària. Els cinc pacients de la família són homozigots pel polimorfisme A45T (FIGURA 11).

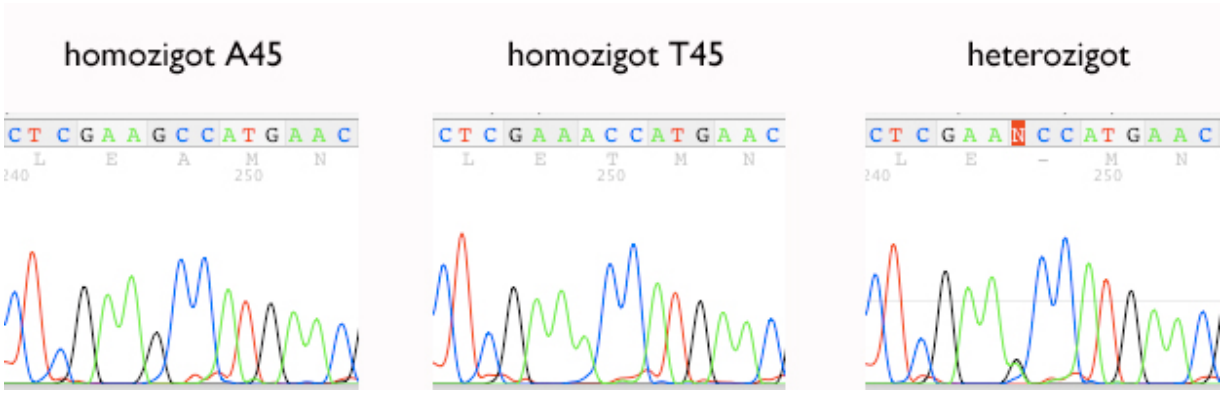



**Figura 11** | Genotips pel polimorfisme A45T del gen *NEUROD1* en la família P2. Els pacients RP són homozigots A/A (Thr en la posició 45); els pares i II.3, II.8, II.11 i II.12 són heterozigots; i II.4 i II.7 són homozigots G/G (Ala en la posició 45).

La presència d'aquest canvi, en heterozigosi, ha estat documentada tant en població control com en individus amb diabetis mellitus de tipus 2, en estudis realitzats sobre aquesta patologia, i s'ha demostrat que no està associat a la malaltia, almenys en heterozigosi (Malecki et al. 1999). Un altre estudi però, sosté que hi ha una associació entre la presència en homozigosi del polimorfisme A45T i el desenvolupament de diabetis mellitus de tipus 2, perquè en població control no s'ha trobat el polimorfisme en homozigosi i si que ha estat identificat en pacients diabètics japonesos (Kanatsuka et

al. 2002). En el nostre cas, tenint en compte la freqüència poblacional d'aquest polimorfisme i l'evidència que els pacients diabètics japonesos homozigots pel polimorfisme no presenten retinitis pigmentària, vam descartar-lo com a causal de la malaltia a la família P2.

### 2.3.3. *ITGAV*

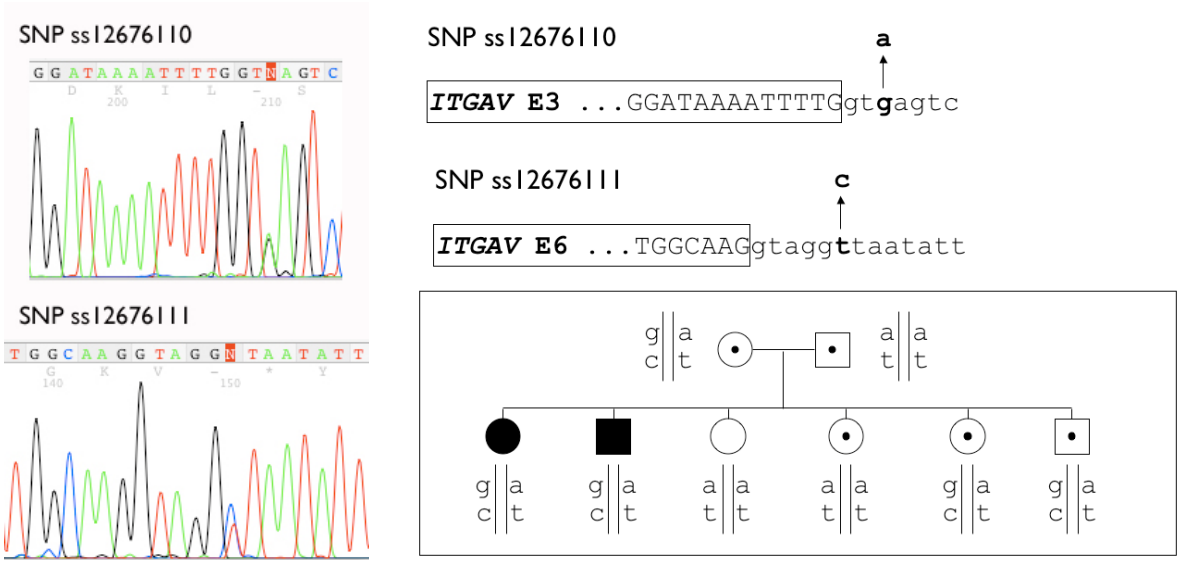



**Figura 12** | Exemple de l'anàlisi de dos dels SNP identificats en el gen *ITGAV* en el curs d'aquest estudi. Els SNP ss12676110 i ss12676111 van ser caracteritzats en la seqüència intrònica a 3' dels exons E3 i E6, respectivament. Els pacients de la família P2 són heterozigots per aquests SNP (vegeu cromatogrames), així com, per tots els altres SNP analitzats en el gen *ITGAV*.

El gen *ITGAV* codifica la subunitat integrina β-V que forma part del receptor de vitronectina (format per βvβ5, o bé βvβ3). A l'epiteli pigmentari de la retina, el receptor integrina βvβ5 és necessari per a que les cèl·lules de l'epiteli pigmentari puguin dur a terme la fagocitosi dels segments externs dels fotoreceptors. El receptor βvβ5 intervé en el reconeixement del segment extern per part de l'epiteli pigmentari, previ a la fagocitosi. En estudis duts a terme en cocultius de cèl·lules d'epiteli pigmentari i fotoreceptors, el bloqueig del receptor de vitronectina amb anticossos contra βvβ5 va reduir la fagocitosi un 85% (Finnemann et al. 1997). El procés de fagocitosi dels segments externs dels fotoreceptors és essencial per a la viabilitat de la funció retinal, com posa de manifest l'alteració del gen *MERTK*, que codifica una proteïna de l'epiteli pigmentari implicada en la fagocitosi i que és responsable de la degeneració retinal, de la soca de rata del *Royal College of Surgeons* (RCS), i d'una forma de retinitis pigmentària autosòmica recessiva en humans. Totes aquestes dades situaven *ITGAV* com un excel·lent candidat funcional.

El gen *ITGAV* té 30 exons, que vam analitzar per seqüenciació directa en els individus de la família P2. No vam identificar cap variant patogènica i, únicament, vam localitzar un polimorfisme d'un únic nucleòtid (SNP), ja documentat a les bases de dades (rs3816386), a la seqüència intrònica que flanqueja l'exó E19, i vam caracteritzar 5 nous SNP a les seqüències intròniques que flanquegen

els exons E1 (ss12676109), E3 (ss12676110), E6 (ss12676111), E10 (ss12676112), i E16 (ss12676113), que vam trametre a la base de dades pública d'SNP, NCBI dbSNP (FIGURA 12).

2.3.4. *GLS*

El quart candidat funcional que vam analitzar va ser el gen *GLS*, que codifica l'enzim glutaminasa. A la retina, el neurotransmissor glutamat és produït per diverses vies, entre les quals la formació a partir de glutamina, per acció de la glutaminasa. Entre les poblacions de neurones retinals que produeixen glutamat mitjançant l'enzim glutaminasa, hi ha les cèl·lules ganglionars i les cèl·lules bipolars (Takatsuna et al. 1994). Sembla que, a la retina, la neurotransmissió "ràpida" entre fotoreceptors, cèl·lules bipolars i cèl·lules ganglionars funciona mitjançant el glutamat, i que les vies laterals de neurotransmissió funcionen a través dels neurotransmissors inhibidors GABA i glicina. D'altra banda, existeixen evidències experimentals que, en retines en degeneració, es produeix una alteració del metabolisme del glutamat prèvia a la pèrdua de fotoreceptors (Fletcher 2000). Per tant, la possible connexió entre una alteració neuroquímica i el procés de degeneració retinal ens va fer considerar *GLS* com a candidat funcional per al locus RP26.

El gen GLS conté 18 exons, els quals van ser analitzats per seqüenciació en els individus de la família P2. No vam identificar cap mutació ni en la seqüència codificant ni en els senyals d'empalmament que flanquegen els exons. La identificació de dos marcadors SNP –rs2355571 i rs2883713, en les seqüències intròniques a 5' dels exons E7 i E10, respectivament–, pels quals els pacients de la família P2 eren heterozigots, remarcava que *GLS* no era el gen RP26. Si teníem en compte la consanguinitat de la família P2, el gen responsable de la malaltia havia d'estar localitzat en una regió on es mantingués l'homozigositat, i el gen *GLS* –a l'igual que anteriorment el gen *ITGAV*– no ho estava.

§

## 2.4. Mapatge d'homozigositat en el locus RP26

Segons les dades derivades de l'anàlisi de lligament inicial, el locus RP26 comprenia un interval de cosegregació d'11 cM entre els marcadors D2S148 i D2S117, tenint en compte les distàncies del mapa genètic de Généthon de 1994. A l'anàlisi inicial, també es va posar de manifest que el locus RP26 contenia una putativa regió d'homozigositat de 7 cM, entre els marcadors D2S148 i D2S161 (Bayés et al. 1998).

Anàlisis posteriors van determinar que el marcador D2S350, descrit inicialment com a homozigot, presentava realment un patró heterozigot en els pacients de la família P2 (FIGURA 13). Un cop situats els 8 microsatèl·lits de l'anàlisi de lligament inicial a l'ordenament del genoma de desembre de 2001 i, tenint en compte les noves dades del marcador D2S350, la putativa regió

d'homozigositat quedava restringida entre els marcadors D2S350 i D2S161, i contenia tres dels marcadors homozigots –D2S118, D2S389, D2S318– prop del marge telomèric de la regió. L'altre marcador homozigot –D2S364– quedava situat al bell mig de la regió crítica esmentada anteriorment. Calia, doncs, incrementar substancialment la densitat de marcadors analitzats dins del locus per tal de discernir quin dels dos blocs contenia una regió d'homozigositat, en la qual cercar el gen *RP26*.

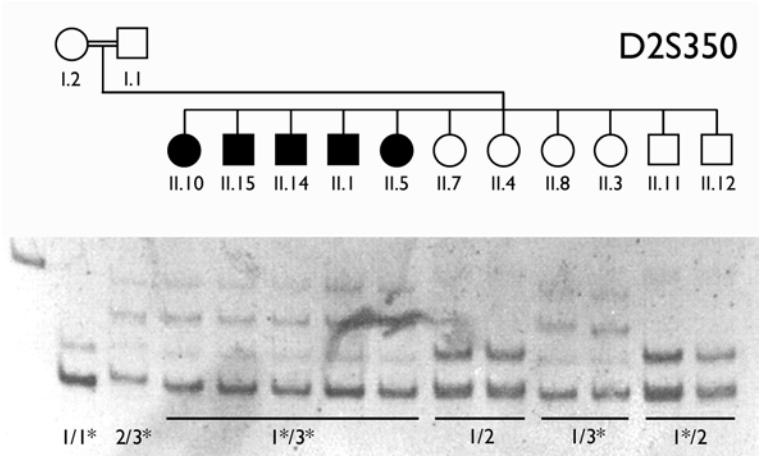



**Figura 13** | Patró al·lèlic del microsatèl·lit D2S350 a la família P2. Sota el patró es mostra la representació numèrica dels al·lels de cada individu. L'asterisc denota l'al·lel lligat a la malaltia.

2.4.1. Anàlisi de marcadors microsatèl·lit

A banda dels marcadors microsatèl·lits emprats per a caracteritzar la posició de les recombinacions que flanquegen el locus, en aquest estudi han estat analitzats 12 microsatèl·lits addicionals. Les posicions d'aquests marcadors, i el genotip dels pacients de la família P2, s'indiquen, a la TAULA 1.

**Taula 1 • Marcadors microsatèl·lit analitzats a 2q31.2–q32.3**

| Marcador | Nom | Situació (a l'ordenament NCBI Build 32) | Genotip als pacients |
|---|---|---|---|
| **D2S148** | **AFM200WA11** | chr2: 176902551–176902745 | Recombinant |
| D2S324 | AFM263XE1 | chr2: 179620249–179620637 | Recombinant |
| D2S2978 | ATA44H04 | chr2: 179788236–179788348 | Recombinant |
| D2S2261 | AFMB072WG1 | chr2: 180180111–180180442 | Recombinant |
| D2S2310 | AFMB355XD5 | chr2: 180849531–180849858 | No informatiu |
| **D2S364** | **AFM303YA9** | chr2: 181717611–181718002 | Homozigot |
| **D2S350** | **AFM292WD1** | chr2: 182532022–182532389 | Heterozigot |
| D2S2273 | AFMB297XC1 | chr2: 182825486–182825797 | Heterozigot |
| D2S2281 | AFMB310XF5 | chr2: 182863973–182864335 | Heterozigot |
| D2S2366 | AFMA057VG9 | chr2: 183175061–183175459 | No informatiu |
| D2S1391 | GATA65C03 | chr2: 183675508–183675835 | Homozigot |
| D2S1361 | CHLC.GATA14E05 | chr2: 184899775–184900087 | Heterozigot |
| D2S152 | AFM207XG1 | chr2: 186914674–186915071 | No informatiu |
| D2S2262 | AFMB082YE1 | chr2: 189985176–189985493 | Heterozigot |
| **D2S118** | **AFM066XC1** | chr2: 190309134–190309312 | Homozigot |
| **D2S389** | **AFM333WF9** | chr2: 190347305–190347628 | Homozigot |
| D2S1775 | CHLC.GATA71C07 | chr2: 190402485–190402614 | Homozigot |
| D2S2246 | AFMB007WC1 | chr2: 191085514–191085746 | Heterozigot |
| **D2S318** | **AFM105XC1** | chr2: 191278428–191278834 | Homozigot |
| **D2S161** | **AFM224ZF4** | chr2: 191431730–191432052 | Heterozigot |
| D2S280 | AFM155YE1 | chr2: 191557728–191557932 | Homozigot |
| D2S315 | AFM081YG5 | chr2: 192076101–192076267 | Homozigot |
| D2S273 | UT5048 | chr2: 192288946–192289538 | Recombinant |
| **D2S117** | **AFM065YF11** | chr2: 194331362–194331699 | Recombinant |

En negreta s'indiquen els microsatèl·lits emprats a l'anàlisi de lligament (Bayés et al. 1998).

La primera conclusió que vam extreure de l'anàlisi d'aquests marcadors va ser que, únicament amb la densitat d'informació que proporcionaven, no podíem determinar de manera concloent l'existència d'una regió d'homozigositat. De l'homozigositat inicial en quedava, per una banda, el microsatèl·lit D2S364 –central en una regió d'aproximadament 2.5 Mb entre D2S2261 i D2S350– i, per l'altra, els microsatèl·lits D2S118 i D2S389, juntament amb un nou microsatèl·lit, D2S1775, pel qual els pacients també eren homozigots (FIGURA 14) –en un segment d'1.1 Mb, entre D2S2262 i D2S2246–.

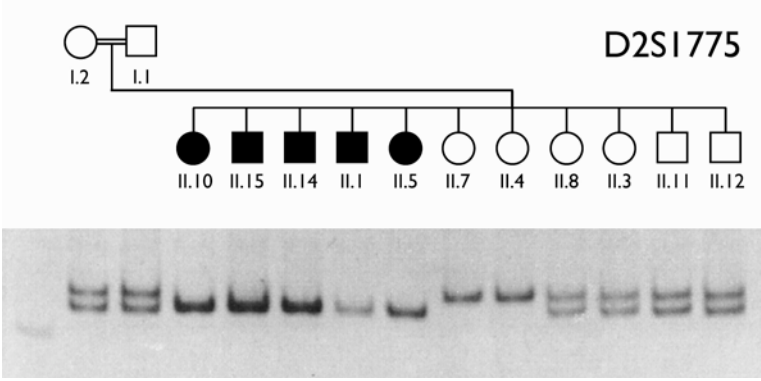
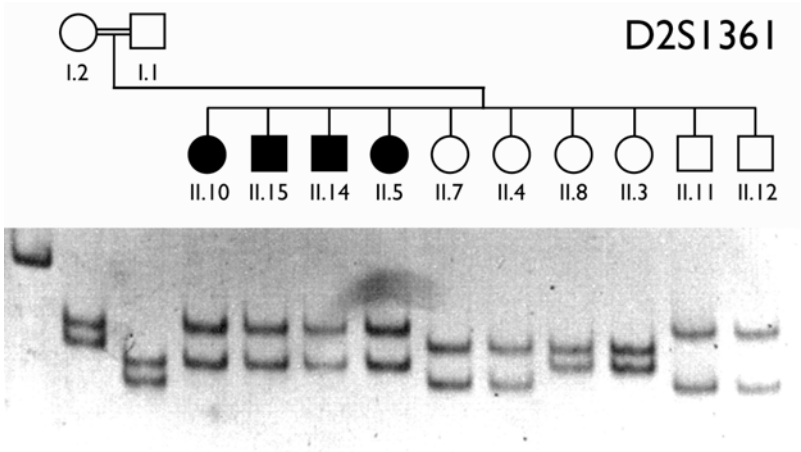



**Figura 14** | Patró al·lèlic del microsatèl·lit D2S1775 en la família P2. D2S1775 està situat molt proper als microsatèl·lits D2S118 i D2S389, que formaven part de la pressumpta regió d'homozigositat de 7 cM descrita originalment en (Bayés et al. 1998).

El microsatèl·lit D2S318, un dels altres marcadors emprats en l'anàlisi de lligament, que era homozigot, quedava separat de D2S118 i D2S389 pel microsatèl·lit D2S2246; heterozigot (FIGURA 16). Hi havia altres regions amb marcadors homozigots, com la que contenia D2S1391, i la de D2S280 i D2S315 però, en resum, el que es desprenia de l'anàlisi de nous marcadors microsatèl·lits era que necessitavem molts més marcadors per a poder discernir si en alguna d'aquestes regions existia un segment en què es mantingués l'homozigositat o si, per contra, ens trobàvem davant d'una regió on, aparentment, se succeien els marcadors homozigots i heterozigots (FIGURES 15 I 16). Per dur a terme aquesta tasca –un mapatge d'homozigositat precís–, ens calia un tipus de marcador polimòrfic que ens permetés cobrir, amb una certa densitat, les 12 Mb del locus, i això només era possible a partir de la localització i anàlisi de polimorfismes d'un únic nucleòtid (SNP).

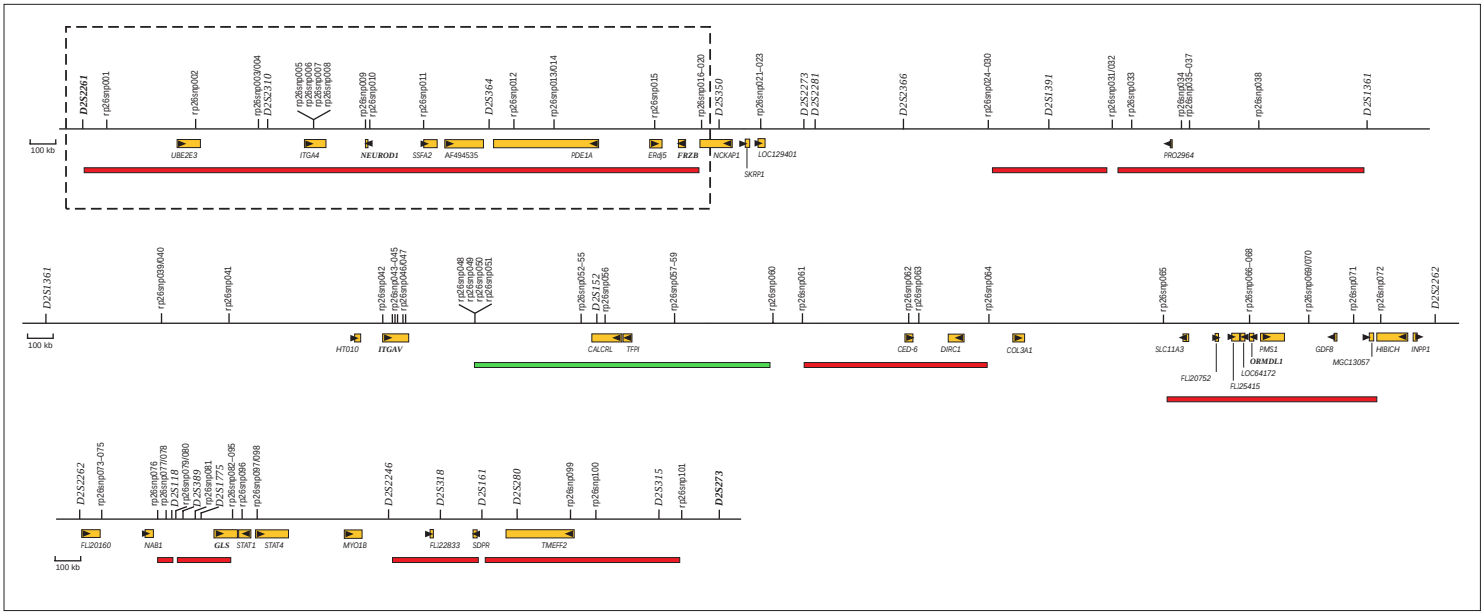



| Microsatèl·lit D2S1361 analitzat en la família P2. Aquest és un dels múltiples punts del locus pels quals els pacients de la família són heterozigots. Es pot veure clarament la cosegregació de la malaltia amb el marcador: els pares són heterozigots per quatre al·lels diferents (1,2 el pare; 3,4 la mare), els malalts són heterozigots 2,4; dues germanes no portadores són heterozigotes 1,3; dues germanes portadores per part de pare són heterozigotes 2,3; i dos germans portadors per part de mare són heterozigots 1,4.

2.4.2. Anàlisi de polimorfismes d'un únic nucleòtid (SNP)

El mapatge d'homozigositat mitjançant SNP va ser realitzat en tres tandes successives, en les quals vam analitzar un total de 101 SNP. Aquestes dades, juntament amb la informació proporcionada pels 24 microsatèl·lits analitzats, ens van permetre disposar d'informació sobre l'homozigositat/heterozigositat de marcadors situats a intervals de menys de 500 kb al llarg de tot el locus RP26 (FIGURA 16; TAULA 2). D'aquesta manera, vam poder localitzar una regió de 2.5 Mb –entre D2S2261 i rs7510– per la qual els pacients de la familia P2 eren homozigots per tots els marcadors (FIGURA 16; TAULA 2). Havíem identificat la regió d'homozigositat en la qual hauríem de cercar el gen responsable de la malaltia. Cal remarcar que aquesta regió estava situada dins de la regió crítica de 3 Mb –definida pel solapament dels intervals de cosegregació de la familia P2 i de la família retinítica xinesa, entre D2S2261 i D2S2366–, de la qual ja havíem seqüenciat 12 gens.

| Mapatge d'homozigositat en el locus RP26. S'ha representat l'interval de cosegregació entre els microsatèl·lits D2S2261 i D2S273 (12 Mb). Sobre la línia s'han indicat els marcadors microsatèl·lit i SNP analitzats. Els SNP s'han numerat de manera consecutiva: la TAULA 2 conté les equivalències entre aquesta numeració i la denominació de cadascun dels SNP. Les barres vermelles representen zones d'homozigositat, i la barra verda una zona de marcadors no informatius. Els gens s'han indicat amb rectangles grocs i les fletxes representen el sentit de la transcripció. El rectangle discontinu remarca la regió candidata de 2.5 Mb identificada gràcies al mapatge d'homozigositat.


D2S2261
rp26mp001
rp26mp002
rp26mp003/004
D2S2310
rp26mp005
rp26mp006
rp26mp007
rp26mp008
rp26mp009
rp26mp010
rp26mp011
D2S364
rp26mp012
rp26mp013/014
rp26mp015
rp26mp016-020
D2S350
rp26mp021-023
D2S2273
D2S2281
D2S2366
rp26mp024-030
D2S1391
rp26mp031/032
rp26mp033
rp26mp034
rp26mp035-037
rp26mp038
D2S1361
100 kb
UBE2E3
ITGA4
NEUROD1
SSFA2
AF494535
PDE1A
ERdj5
FRZB
NCKAP1
LOC129401
SKRP1
PRO2964
D2S1361
rp26mp039/040
rp26mp041
rp26mp042
rp26mp043-045
rp26mp046/047
rp26mp048
rp26mp049
rp26mp050
rp26mp051
rp26mp052-55
D2S152
rp26mp056
rp26mp057-59
rp26mp060
rp26mp061
rp26mp062
rp26mp063
rp26mp064
rp26mp065
rp26mp066-068
rp26mp069/070
rp26mp071
rp26mp072
D2S2262
100 kb
HT010
ITGAV
CALCRL
TFR
CED-6
DIRC1
COL3A1
SLC11A3
FLI20752
PMS1
ORMDL1
LOC64172
FLI25415
GDF9
MGC13057
HIBCH
INPP1
D2S2262
rp26mp073-075
rp26mp076
rp26mp077/078
D2S118
rp26mp079/080
D2S389
rp26mp081
D2S1775
rp26mp082-085
rp26mp086
rp26mp087/088
D2S2246
D2S318
D2S161
D2S280
rp26mp099
rp26mp100
D2S315
rp26mp101
D2S273
FLI20160
NAB1
GLS
STAT1
STAT4
MYO18
FLI22833
SDPR
TMEFF2
100 kb

## Taula 2 • Marcadors microsatèl·lit i SNP analitzats a 2q31.2–q32.3

| Microsatèl·lits | SNP | Situació (a l'ordenament NCBI Build 32) | Locus | Genotip als pacients de la família P2 |
|---|---|---|---|---|
| D2S148; AFM200WA11 | | chr2: 176902551–176902745 | | Rec |
| D2S324; AFM263XE1 | | chr2: 179620249–179620637 | | Rec |
| D2S2978; ATA44H04 | | chr2: 179788236–179788348 | | Rec |
| D2S2261; AFMB072WG1 | | chr2: 180180111–180180442 | | Rec |
| | (01) rs1002207; TSC0036660 | chr2: 180237706 | | Ni |
| | (02) rs1949453; TSC1026876 | chr2: 180594170 | *UBE2E3* | **H** |
| | (03) rs2077783; TSC0046543 | chr2: 180834855 | | **H** |
| | (04) rs720453; TSC0046544 | chr2: 180834909 | | **H** |
| D2S2310; AFMB355XD5 | | chr2: 180849531–180849858 | | Ni |
| | (05) rs155100; TSC0531809 | chr2: 181033165 | *ITGA4* | **H** |
| | (06) rs155101; TSC0531810 | chr2: 181033223 | *ITGA4* | **H** |
| | (07) rs155102 | chr2: 181033525 | *ITGA4* | **H** |
| | (08) rs155103 | chr2: 181033821 | *ITGA4* | **H** |
| | (09) rs1801262 | chr2: 181226533 | *NEUROD1* | **H** |
| | (10) rs2583016 | chr2: 181228296 | *NEUROD1* | **H** |
| | (11) rs1157595; TSC0326375 | chr2: 181432646 | | **H** |
| D2S364; AFM303YA9 | | chr2: 181717611–181718002 | *PDE1A* | **H** |
| | (12) rs3754929 | chr2: 181771662 | *PDE1A* | **H** |
| | (13) rs1438065; TSC0655111 | chr2: 181902071 | *PDE1A* | **H** |
| | (14) rs833152; TSC0655110 | chr2: 181902179 | *PDE1A* | Ni |
| | (15) rs288254; TSC0681111 | chr2: 182291705 | *ERdj5* | Ni |
| | (16) rs7510 | chr2: 182472727 | *NCKAP1* | Het |
| | (17) ss12676100 | chr2: 182476461 | *NCKAP1* | Het |
| | (18) ss12676101 | chr2: 182476476 | *NCKAP1* | Het |
| | (19) ss12676102 | chr2: 182482636 | *NCKAP1* | Het |
| | (20) rs2271671 | chr2: 182510096 | *NCKAP1* | Het |
| D2S350; AFM292WD1 | | chr2: 182532022–182532389 | *NCKAP1* | het |
| | (21) ss12676103 | chr2: 182678153 | LOC129401 | het |
| | (22) ss12676104 | chr2: 182678197 | LOC129401 | het |
| | (23) ss12676105 | chr2: 182678220 | LOC129401 | het |
| D2S2273; AFMB297XC1 | | chr2: 182825486–182825797 | | het |
| D2S2281; AFMB310XF5 | | chr2: 182863973–182864335 | | het |
| D2S2366; AFMA057VG9 | | chr2: 183175061–183175459 | | ni |
| | (24) rs826135; TSC1254035 | chr2: 183459257 | | ni |
| | (25) rs826134; TSC0476562 | chr2: 183459267 | | ni |
| | (26) ss12676106 | chr2: 183459509 | | het |
| | (27) rs826133; TSC0658245 | chr2: 183459526 | | ni |
| | (28) ss12676107 | chr2: 183459659 | | het |
| | (29) ss12676108 | chr2: 183459776 | | het |
| | (30) rs826132; TSC0658244 | chr2: 183459802 | | het |
| D2S1391; GATA65C03 | | chr2: 183675508–183675835 | | H |
| | (31) rs1443021; TSC0664031 | chr2: 183910062 | | het |
| | (32) rs1443022; TSC0664032 | chr2: 183910149 | | het |
| | (33) rs768352; TSC0076040 | chr2: 184012769 | | H |
| | (34) rs994653; TSC0317578 | chr2: 184193063 | | H |
| | (35) rs728534; TSC0064305 | chr2: 184484995 | | H |
| | (36) rs3046266; TSC1530131 | chr2: 184485293 | | ni |
| | (37) rs1366842; TSC0514415 | chr2: 184485321 | | ni |
| | (38) rs3731834 | chr2: 184486442 | | H |
| D2S1361; GATA14E05 | | chr2: 184899775–184900087 | | het |
| | (39) rs1016410; TSC0098979 | chr2: 185334459 | | het |
| | (40) rs1019430; TSC0223209 | chr2: 185334589 | | het |
| | (41) rs878845; TSC0212635 | chr2: 185580725 | | het |
| | (42) ss12676109 | chr2: 186138356 | *ITGAV* | het |
| | (43) ss12676110 | chr2: 186170238 | *ITGAV* | het |
| | (44) ss12676111 | chr2: 186181185 | *ITGAV* | het |
| | (45) ss12676112 | chr2: 186186304 | *ITGAV* | het |
| | (46) ss12676113 | chr2: 186202416 | *ITGAV* | het |
| | (47) rs3816386 | chr2: 186211713 | *ITGAV* | het |
| | (48) rs2018302; TSC0091486 | chr2: 186467891 | | het |
| | (49) rs1467990; TSC0384843 | chr2: 186468012 | | ni |
| | (50) rs2018314; TSC0085681 | chr2: 186468088 | | het |
| | (51) rs1000623; TSC0033486 | chr2: 186468142 | | het |
| | (52) rs840570; TSC1028023 | chr2: 186855755 | | ni |
| | (53) rs2308091 | chr2: 186856059 | | ni |
| | (54) rs84069 | chr2: 186856203 | | ni |
| | (55) rs84068 | chr2: 186856285 | | ni |
| D2S152; AFM207XG1 | | chr2: 186914674–186915071 | *CALCRL* | ni |
| | (56) rs1398061; TSC0576737 | chr2: 186941063 | *CALCRL* | ni |
| | (57) rs1464338; TSC0378776 | chr2: 187177069 | | ni |
| | (58) rs1464337; TSC0378775 | chr2: 187177070 | | ni |
| | (59) rs2176858; TSC1217605 | chr2: 187177146 | | ni |
| | (60) rs2033838; TSC1047282 | chr2: 187551834 | | het |
| | (61) rs2028374; TSC1091521 | chr2: 187666577 | | het |
| | (62) rs1354906; TSC0490640 | chr2: 188129892 | CED-6 | H |
| | (63) rs1007120: TSC0084553 | chr2: 188175289 | | ni |
| | (64) rs925825; TSC0241755 | chr2: 188431720 | | het |
| | (65) rs893407; TSC0165376 | chr2: 189033193 | | het |
| | (66) rs2289404 | chr2: 189318673 | *ORMDL1* | ni |
| | (67) rs6942; TSC0023616 | chr2: 189319070 | *ORMDL1* | H |
| | (68) rs3791767 | chr2: 189322574 | *ORMDL1* | ni |
| | (69) rs288817; TSC1078112 | chr2: 189529648 | | ni |
| | (70) ss12676114 | chr2: 189529690 | | H |
| | (71) rs998173; TSC0011271 | chr2: 189694293 | | H |
| | (72) rs3791789 | chr2: 189787488 | *HIBCH* | het |
| D2S2262; AFMB082YE1 | | chr2: 189985176–189985493 | | het |
| | (73) rs7721 | chr2: 190069246 | | het |

| | | | | |
|---|---|---|---|---|
| | (74) rs1128723 | chr2: 190069374 | | het |
| | (75) rs8962 | chr2: 190074316 | | ni |
| | (76) rs1468685; TSC0385949 | chr2: 190268436 | | het |
| | (77) rs1263136; TSC1488116 | chr2: 190298357 | | ni |
| | (78) rs1263125; TSC0446273 | chr2: 190303089 | | H |
| D2S118; AFM066XC1 | | chr2: 190309134–190309312 | | H |
| | (79) rs1263100; TSC0398872 | chr2: 190327138 | | het |
| | (80) rs1476896; TSC0398871 | chr2: 190327184 | | H |
| D2S389; AFM333WF9 | | chr2: 190347305–190347628 | | H |
| | (81) rs1882395; TSC0896111 | chr2: 190387668 | | H |
| D2S1775; GATA71C07 | | chr2: 190402485–190402614 | | H |
| | (82) rs3217036 | chr2: 190462343 | *GLS* | ni |
| | (83) rs2355571 | chr2: 190477428 | *GLS* | het |
| | (84) rs2883713 | chr2: 190491082 | *GLS* | het |
| | (85) rs3199237 | chr2: 190497863 | *GLS* | ni |
| | (86) rs3207595 | chr2: 190497884 | *GLS* | ni |
| | (87) rs3207596 | chr2: 190497887 | *GLS* | ni |
| | (88) rs3199238 | chr2: 190497893 | *GLS* | ni |
| | (89) rs3207597 | chr2: 190497917 | *GLS* | ni |
| | (90) rs1801893 | chr2: 190498994 | *GLS* | ni |
| | (91) rs3215259 | chr2: 190521768 | *GLS* | ni |
| | (92) rs1058589 | chr2: 190522002 | *GLS* | ni |
| | (93) rs1058590 | chr2: 190522011 | *GLS* | ni |
| | (94) rs1058591 | chr2: 190522017 | *GLS* | ni |
| | (95) rs1058592 | chr2: 190522020 | *GLS* | ni |
| | (96) rs1547550; TSC0429185 | chr2: 190548390 | *STAT1* | het |
| | (97) rs925847; TSC0241788 | chr2: 190600205 | *STAT4* | ni |
| | (98) rs3024891 | chr2: 190601614 | *STAT4* | het |
| D2S2246; AFMB007WC1 | | chr2: 191085514–191085746 | | het |
| D2S318; AFM105XC1 | | chr2: 191278428–191278834 | | H |
| D2S161; AFM224ZF4 | | chr2: 191431730–191432052 | | het |
| D2S280; AFM155YE1 | | chr2: 191557728–191557932 | *TMEFF2* | H |
| | (99) rs3738882 | chr2: 191751560 | *TMEFF2* | H |
| | (100) rs935367; TSC0352562 | chr2: 191842107 | | ni |
| D2S315; AFM081YG5 | | chr2: 192076101–192076267 | | H |
| | (101) rs717621; TSC0040380 | chr2: 192153621 | | het |
| D2S273; UT5048 | | chr2: 192288946–192289538 | | Rec |
| D2S117; AFM065YF11 | | chr2: 194331362–194331699 | | Rec |

Rec - recombinant; ni - no informatiu; H - homozigot; het -heterozigot. En negreta- regió d'homozigositat candidata.
ss# - SNP identificats en aquest treball. Els nombres entre parèntesi, que precedeixen el nom dels SNP a la columna SNP, fan referència a la numeració amb què són indicats a la FIGURA 16.

## 2.5. Anàlisi en profunditat de la regió candidata

Com hem vist a l'apartat anterior, l'estrategia del mapatge d'homozigositat duta a terme al llarg de la tardor del 2002 ens va permetre, finalment, definir una regió candidata de 2.5 Mb entre el microsatèl·lit D2S2261 i l'SNP rs7510. Ja hem apuntat que la primera aproximació per a la cerca del gen *RP26* –l'anàlisi de gens a la regió crítica definida pels intervals de cosegregació en les famílies P2 i xinesa– ens va portar a analitzar 12 gens de la regió crítica sense identificar cap variant patogènica. Els gens que havíem analitzat fins llavors eren els següents: *ZNF533*, eIF4G, *UBE2E3*, *PRO1316*, *NEUROD1*, *SSFA2*, *PDE1A*, PDIP5 (*ERdj5*), *FRZB*, *NCKAP1*, *SKRP1*, i MP44. Quan van ser analitzats, tots aquests gens es trobaven dins l'interval comprès entre D2S2261 i D2S2366 –límits de la regió crítica–. A finals de 2002, l'ordenament del genoma en curs era el que havia estat construït a partir de les dades de juny del 2002. En aquest ordenament, la posició relativa d'alguns dels gens de la regió havia canviat respecte a ordenaments anteriors. Així, els gens *ZNF533* i eIF4G, que a l'ordenament d'abril del 2002 estaven inclosos a la regió crítica, en direcció 3' del marcador D2S2261, a l'ordenament de juny del 2002 quedaven situats fora de la regió (FIGURA 17). A partir de l'ordenament de juny del 2002, i en els successius –novembre 2002, abril 2003 i juliol 2003– les posicions relatives dels gens de la regió no s'han modificat.

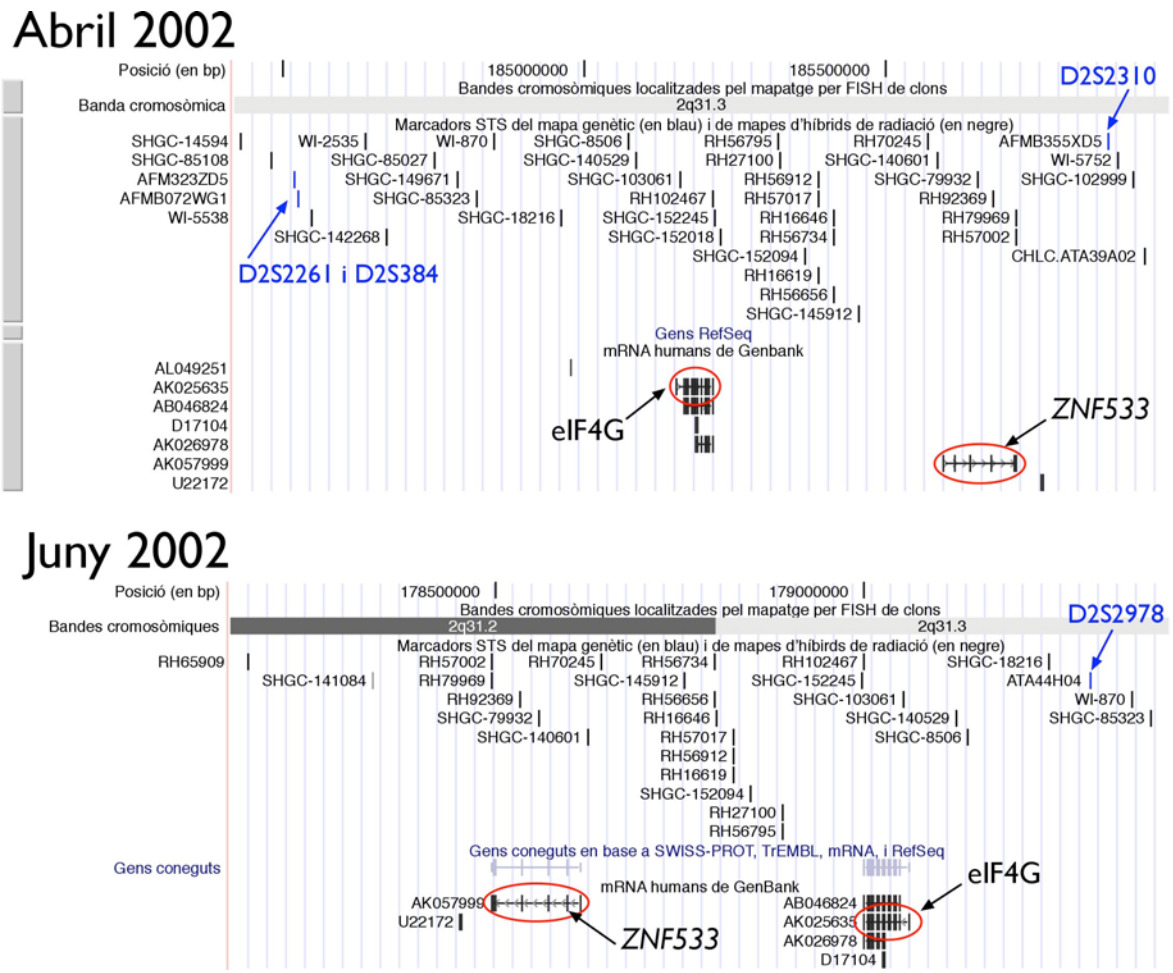



**Figura 17** | Posició dels gens eIF4G i *ZNF533* als ordenaments del genoma d'abril de 2002 i de juny de 2002. Al primer dels ordenaments, els dos gens es troben inclosos dins de la regió candidata, entre els microsatèl·lits D2S2261 (recombinant) i D2S2310. En canvi, a partir de l'ordenament de juny de 2002, els dos gens són fora de la regió candidata (D2S2978 és recombinant, vegeu la FIGURA 7). Noteu, per la posició dels marcadors i l'orientació dels gens, que el contig havia estat mapat invertit a l'ordenament d'abril de 2002. Mapes físics adaptats d'UCSC Genome Bioinformatics.

En tots aquests ordenaments, la regió candidata –definida per homozigositat, entre D2S2261 i rs7510– abarcava des del gen *UBE2E3* fins a *FRZB* (FIGURA 18). Tots ells incloïen els gens *UBE2E3* i *PRO1316* dins de la regió candidata, de la qual n'havien quedat exclosos a l'ordenament d'abril del 2002. El marge telomèric de la regió candidata, definit per l'SNP rs7510, està localitzat en la regió 3' no traduïda del gen *NCKAP1*, la pauta de lectura del qual està situada en la cadena reversa. Per tant, *NCKAP1* –que havíem inclòs al llistat de gens de la regió crítica i seqüenciat– quedava ara exclòs, tot sencer, de la regió candidata basant-nos en el criteri d'homozigositat. De la mateixa manera, quedaven també exclosos el gens *SKRP1* (*DUSP19*) i MP44 (*LOC129401*) que, com *NCKAP1*, havíem considerat inicialment com a candidats posicionals (FIGURA 18). Així doncs, la regió candidata de 2.5 Mb contenia 9 gens, dels quals 7 havien estat seqüenciats –*UBE2E3*, *PRO1316*, *NEUROD1*, *SSFA2*, *PDE1A*, PDIP5 (*ERdj5*), *FRZB*– i 2 no havien estat analitzats perquè no n'havíem detectat expressió a la retina –ITGA4 i PP1I (AF494535)–.

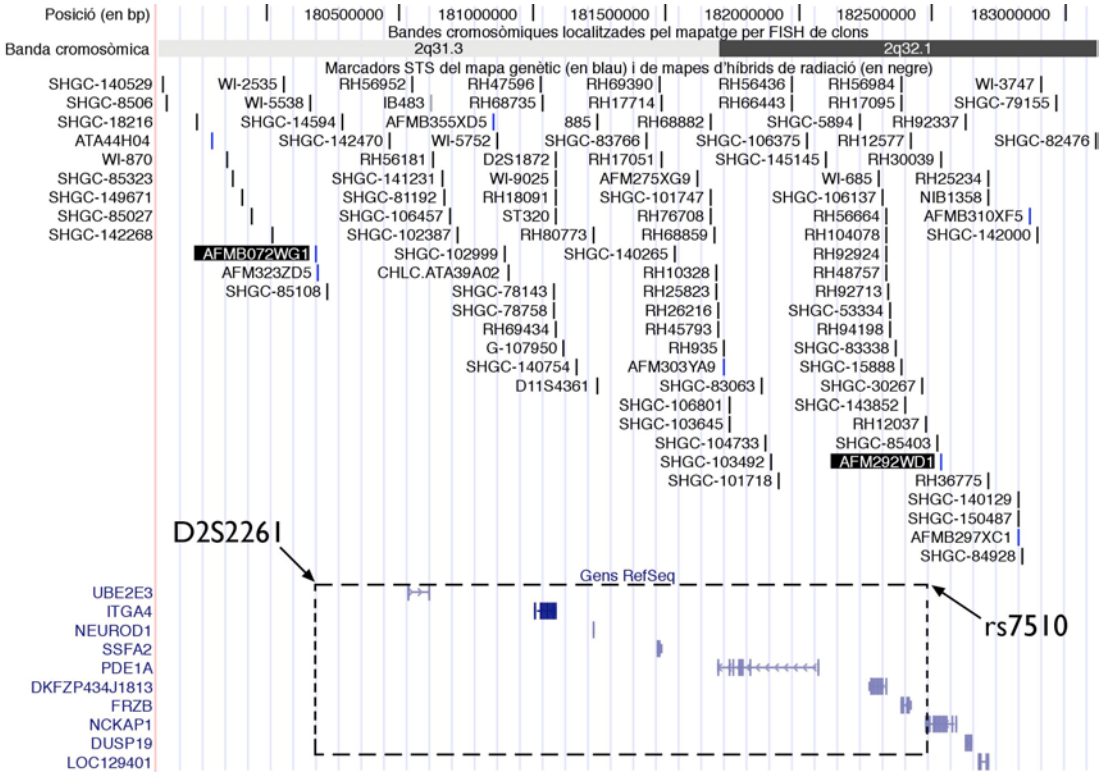



**Figura 18** | Mapa de la regió candidata a l'ordenament del genoma de novembre de 2002. La regió candidata, definida pel mapatge d'homozigositat, comprèn un segment de 2.5 Mb entre els marcadors D2S2261 i rs7510 (requadre discontinu). En aquesta regió hi havia 7 gens indexats en RefSeq (*UBE2E3*, *ITGA4*, *NEUROD1*, *SSFA2*, *PDE1A*, *ERdj5* –a la figura *DKFZP434J1813*– i *FRZB*) i 2 mRNAs que no es mostren (PRO1316 i AF494535). Mapa físic extret i adaptat d'UCSC Genome Bioinformatics.

### 2.5.1. Anàlisi mutacional dels gens que havien quedat descartats pel criteri d'expressió

Un cop definida la regió candidata, i tenint en compte que cap dels gens que havíem analitzat no era el gen *RP26*, vam decidir analitzar també els gens que inicialment havíem descartat en no detectar-ne expressió a la retina. Aquests gens eren *ITGA4* (precursor de la subunitat β de la integrina 4) i l'mRNA AF494535 (PP1I), que codificava una proteïna similar a la subunitat 1A reguladora (inhibidora) de la proteïna fosfatasa 1. Aquests gens contenen 28 i 6 exons, respectivament, que vam seqüenciar i en els quals no vam identificar cap mutació.

### 2.5.2. Seqüenciació d'exons addicionals de gens analitzats prèviament

En paral·lel, vam localitzar a GenBank noves entrades d'mRNA de gens que ja havíem analitzat. En concret, nous mRNA dels gens *UBE2E3* i *SSFA2* que contenien exons addicionals que no havíem seqüenciat en la primera anàlisi. El nou mRNA del gen *UBE2E3* contenia tres exons més, a banda dels tres ja seqüenciats, i per tant aquest gen passava a tenir-ne sis en total. Per la seva banda, el nou mRNA del gen *SSFA2* allargava considerablement l'ORF del gen, que quan l'havíem analitzat constava de cinc exons i que ara, amb l'addició de 13 exons extres, passava a tenir-ne 18 (FIGURA 19).

L'anàlisi d'aquests exons addicionals tampoc ens va permetre identificar cap variant patogènica en aquests dos gens.

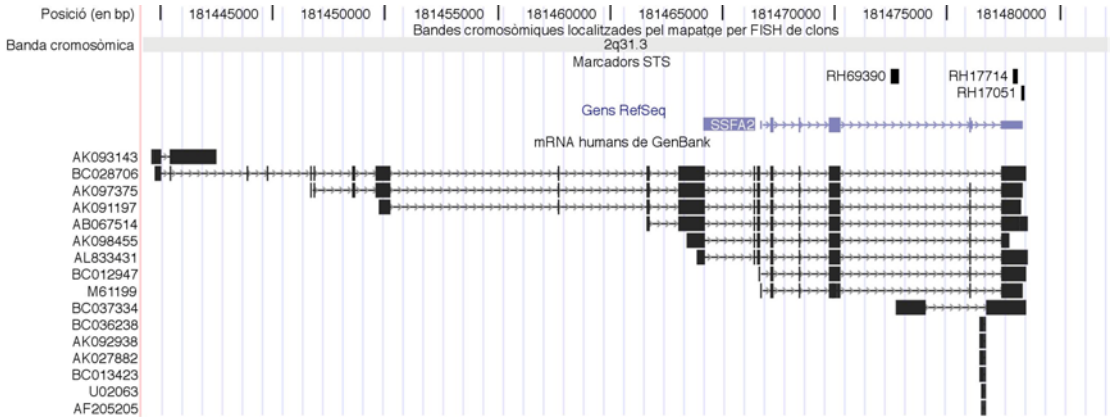



**Figura 19** | Localització genòmica del gen *SSFA2* a l'ordenament de novembre de 2002. Nous mRNA d'aquest gen, a l'igual que del gen *UBE2E3*, van allargar la regió codificant del gen amb la incorporació d'exons addicionals, que van ser analitzats per seqüenciació. Mapa físic extret i adaptat d'UCSC Genome Bioinformatics.

Un cop seqüenciats els exons addicionals d'*UBE2E3* i *SSFA2*, havíem analitzat completament les ORF de tots els gens de la regió candidata que, en aquell moment, estaven documentats en les bases de dades, i cap d'ells era el gen *RP26*. A partir d'aquí, ens vam plantejar la cerca *in silico* en la regió candidata d'EST o cDNA parcials amb l'objectiu de caracteritzar gens desconeguts i definir nous candidats posicionals per a la cerca del gen del locus RP26.

## 2.6. Cerca de nous gens a la regió candidata

La cerca d'EST *in silico* ens va permetre identificar una EST (no. de GenBank BE797822) i un cDNA parcial (no. de GenBank BC020465) no caracteritzats, que estaven localitzats propers l'un de l'altre, en el segment genòmic entre els gens *ITGA4 NEUROD1* (FIGURA 20). El cDNA incomplet havia estat aïllat d'una llibreria d'ull/retinoblastoma. Les dues seqüències estaven compostes per diversos exons, com s'evidenciava en comparar-les amb el DNA genòmic de la regió: en concret, el cDNA parcial contenia 6 exons i l'EST 7 exons. Ambdues presentaven una pauta de lectura oberta. Ens trobàvem, doncs, davant la possibilitat que, en aquesta regió, hi hagués un o dos gens nous, entre *ITGA4* i *NEUROD1*. La localització d'aquestes dues seqüències en el genòmic de la regió posava de manifest que eren molt properes tot i que, aparentment, no se solapaven, i indicava que possiblement pertanyien a un únic gen encara no caracteritzat. L'escrutini en detall ens va permetre identificar que, en un dels extrems, el cDNA i l'EST compartien 18 nucleòtids.

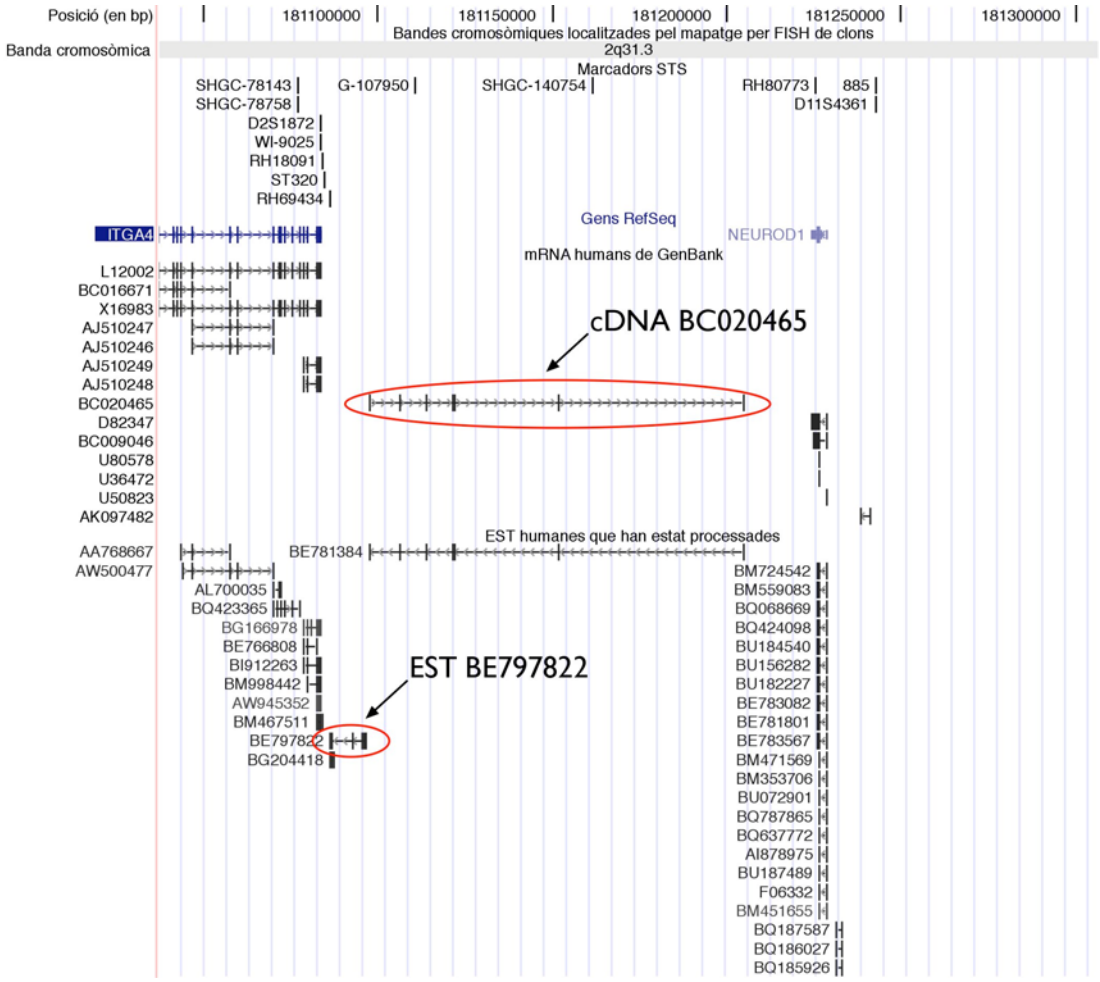



**Figura 20** | Posició del cDNA BC020465 i de l'EST BE797822 en la regió genòmica entre *ITGA4* i *NEUROD1* (ordenament de novembre de 2002). La cerca als bancs d'EST i mRNA posava de manifest l'existència de nous gens en la regió candidata del locus RP26. Mapa físic adaptat d'UCSC Genome Bioinformatics.

Un cop encavalcades les dues seqüències *in silico*, el cDNA resultant contenia una pauta de lectura oberta de 1596 nucleòtids. La proteïna predita, resultant de la traducció conceptual d'aquest cDNA, contenia 532 aminoàcids i presentava homologia amb proteïnes quinases de ceramida, esfingosina i diacilglicerol. En concret, la semblança més alta que mostrava –29% identitat, 50% semblança– era amb la ceramida quinasa humana, CERK (Sugiura et al. 2002). El domini conservat estava codificat a cavall entre el cDNA parcial (regió aminoterminal de la proteïna) i l'EST (regió carboxiterminal), i aquest fet demostrava que eren seqüències parcials del mateix gen (FIGURA 21).

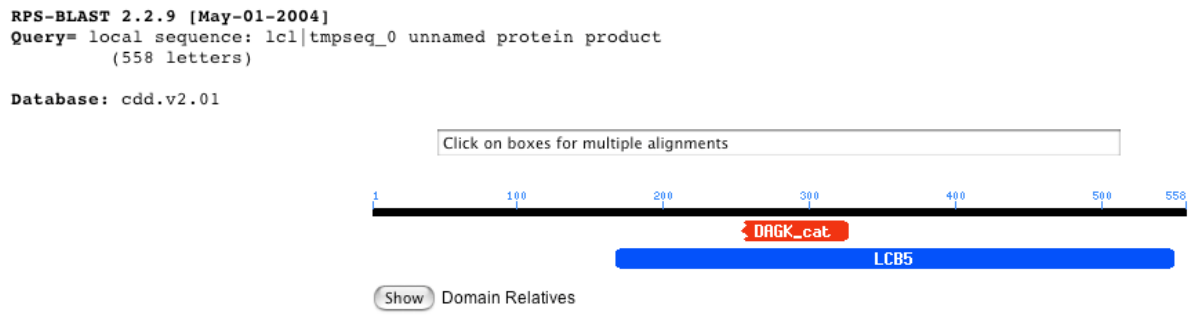



**Figura 21** | Predicció de dominis conservats en la seqüència solapada mitjançant RPS-BLAST 2.2.9, que evidenciava que el cDNA parcial BC020465 i l'EST BE797822 formaven part d'un mateix gen. En aquest cas, l'anàlisi s'ha realitzat a partir de la isoforma llarga d'empalmament alternatiu del gen (vegeu el CAPÍTOL 3 de RESULTATS), en la qual es detecta el domini LCB5 d'homologia amb ceramida, esfingosina i diacilglicerol quinases –que també es detecta en analitzar la seqüència solapada del cDNA i l'EST– i un segment del domini DAG quinasa (DAGK_cat).

### 2.6.1. Anàlisi mutacional de la ceramida quinasa a la família P2

Aquest nou gen presentava 13 exons, que vam analitzar en tots els membres de la família P2 per seqüenciació directa. En tots els pacients RP de la família vam identificar, en homozigosi, una mutació sense sentit en l'exó 5 (R257X; CGA→TGA), que truncava la proteïna al bell mig del putatiu domini catalític [FIGURA 1 I 4 EN (TUSON ET AL. 2004)]. La mutació cosegregava amb la malaltia en aquesta família i no vam poder-la detectar en una anàlisi realitzada en 340 cromosòmes control d'individus no relacionats. Acabàvem, doncs, d'identificar un nou gen responsable de retititis pigmentària autosòmica recessiva, el gen *RP26*, que codificava una ceramida quinasa, i que des de llavors anomenaríem *CERKL* (*ceramide kinase–like*).

### 2.6.2. Anàlisi de la ceramida quinasa en el panell de famílies arRP

Per tal d'avaluar la importància del gen *CERKL* en la patogènia de la retinitis pigmentària a la població espanyola, vam realitzar una anàlisi de cosegregació amb els dos microsatèl·lits que flanquegen el gen, D2S2310 i D2S364, en sis famílies espanyoles –B6, M150, i M231, no consanguínies, i B2, P1 i E1, consanguínies– del panell de 55 famílies amb retinitis pigmentària autosòmica recessiva de què disposem. Amb anterioritat a aquest treball, en les famílies del panell s'havien realitzat anàlisis de cosegregació amb marcadors del locus RP26 i en sis d'elles no s'havia pogut excloure que la malaltia pogués ser causada pel gen *RP26* (Bayés et al. 1998). L'anàlisi realitzada en aquest treball de tesi, amb els microsatèl·lits més propers al gen (D2S2310 i D2S364), vam poder descartar cinc de les sis famílies, excepte la família E1, que s'ajustava als criteris de cosegregació i homozigositat que la feien idònia per a la cerca de mutacions en el gen de la ceramida quinasa (FIGURA 22).

La seqüenciació del gen *CERKL* en els membres de la família E1, ens va permetre identificar la mateixa mutació (R257X) que havíem caracteritzat en la família P2. Aparentment, les dues famílies no estaven relacionades i, si ens fonamentàvem tant en la seva procedència geogràfica com en el

llinatge d'ambdues, ens havíem de trobar davant de fenòmens mutacionals independents. Emperò, vam voler analitzar l'haplotip de marcadors a banda i banda de la mutació per tal de determinar si realment les dues mutacions eren el reflex d'un mateix esdeveniment mutacional. Així, vam analitzar els 12 marcadors més propers –10 SNP i 2 microsatèl·lits– que envolten el gen en una regió d'aproximadament 600 kb. Les dues famílies no compartien l'haplotip a banda i banda de la mutació, i per tant vam concloure que es tractava de mutacions independents.

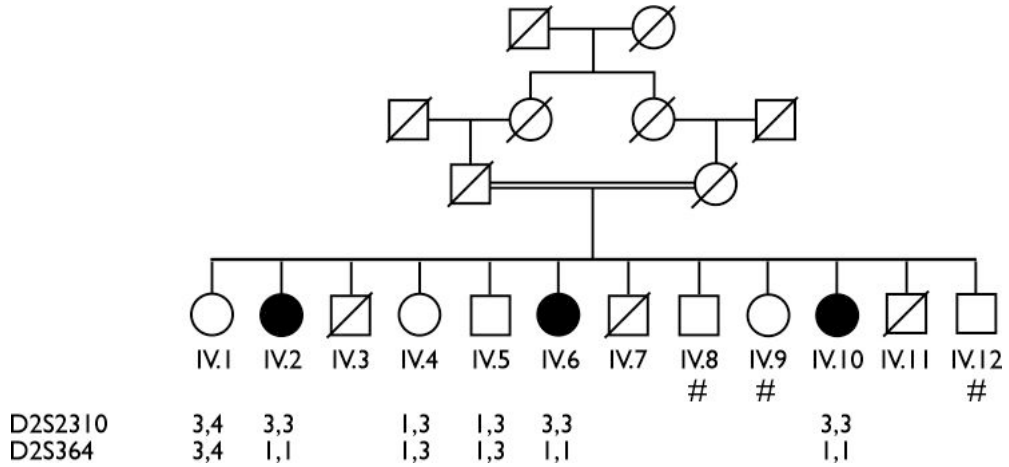



**Figura 22** | Anàlisi de cosegregació realitzada en la família E1. Sota de pedigrí s'indiquen els genotips obtinguts per als microsatèl·lits D2S2310 i D2SS364. L'asterisc denota individus que no participen en l'estudi.

*Am. J. Hum. Genet. 74:128–138, 2004*

# Mutation of *CERKL,* a Novel Human Ceramide Kinase Gene, Causes Autosomal Recessive Retinitis Pigmentosa (RP26)

Miquel Tuson, Gemma Marfany, and Roser Gonzàlez-Duarte

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona

**Retinitis pigmentosa (RP), the main cause of adult blindness, is a genetically heterogeneous disorder characterized by progressive loss of photoreceptors through apoptosis. Up to now, 39 genes and loci have been implicated in nonsyndromic RP, yet the genetic bases of >50% of the cases, particularly of the recessive forms, remain unknown. Previous linkage analysis in a Spanish consanguineous family allowed us to define a novel autosomal recessive RP (arRP) locus, RP26, within an 11-cM interval (17.4 Mb) on 2q31.2-q32.3. In the present study, we further refine the RP26 locus down to 2.5 Mb, by microsatellite and single-nucleotide polymorphism (SNP) homozygosity mapping. After unsuccessful mutational analysis of the nine genes initially reported in this region, a detailed gene search based on expressed-sequence-tag data was undertaken. We finally identified a novel gene encoding a ceramide kinase (*CERKL*), which encompassed 13 exons. All of the patients from the RP26 family bear a homozygous mutation in exon 5, which generates a premature termination codon. The same mutation was also characterized in another, unrelated, Spanish pedigree with arRP. Human *CERKL* is expressed in the retina, among other adult and fetal tissues. A more detailed analysis by in situ hybridization on adult murine retina sections shows expression of *Cerkl* in the ganglion cell layer. Ceramide kinases convert the sphingolipid metabolite ceramide into ceramide-1-phosphate, both key mediators of cellular apoptosis and survival. Ceramide metabolism plays an essential role in the viability of neuronal cells, the membranes of which are particularly rich in sphingolipids. Therefore, CERKL deficiency could shift the relative levels of the signaling sphingolipid metabolites and increase sensitivity of photoreceptor and other retinal cells to apoptotic stimuli. This is the first genetic report suggesting a direct link between retinal neurodegeneration in RP and sphingolipid-mediated apoptosis.**

## Introduction

Inherited retinal degenerative diseases are characterized by the progressive loss of mature photoreceptor cells through apoptosis. Among them, retinitis pigmentosa (RP [MIM 268000]), the most common hereditary cause of blindness, comprises a clinically and genetically heterogeneous group of retinal disorders that affects ~1.5 million people worldwide (Sullivan and Daiger 1996). Patients with RP typically show night blindness, gradual visual impairment, and bone spicule–like pigment deposits in the retina.

RP is a paradigm of monogenic diseases with extremely high genetic heterogeneity (Hims et al. 2003). This complexity has clearly hindered the identification of causative genes, particularly in the recessive forms, since large nuclear families suitable for linkage studies are scarce. In fact, although 39 genes and loci have been described, many others, still unknown, could account for the remaining unassigned cases. At present, the reported RP causative genes fall into four categories: (*a*) genes directly involved in the phototransduction cascade, (*b*) genes encoding proteins responsible for the structure and polarity of the photoreceptors, (*c*) genes encoding proteins of the visual cycle, and (*d*) regulatory genes (such as transcription and splicing factors). In the study of autosomal recessive RP (arRP), mutations on 16 genes and linkage to five chromosomal regions have been reported (Wang et al. 2001; RetNet Web site). Therefore, RP gene identification is still a priority in this field, since it not only is needed for diagnosis but also is crucial for giving insights into the molecular basis of the disease.

RP26, one of the remaining uncharacterized loci, was previously mapped to an 11-cM interval on 2q31.2-q32 in a consanguineous family with arRP (Bayés et al. 1998). Later, this 11-cM interval, between markers D2S148 and D2S117, was ascribed to a physical distance of 17.4 Mb that comprised more than 50 genes. As the incoming human genome data were made available in the public databases, new markers—either microsatellites or, mainly, SNPs—were retrieved and analyzed, first for cosegregation and finally for homozygosity mapping. All

Received September 5, 2003; accepted for publication October 27, 2003; electronically published December 16, 2003.

Address for correspondence and reprints: Dr. Roser Gonzàlez-Duarte, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E-08028 Barcelona, Spain. E-mail: rgonzalez@ub.edu


0002-9297/2004/7401-0013$15.00

of this new information allowed us to narrow the candidate region to 2.5 Mb, and mutational screening was subsequently performed to identify the RP26 gene.

## Material and Methods

### *Families and DNA*

The two families with arRP we analyzed, P2 and E1, have been described elsewhere (Bayés et al. 1996, 1998). Blood samples were collected from family members, and DNA was extracted using a commercial kit (Wizard Genomic DNA Purification Kit [Promega]). Informed consent was gathered from all family members. This research followed the tenets of the Declaration of Helsinki.

### *Cosegregation and Homozygosity Mapping*

The RP26 locus was further refined using three flanking markers: D2S2978, D2S2261, and D2S273 (table 1). For subsequent homozygosity mapping, additional markers (13 microsatellites and 101 SNPs) were chosen on the basis of their physical position according to the National Center for Biotechnology Information (NCBI) Human Builds 29–33 at the University of California Santa Cruz Genome Bioinformatics Web site (table 1). The SNP sequences were retrieved from the NCBI dbSNP Home Page and The SNP Consortium database. New SNPs observed during this study were submitted and deposited in the NCBI dbSNP database (see table 1). SNP analyses involved PCR assays with flanking primers and subsequent sequencing to verify the genotype. Three amplification steps (30 s each), with the annealing temperature ranging from 48°C to 61°C, were performed. The reaction mixture (50 $\mu$l) contained 10 $\mu$M of each primer, 2 $\mu$M of dNTPs, 1.5 mM $MgCl_2$, and 1 U of *Taq* polymerase (Innogenetics).

### *Mutation Screening*

Primers located at the flanking intron regions were designed to amplify each coding exon for all the analyzed genes in all samples. Three amplification steps (30 s each), with the annealing temperature ranging from 45°C to 60°C, were performed. Sequencing was performed using the BigDye Terminator Cycle Sequencing v3.1 kit and the automatic sequencers ABI 3700 and ABI 3730 (Applied Biosystems).

### *Mutation Restriction Analysis*

A forward primer (5′-GAAGAATGCTGGGATGGAAACGGAC-3′) containing a mismatch nucleotide (underlined) that created an *Ava*II restriction site in the exon 5 wild-type sequence was designed. A three-step PCR (annealing temperature 50°C) was performed on the genomic DNA of the P2 and E1 family members, through use of the mismatch forward primer and exon 5 reverse primer (5′-GTTGTGCTGTCTAGATTAGC-3′), under the conditions mentioned above. The 157-bp PCR fragments were digested with *Ava*II and were resolved in a 4% low-melting-temperature agarose gel.

### *Human and Murine* CERKL *cDNA Characterization*

Database searches from the available human genomic sequence allowed the identification of a truncated EST (GenBank accession number BE797822) and a partial cDNA (GenBank accession number BC020465), which overlapped 18 bp. To characterize the full-length cDNA, a human retina cDNA library (BD Biosciences) was screened ($1.2 \times 10^6$ plaque-forming units) with a mixture of all the *CERKL* coding exons as a probe. The hybridization was performed overnight under highly stringent conditions (42°C in 50% formamide, 5 × Denhardt's, 5 × SSPE, 0.1% sodium dodecyl sulfate, and 100 $\mu$g/ml salmon sperm DNA), and several washes were performed at 55°C in 2 × SSC and 1 × SSC before autoradiography. Only a partial *CERKL* cDNA was obtained. For the full-length coding sequence, we devised a PCR strategy using two specific primers flanking the conceptual ATG and STOP codons on the Marathon-Ready retina cDNA (BD Biosciences) under the following conditions: 94°C for 30 s, 60°C for 30 s, and 72°C for 2 min for 40 cycles. The reaction mixture (50 $\mu$l) contained 10 $\mu$M of each primer, 2 $\mu$M of dNTPs, 1.5 mM $MgCl_2$, 1 U of *Taq* polymerase (Innogenetics), and 5% DMSO.

TBLASTN database searches at the NCBI BLAST server revealed two murine sequences homologous to human *CERKL,* one EST (GenBank accession number BY742285) and one partial cDNA (GenBank accession number BC046474), which had 377 nt of overlap with one another. Protein alignment was performed with CLUSTALW 1.8 at the BCM Search Launcher Web site and was depicted with BOXSHADE 3.21 at the EMBnet BoxShade Server.

### *Phylogenetic Analysis*

Protein sequences from several members of the diacylglycerol, sphingosine, and ceramide kinase families were retrieved from public databases and aligned with human and murine CERKL through use of CLUSTALX version 1.64 (Thompson et al. 1997). The resulting phylogenetic tree was constructed using the neighbor-joining algorithm and considering the diacylglycerol kinase sequences as outgroup.

### *RT-PCRs*

Two specific primers on exons 8 (5′-CCATTTAACAGCTCTGATGATGTGCAAG-3′) and 12 (5′-GACCTCTGATGCAACTTCCATTAAGTCAC-3′), which ampli-

## Table 1

**Microsatellite and SNP Markers Analyzed on 2q31.2-q32.3**

| Microsatellite or SNP ID[a] | Location[b] | Locus | Genotype in RP26 Patients[c] |
|---|---|---|---|
| D2S148; AFM200WA11 | chr2: 176902551–176902745 | | Recombinant |
| D2S2978; ATA44H04 | chr2: 179788236–179788348 | | Recombinant |
| D2S2261; AFMB072WG1 | chr2: 180180111–180180442 | | Recombinant |
| rs1002207; TSC0036660 | chr2: 180237706 | | Noninformative |
| rs1949453; TSC1026876 | chr2: 180594170 | *UBE2E3* | ***Homozygous*** |
| *rs2077783; TSC0046543 | chr2: 180834855 | | ***Homozygous*** |
| *rs720453; TSC0046544 | chr2: 180834909 | | ***Homozygous*** |
| *D2S2310; AFMB355XD5 | chr2: 180849531–180849858 | | Noninformative |
| *rs155100; TSC0531809 | chr2: 181033165 | *ITGA4* | ***Homozygous*** |
| *rs155101; TSC0531810 | chr2: 181033223 | *ITGA4* | ***Homozygous*** |
| *rs155102 | chr2: 181033525 | *ITGA4* | ***Homozygous*** |
| *rs155103 | chr2: 181033821 | *ITGA4* | ***Homozygous*** |
| rs1801262 | chr2: 181226533 | *NEUROD1* | ***Homozygous*** |
| rs2583016 | chr2: 181228296 | *NEUROD1* | ***Homozygous*** |
| *rs1157595; TSC0326375 | chr2: 181432646 | | ***Homozygous*** |
| *D2S364; AFM303YA9 | chr2: 181717611–181718002 | *PDE1A* | ***Homozygous*** |
| *rs3754929 | chr2: 181771662 | *PDE1A* | ***Homozygous*** |
| *rs1438065; TSC0655111 | chr2: 181902071 | *PDE1A* | ***Homozygous*** |
| *rs833152; TSC0655110 | chr2: 181902179 | *PDE1A* | Noninformative |
| rs288254; TSC0681111 | chr2: 182291705 | *ERdj5* | Noninformative |
| rs7510 | chr2: 182472727 | *NCKAP1* | Heterozygous |
| ss12676100 | chr2: 182476461 | *NCKAP1* | Heterozygous |
| ss12676101 | chr2: 182476476 | *NCKAP1* | Heterozygous |
| ss12676102 | chr2: 182482636 | *NCKAP1* | Heterozygous |
| rs2271671 | chr2: 182510096 | *NCKAP1* | Heterozygous |
| D2S350; AFM292WD1 | chr2: 182532022–182532389 | *NCKAP1* | Heterozygous |
| ss12676103 | chr2: 182678153 | LOC129401 | Heterozygous |
| ss12676104 | chr2: 182678197 | LOC129401 | Heterozygous |
| ss12676105 | chr2: 182678220 | LOC129401 | Heterozygous |
| D2S2273; AFMB297XC1 | chr2: 182825486–182825797 | | Heterozygous |
| D2S2281; AFMB310XF5 | chr2: 182863973–182864335 | | Heterozygous |
| D2S2366; AFMA057VG9 | chr2: 183175061–183175459 | | Noninformative |
| rs826135; TSC1254035 | chr2: 183459257 | | Noninformative |
| rs826134; TSC0476562 | chr2: 183459267 | | Noninformative |
| ss12676106 | chr2: 183459509 | | Heterozygous |
| rs826133; TSC0658245 | chr2: 183459526 | | Noninformative |
| ss12676107 | chr2: 183459659 | | Heterozygous |
| ss12676108 | chr2: 183459776 | | Heterozygous |
| rs826132; TSC0658244 | chr2: 183459802 | | Heterozygous |
| D2S1391; GATA65C03 | chr2: 183675508–183675835 | | Homozygous |
| rs1443021; TSC0664031 | chr2: 183910062 | | Heterozygous |
| rs1443022; TSC0664032 | chr2: 183910149 | | Heterozygous |
| rs768352; TSC0076040 | chr2: 184012769 | | Homozygous |
| rs994653; TSC0317578 | chr2: 184193063 | | Homozygous |
| rs728534; TSC0064305 | chr2: 184484995 | | Homozygous |
| rs3046266; TSC1530131 | chr2: 184485293 | | Noninformative |
| rs1366842; TSC0514415 | chr2: 184485321 | | Noninformative |
| rs3731834 | chr2: 184486442 | | Homozygous |
| D2S1361; GATA14E05 | chr2: 184899775–184900087 | | Heterozygous |
| rs1016410; TSC0098979 | chr2: 185334459 | | Heterozygous |
| rs1019430; TSC0223209 | chr2: 185334589 | | Heterozygous |
| rs878845; TSC0212635 | chr2: 185580725 | | Heterozygous |
| ss12676109 | chr2: 186138356 | *ITGAV* | Heterozygous |
| ss12676110 | chr2: 186170238 | *ITGAV* | Heterozygous |
| ss12676111 | chr2: 186181185 | *ITGAV* | Heterozygous |
| ss12676112 | chr2: 186186304 | *ITGAV* | Heterozygous |
| ss12676113 | chr2: 186202416 | *ITGAV* | Heterozygous |
| rs3816386 | chr2: 186211713 | *ITGAV* | Heterozygous |
| rs2018302; TSC0091486 | chr2: 186467891 | | Heterozygous |
| rs1467990; TSC0384843 | chr2: 186468012 | | Noninformative |
| rs2018314; TSC0085681 | chr2: 186468088 | | Heterozygous |
| rs1000623; TSC0033486 | chr2: 186468142 | | Heterozygous |
| rs840570; TSC1028023 | chr2: 186855755 | | Noninformative |
| rs2308091 | chr2: 186856059 | | Noninformative |
| rs84069 | chr2: 186856203 | | Noninformative |
| rs84068 | chr2: 186856285 | | Noninformative |

*(continued)*

**Table 1 (continued)**

| Microsatellite or SNP ID[a] | Location[b] | Locus | Genotype in RP26 Patients[c] |
|---|---|---|---|
| D2S152; AFM207XG1 | chr2: 186914674–186915071 | *CALCRL* | Noninformative |
| rs1398061; TSC0576737 | chr2: 186941063 | *CALCRL* | Noninformative |
| rs1464338; TSC0378776 | chr2: 187177069 | | Noninformative |
| rs1464337; TSC0378775 | chr2: 187177070 | | Noninformative |
| rs2176858; TSC1217605 | chr2: 187177146 | | Noninformative |
| rs2033838; TSC1047282 | chr2: 187551834 | | Heterozygous |
| rs2028374; TSC1091521 | chr2: 187666577 | | Heterozygous |
| rs1354906; TSC0490640 | chr2: 188129892 | CED-6 | Homozygous |
| rs1007120: TSC0084553 | chr2: 188175289 | | Noninformative |
| rs925825; TSC0241755 | chr2: 188431720 | | Heterozygous |
| rs893407; TSC0165376 | chr2: 189033193 | | Heterozygous |
| rs2289404 | chr2: 189318673 | *ORMDL1* | Noninformative |
| rs6942; TSC0023616 | chr2: 189319070 | *ORMDL1* | Homozygous |
| rs3791767 | chr2: 189322574 | *ORMDL1* | Noninformative |
| rs288817; TSC1078112 | chr2: 189529648 | | Noninformative |
| ss12676114 | chr2: 189529690 | | Homozygous |
| rs998173; TSC0011271 | chr2: 189694293 | | Homozygous |
| rs3791789 | chr2: 189787488 | *HIBCH* | Heterozygous |
| D2S2262; AFMB082YE1 | chr2: 189985176–189985493 | | Heterozygous |
| rs7721 | chr2: 190069246 | | Heterozygous |
| rs1128723 | chr2: 190069374 | | Heterozygous |
| rs8962 | chr2: 190074316 | | Noninformative |
| rs1468685; TSC0385949 | chr2: 190268436 | | Heterozygous |
| rs1263136; TSC1488116 | chr2: 190298357 | | Noninformative |
| rs1263125; TSC0446273 | chr2: 190303089 | | Homozygous |
| D2S118; AFM066XC1 | chr2: 190309134–190309312 | | Homozygous |
| rs1263100; TSC0398872 | chr2: 190327138 | | Heterozygous |
| rs1476896; TSC0398871 | chr2: 190327184 | | Homozygous |
| D2S389; AFM333WF9 | chr2: 190347305–190347628 | | Homozygous |
| rs1882395; TSC0896111 | chr2: 190387668 | | Homozygous |
| D2S1775; GATA71C07 | chr2: 190402485–190402614 | | Homozygous |
| rs3217036 | chr2: 190462343 | *GLS* | Noninformative |
| rs2355571 | chr2: 190477428 | *GLS* | Heterozygous |
| rs2883713 | chr2: 190491082 | *GLS* | Heterozygous |
| rs3199237 | chr2: 190497863 | *GLS* | Noninformative |
| rs3207595 | chr2: 190497884 | *GLS* | Noninformative |
| rs3207596 | chr2: 190497887 | *GLS* | Noninformative |
| rs3199238 | chr2: 190497893 | *GLS* | Noninformative |
| rs3207597 | chr2: 190497917 | *GLS* | Noninformative |
| rs1801893 | chr2: 190498994 | *GLS* | Noninformative |
| rs3215259 | chr2: 190521768 | *GLS* | Noninformative |
| rs1058589 | chr2: 190522002 | *GLS* | Noninformative |
| rs1058590 | chr2: 190522011 | *GLS* | Noninformative |
| rs1058591 | chr2: 190522017 | *GLS* | Noninformative |
| rs1058592 | chr2: 190522020 | *GLS* | Noninformative |
| rs1547550; TSC0429185 | chr2: 190548390 | *STAT1* | Heterozygous |
| rs925847; TSC0241788 | chr2: 190600205 | *STAT4* | Noninformative |
| rs3024891 | chr2: 190601614 | *STAT4* | Heterozygous |
| D2S2246; AFMB007WC1 | chr2: 191085514–191085746 | | Heterozygous |
| D2S318; AFM105XC1 | chr2: 191278428–191278834 | | Homozygous |
| D2S161; AFM224ZF4 | chr2: 191431730–191432052 | | Heterozygous |
| D2S280; AFM155YE1 | chr2: 191557728–191557932 | *TMEFF2* | Homozygous |
| rs3738882 | chr2: 191751560 | *TMEFF2* | Homozygous |
| rs935367; TSC0352562 | chr2: 191842107 | | Noninformative |
| D2S315; AFM081YG5 | chr2: 192076101–192076267 | | Homozygous |
| rs717621; TSC0040380 | chr2: 192153621 | | Heterozygous |
| D2S273; UT5048 | chr2: 192288946–192289538 | | Recombinant |
| D2S117; AFM065YF11 | chr2: 194331362–194331699 | | Recombinant |

[a] Microsatellite IDs are those that begin with "D," and SNP IDs are those that begin with "rs" or "ss." SNPs IDs that begin with "ss" are those identified in the present study and are deposited in the NCBI dbSNP database. Microsatellite and SNP IDs preceded by an asterisk (*) are the flanking *CERKL* markers used for haplotype analysis and comparison in families P2 and E1.

[b] Location in NCBI Build 32.

[c] Boldface italic type indicates candidate homozygous region.

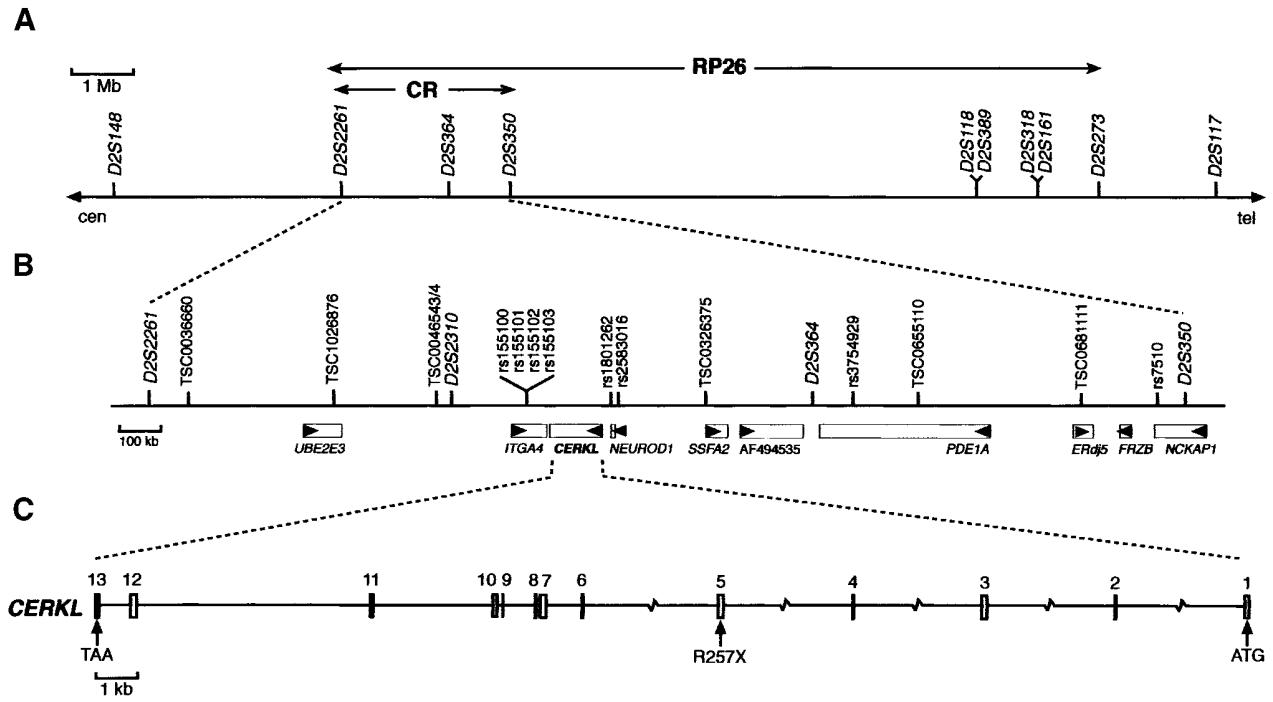



**Figure 1** The RP26 locus and the *CERKL* gene and protein. *A*, RP26 physical map of the 12.5-Mb cosegregation interval between markers D2S2261 and D2S273. *B*, Localization of *CERKL* within the candidate region ("CR") defined by homozygosity mapping. Genes (*rectangles*) and the transcription direction (*arrowheads*) are indicated. *C*, *CERKL* exon/intron structure. The location of the R275X null mutation in exon 5 is shown.

fied 429 bp, were used to detect *CERKL* expression on a panel of first-strand cDNAs of several human adult and fetal tissues (BD Biosciences). Specific *CERKL* expression was confirmed by sequencing the amplified products. Amplification of *GAPD* was used to compare and normalize the samples, as suggested by the manufacturer. PCR conditions were as follows: 35 cycles of 94°C for 30 s and 65°C for 60 s. The reaction mixture (50 $\mu$l) contained 10 $\mu$M of each primer, 2 $\mu$M of dNTPs, 1.5 mM $MgCl_2$, and 1 U *Taq* pol (Innogenetics).

### *In Situ Hybridizations*

For in situ hybridization, adult mouse eye slides from the C57BL/6J strain were obtained from Novagen. Eyes were fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned at 7 $\mu$m. Sections were washed for 7 min three times with xylene to remove paraffin, for 3 min twice with ethanol 100%, for 3 min with ethanol 90%, for 3 min with ethanol 70%, and for 3 min with diethyl pyrocarbonate–treated water, and then were fixed with 4% paraformaldehyde for 20 min. Sections were treated with 2 $\mu$g/ml proteinase K for 30 min, washed for 5 min twice with phosphate-buffered saline, and postfixed with 4% paraformaldehyde. Acetylation with 0.1 M triethanolamine-HCl (pH 8.0) containing 0.25% acetic anhydride was performed for 10 min. Slides were subsequently washed, ethanol dehydrated, and air dried. Hybridization was performed overnight at 55°C with 2 $\mu$g/ml digoxigenin-labeled riboprobes in 50% formamide, 1 × Denhardt's solution, 10% dextran sulfate, 0.3 M NaCl, 10 mM Tris-HCl (pH 8.0), 5 mM EDTA (pH 8.0), 10 mM $NaH_2PO_4$, and 200 $\mu$g/ml salmon sperm DNA.

The riboprobes were generated from the T7 and T3 promoters of a pBluescript II KS vector containing either a 201-bp *Hin*dIII-*Pst*I fragment, a 350-bp *Hin*dIII fragment, a 513-bp *Pst*I-*Sph*I fragment of the murine *Cerkl* cDNA IMAGE 4504238, or a 198-bp fragment from coding exon 4 of murine rhodopsin.

After hybridization, the slides were washed in 2 × SSC for 20 min; washed for 5 min twice in a solution of 50% formamide and 2 × SSC (washes at 50°C); equilibrated in NTE (0.5 M NaCl, 10 mM Tris-HCl pH 8.0, 5 mM EDTA) at 37°C; and then treated with 10 $\mu$g/ml RNase A in NTE at 37°C for 30 min. Subsequently, the sections were washed in NTE, 2 × SSC, and 0.1 × SSC (15 min each), equilibrated in buffer 1 (100 mM Tris-HCl pH 7.5, 150 mM NaCl), and blocked in blocking buffer (1% BSA and 0.1% Triton X-100 in buffer 1) for 30 min. An anti-digoxigenin-AP conjugate antibody (1:500; Roche Applied Science) in blocking buffer was incubated

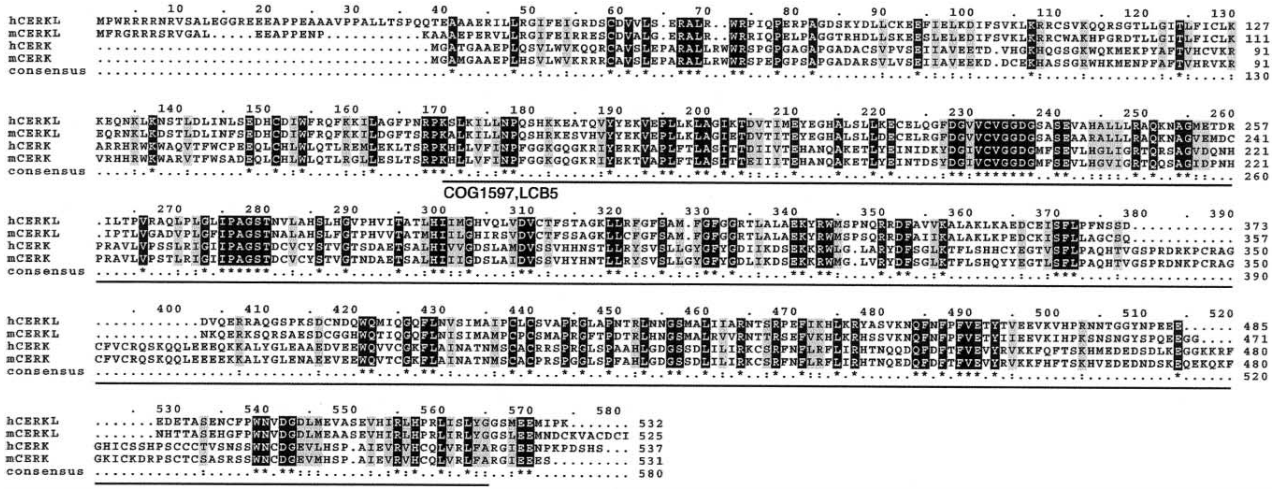



**Figure 2** Protein alignment of human ("h") and murine ("m") CERKL and CERK. Identical residues are highlighted in black, and conservative positions are in gray. The solid black line underlines the conserved COG1597 (LCB5) domain detected after comparison with the NCBI Conserved Domain Database. Murine CERKL sequence was assembled from GenBank sequences BY742285 and BC046474, human CERK sequence from AB079066, and murine CERK sequence from AB79067.

overnight at 4°C. The sections were washed twice in buffer 1 for 15 min and in buffer 2 (100 mM Tris-HCl pH 9.5, 150 mM NaCl, and 50 mM $MgCl_2$) for 10 min prior to exposure to the alkaline phosphatase substrate, nitroblue tetrazolium–5-bromo-4-chloro-3-indoyl phosphate (NBT-BCIP; Roche Applied Science). The reaction was stopped by several washes in distilled water. The sections were cover-slipped with 70% glycerol in PBS and were photographed using a Nomarski optics microscope (Axioplan, Carl Zeiss) equipped with a Nikon Coolpix digital camera.

## Results

### *Refinement of the Cosegregation Region and Candidate Gene Analysis*

The analysis of two proximal (D2S2978 and D2S2261) and one distal (D2S273) flanking markers (table 1) allowed further refinement of the RP26 locus to a 12.5-Mb interval (fig. 1*A*). As a first attempt to identify the causative gene, a mutational screening of candidate genes selected on the basis of their known retinal function or the phenotype of *Drosophila* mutants or knockout mice was performed. The analysis of five candidates (*ORMDL1, ITGAV, GLS, NEUROD1,* and *FRZB*) within this interval did not reveal any pathogenic mutation.

### *Homozygosity Mapping and Mutation Screening*

The reported consanguinity of the RP26 family and the availability of new markers after the sequencing of the human genome made homozygosity mapping feasible as a means to reduce the candidate region. Sixteen microsatellite markers and 101 SNPs located within the 12.5-Mb interval (table 1) were analyzed in all P2 family members. Eventually, only one homozygous segment spanning >1 Mb, between markers D2S2261 and D2S350 (fig. 1*A*), was identified. This candidate region, spanning 2.5 Mb, harbored five annotated genes and four partially characterized mRNAs (fig. 1*B*), which were prioritized for mutational analysis on the basis of (*a*) retinal expression, as assessed by PCR on a retina cDNA library; and (*b*) availability of functional data. All of them were eventually sequenced (in fact, two genes, *NEUROD1* and *FRZB,* were analyzed in the previous section), but none showed any pathogenic variant.

### *Identification of a Novel Gene,* CERKL

A more exhaustive database search for uncharacterized cDNAs and spliced ESTs within the candidate 2.5-Mb region was undertaken and revealed two close sequences, one EST (GenBank accession number BE797822) and one eye/retinoblastoma incomplete cDNA (GenBank accession number BC020465), located between the *ITGA4* and the *NEUROD1* genes. The conceptual protein from the *in silico* assembly of both sequences showed homology with eukaryotic ceramide, sphingosine, and diacylglycerol kinases and strongly suggested that the sequences analyzed corresponded to a novel, unreported gene. After screening a human retina cDNA library and performing serial PCR assays on retina cDNA, we obtained the full-length 1,596-nt coding sequence, which spanned 13 exons (fig. 1*C*). The predicted 532-aa protein showed the highest similarity (29% identity; 50% similarity) with the human

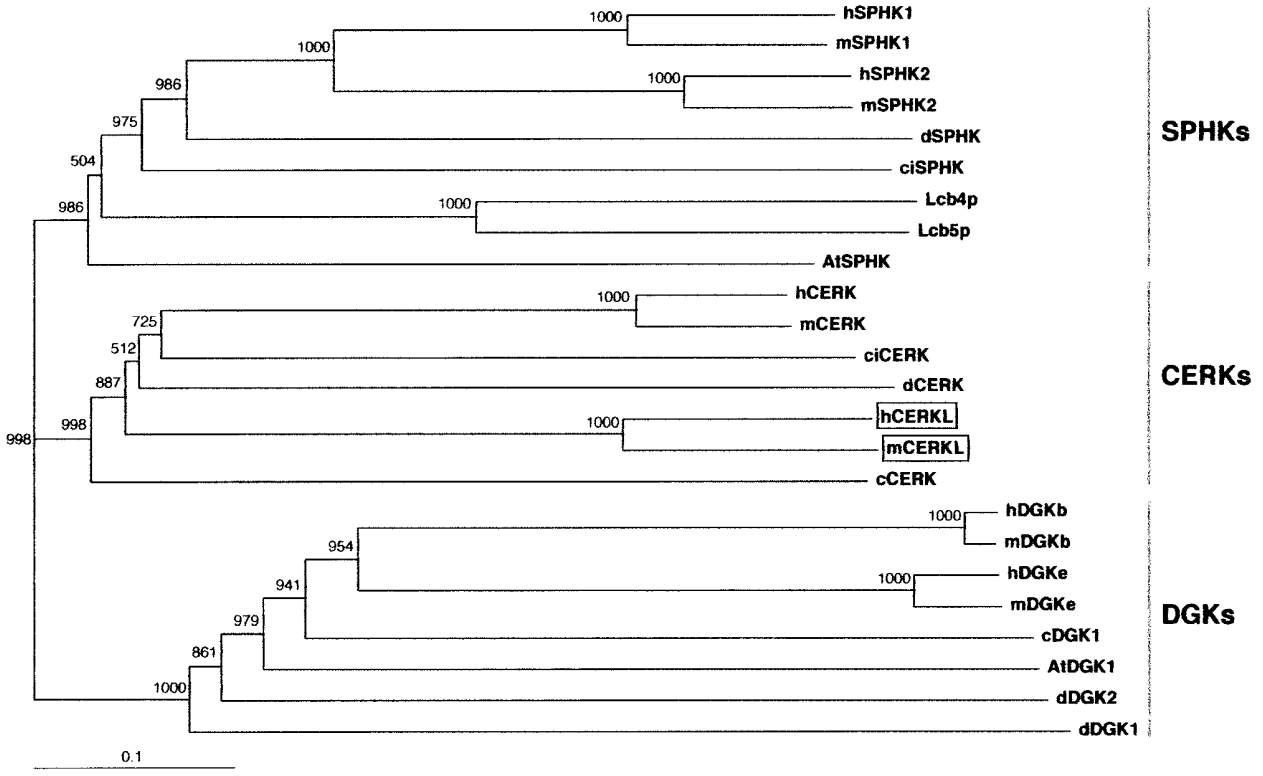



**Figure 3** Unrooted phylogenetic tree of several members of the diacylglycerol, sphingosine, and ceramide kinases, including the human and murine CERKL sequences described in the present study (*boxed*). The tree was obtained with the neighbor-joining algorithm of the program ClustalX version 1.64b. Numbers show the values for the bootstrap analysis. SPHKs = sphingosine kinases; CERKs = ceramide kinases; and DGKs = diacylglycerol kinases. Abbreviations and GenBank accession numbers are as follows: human sphingosine kinase 1 (hSPHK1), AAF73423; murine sphingosine kinase 1 (mSPHK1), AAC61697; human sphingosine kinase 2 (hSPHK2), AAF74124; murine sphingosine kinase 2 (mSPHK2), AAF74125; *Drosophila* sphingosine kinase (dSPHK), AAF48045; *Ciona intestinalis* sphingosine kinase (ciSPHK), AK112588; yeast sphingosine kinase Lcb4 (Lcb4p), NP_014814; yeast sphingosine kinase Lcb5 (Lcb5p), NP_013361; *Arabidopsis thaliana* sphingosine kinase (AtSPHK), AY128394; human ceramide kinase (hCERK), AB079066; murine ceramide kinase (mCERK), AB079067; *Ciona intestinalis* ceramide kinase (ciCERK), AK112750; *Drosophila* ceramide kinase (dCERK), AAF52040; human ceramide kinase-like (hCERKL), AY357073; murine ceramide kinase-like (mCERKL), BY742285 and BC046474; *Caenorhabditis elegans* ceramide kinase (cCERK), AAC67466; human diacylglycerol kinase $\beta$ subunit (hDGKb), Q9Y6T7; murine diacylglycerol kinase $\beta$ subunit (mDGKb), XP_147651; human diacylglycerol kinase epsilon subunit (hDGKe), NP_003638; murine diacylglycerol kinase epsilon subunit (mDGKe), NP_062378); *Caenorhabditis elegans* diacylglycerol kinase 1 (cDGK1), NP_508190; *Arabidopsis thaliana* diacylglycerol kinase 1 (AtDGK1), Q39017; *Drosophila* diacylglycerol kinase 2 (dDGK2), Q09103; and *Drosophila* diacylglycerol kinase 1 (dDGK1), Q01583.

ceramide kinase (CERK [Sugiura et al. 2002]) (fig. 2) and, therefore, the gene was named "human ceramide kinase–like" (*CERKL* [HUGO-approved nomenclature]). In addition, database searches revealed one murine EST (GenBank accession number BY742285) and one partial cDNA (GenBank accession number BC046474), which, upon assembly and conceptual translation, produced a protein that was highly homologous to human CERKL (75% identity; 85.6% similarity) (fig. 2). To verify that CERKL belonged to the ceramide kinase subfamily of lipid kinases, a phylogenetic tree was constructed using 24 protein sequences from different eukaryotic phyla. Human and murine CERKL sequences clearly clustered within the branch of ceramide kinases (fig. 3).

### *Premature Truncation of* CERKL *Causes arRP*

The *CERKL* coding exons were sequenced in all members of the RP26 family. All patients were homozygous for a nonsense mutation (R257X; CGA→TGA) in exon 5, which prematurely truncates the protein within the predicted catalytic domain (fig. 4*A* and 4*B*). This mutation was not observed in 170 unrelated, ethnically matched control individuals. We subsequently evaluated cosegregation of RP with two *CERKL* flanking markers (D2S2310 and D2S364) in a panel of small nuclear Spanish families with RP whose causative gene was still unknown. An additional cosegregating pedigree, unrelated to the RP26 family, was identified. The *CERKL*

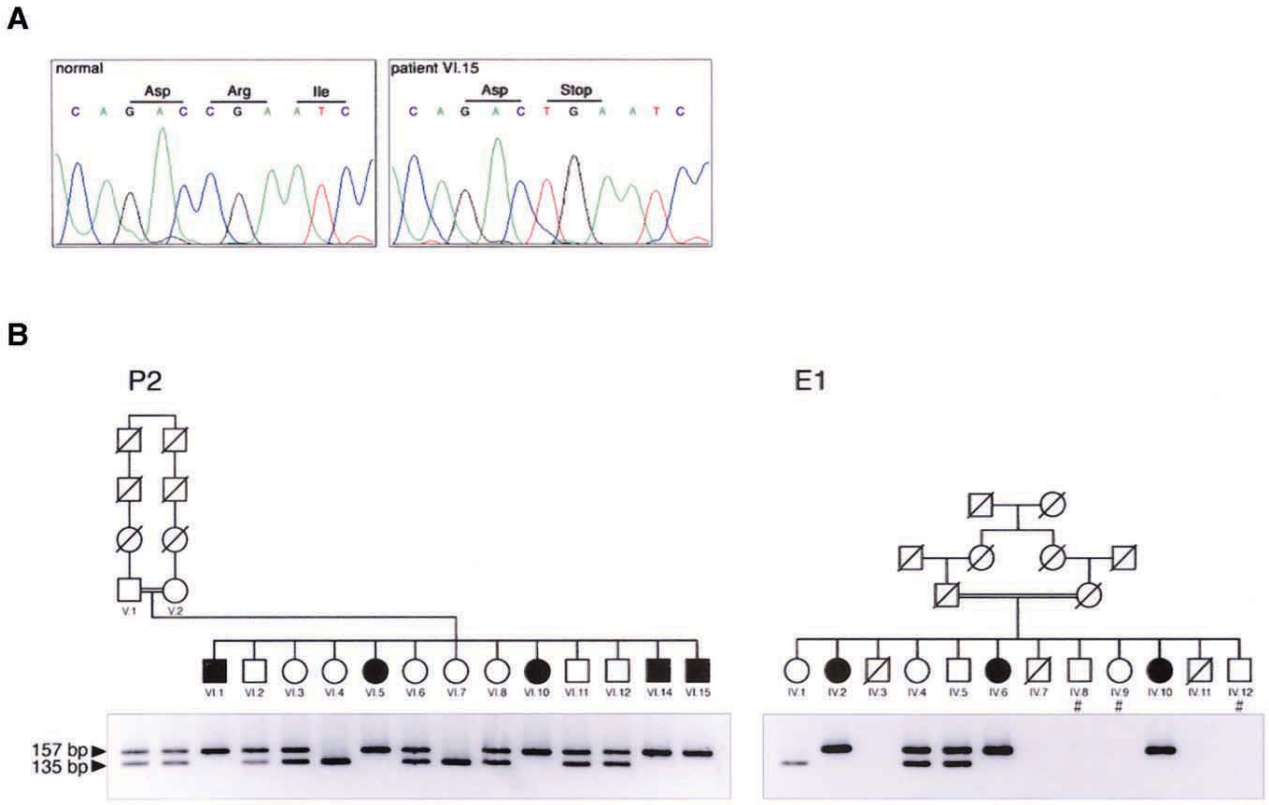



**Figure 4** *A,* DNA sequence of exon 5 in an unaffected individual and patient VI.15 of the P2 family, showing the homozygous nonsense R257X mutation. *B,* Pedigrees of P2 and E1 families, showing the ARMS analysis of R257X mutation. A specific primer was designed to generate an *Ava*II restriction site in the amplified wild-type allele. All RP-affected individuals are homozygous for the mutant, nonrestricted allele. Individuals whose DNA samples were not available are indicated by a number sign (#).

sequence analysis showed that the affected members were also homozygous for the R257X mutation (fig. 4*B*), thus reinforcing the pathogenicity of this CERKL variant. Although both families share the same mutation, the haplotype analysis of the 12 nearest flanking markers (see table 1), stretching 600 kb around *CERKL,* does not support a common ancestry (data not shown).

### CERKL *Shows a Tissue-Specific Pattern of Expression*

Northern expression analysis of *CERKL* did not reveal any transcript on two panels of several human tissues, even after long exposure times (data not shown). This suggested either specific expression in nontested tissues or very low transcriptional levels, in contrast to the higher and more ubiquitous expression of the *CERK* gene (Sugiura et al. 2002). The more sensitive RT-PCR assays revealed moderate expression in adult human retina, kidney, lung, and pancreas, and low levels in brain, placenta, and liver (fig. 5). The RT-PCRs on human fetal tissues also showed expression in lung, kidney, and brain (fig. 5).

To assess *CERKL* expression in the retina more accurately, we performed in situ hybridization on mouse eye sections. The mouse *CERKL* ortholog is predominantly expressed in the retina ganglion cell layer, although a faint signal is also detected in the inner nuclear and photoreceptor cell layers (fig. 6*A* and 6*B*). The axons from the ganglion cells merge in the optic nerve and, therefore, constitute the last cellular component of phototransduction in the retina.

As a positive control, we assayed rhodopsin, which is highly expressed in the photoreceptor cells (fig. 6*C*). It is remarkable that, at a comparable specific labeling and concentration probe, rhodopsin hybridization signal was much more intense and was detected much earlier than *Cerkl* expression, again supporting the low transcriptional levels of the latter.

## Discussion

The identification of *CERKL* and its characterization as an RP causative gene opens new avenues for evaluating

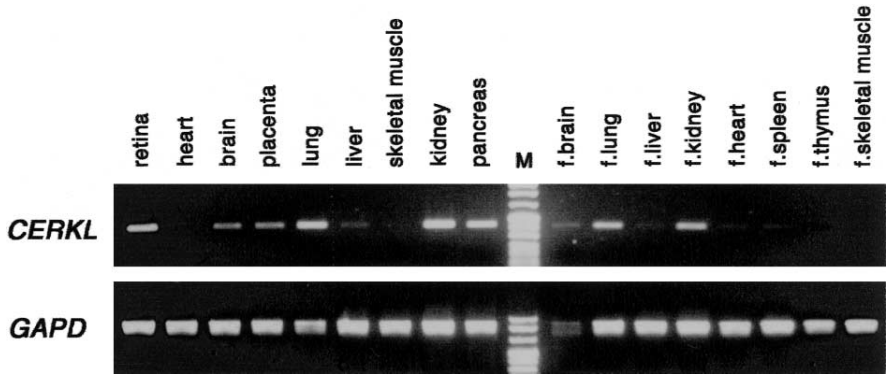



**Figure 5** *CERKL* and *GAPD* (control) RT-PCR expression analysis on several human tissues, including retina

its contribution—as well as that of other sphingolipid (SL) metabolism enzymes—to retinal function and disease. The pathogenicity of the R257X mutation is supported not only by the absence of this variant in the control population (the test on 340 chromosomes allows the detection of a polymorphism of 1% with 95% statistical confidence [Collins and Schwartz 2002]) but also, and with more relevance to function, by the premature truncation of the protein (position 257 out of 532 aa), well within the strictly conserved catalytic domain of ceramide, sphingosine, and diacylglycerol kinases (fig. 2).

During the last decade and after exhaustive cosegregation analyses using internal and flanking markers of all the reported arRP genes and loci (including RP26), we have identified the causative gene in only 8 (~15%) families out of 52 Spanish pedigrees with arRP. When we consider all our data, *CERKL* (responsible for the disease in two families [present study]) appears to contribute to arRP to a similar extent as other well-known arRP genes, such as *PDE6B* (three families [Bayés et al. 1996]), *TULP1* (one family [Paloma et al. 2000]), *ABCA4* (one family [Martínez-Mir et al. 1998]), and *CNGA1* (one family [Paloma et al. 2002]). Although it is too early to assess the real contribution of *CERKL* to retinal disorders, we surmise that, after this first study, additional families and new mutations will be readily described.

When we consider CERKL function, we should remember that SLs are structural membrane components, particularly abundant in neurons. SL accumulation is associated with severe neurodegenerative disorders (Gaucher, Tay-Sachs, and Niemann-Pick A and B diseases [Buccoliero and Futerman 2003]). Moreover, increasing evidence points to both nonphosphorylated and phosphorylated sphingosine and ceramide as essential second messengers in cellular stress, senescence, growth, and apoptosis (Luberto et al. 2002). SLs and their metabolic products act as biosensors of the cellular state, and the enzymes that regulate SL metabolism appear to be fine-tuned cellular switches, connecting several pathways with antagonistic properties (Hannun and Obeid 2002). Thus, ceramide and sphingosine lead to cell growth arrest and apoptosis, whereas sphingosine-1-phosphate and ceramide-1-phosphate have antiapoptotic and neuroprotective effects (Frago et al. 1998, 2003; Hannun and Obeid 2002; Spiegel and Milstien 2002, 2003). In this context, sensory neuropathies (hereditary sensory neuropathy type 1 [Dawkins et al. 2001]) and other neurodegenerative diseases (Batten disease [Puranam et al. 1999]) have recently been associated with ceramide metabolism enzyme deficiencies. It is puzzling that some key enzymes regulating intracellular ceramide levels have remained elusive (Hannun and Obeid 2002)—for example, the first ceramide kinase to be characterized was reported only recently (Sugiura et al. 2002). However, despite the short lapse of time, data on the function of ceramide-1-phosphate and ceramide kinase in several scenarios have increased, stressing the roles of the relative levels of

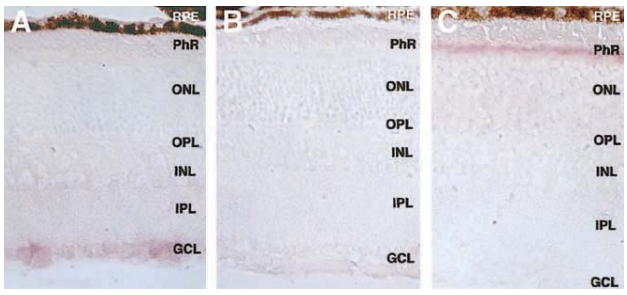



**Figure 6** *A, Cerkl* in situ hybridization on murine eye sections with the antisense riboprobe. *B, Cerkl* in situ hybridization on murine eye sections with the sense riboprobe (negative control). *C, Rhodopsin* in situ hybridization on murine eye sections (positive control). RPE = retinal pigment epithelium; PhR = photoreceptor cell layer; ONL = outer nuclear layer; OPL = outer plexiform layer; INL = inner nuclear layer; IPL = inner plexiform layer; GCL = ganglion cell layer.

ceramide and ceramide-1-phosphate as intracellular indicators for cell death/survival decisions (Frago et al. 2003) and adding a prominent role in the regulation/modulation of inflammatory responses (Pettus et al. 2003).

Photoreceptor neurodegeneration in RP has been reported to proceed by apoptosis triggered by cellular stress. However, none of the described genetic defects underlying inherited retinal dystrophies has provided a direct link to apoptosis. It is interesting that a very recent report showed that a decrease of the intracellular ceramide pools rescued photoreceptor degeneration in *Drosophila* mutants (Acharya et al. 2003). Ceramide may be produced in the photoreceptors in response to stress. In this case, deficiency of CERKL, a specific retinal ceramide kinase, would increase the ceramide pool or decrease the ceramide-1-phosphate levels, rendering the photoreceptor cells more susceptible to stress and, thus, more sensitive to apoptotic signals. The cumulative effects of this enhanced susceptibility could lead to progressive depletion of photoreceptors and, eventually, to blindness. Failure to process the ceramide to ceramide-1-phosphate may trigger cell death not only by increasing the endogenous ceramide pool but also by failing to activate the corresponding downstream signalling pathway. Whether *CERKL* is transcriptionally modulated in response to cellular stress stimuli remains an open question.

This study—to our knowledge, the first genetic report linking RP to ceramide-induced apoptosis—highlights the SL metabolism enzymes as new functional candidates and establishes novel targets for therapeutic intervention of retinal diseases.

## Acknowledgments

We are indebted to the families with arRP who generously contributed to this study, particularly to all members of the P2 family and to Isabel Tejada from the Hospital General de Basurto (Bilbao, Spain) for providing the E1 family samples. This research followed the tenets of the Convention of Helsinki. We are grateful to Rebeca Valero and Olga Gonzàlez-Angulo, for technical help; to Susana Balcells, for valuable suggestions; and to Robin Rycroft, for revising the English. We thank the Serveis Científico-Tècnics (Universitat de Barcelona) for the sequencing and in situ hybridization facilities. This work was financed by Fundaluce (2001) and by Ministerio de Ciencia y Tecnología grant PM99-0168 (to R.G.-D.). M.T. was in receipt of a Formació d'Investigadors fellowship from the Generalitat de Catalunya.

## Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

BCM Search Launcher, http://searchlauncher.bcm.tmc.edu/

dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/ (for the new SNPs observed in the present study, which are deposited under the following identification numbers: ss12676100, ss12676101, ss12676102, ss12676103, ss12676104, ss12676105, ss12676106, ss12676107, ss12676108, ss12676109, ss12676110, ss12676111, ss12676112, ss12676113, and ss12676114)

EMBnet BoxShade Server, http://www.ch.embnet.org/software/BOX_form.html

GenBank, http://www.ncbi.nih.gov/Genbank/ (for truncated EST [accession number BE797822], partial cDNA [accession number BC020465], human SPHK1 [accession number AAF73423], murine SPHK1 [accession number AAC61697], human SPHK2 [accession number AAF74124], murine SPHK2 [accession number AAF74125], *Drosophila* SPHK [accession number AAF48045], *C. intestinalis* SPHK [accession number AK112588], yeast Lcb4p [accession number NP_014814], yeast Lcb5p [NP_013361], *A. thaliana* SPHK [accession number AY128394], human CERK [accession number AB079066], murine CERK [accession number AB079067], *C. intestinalis* CERK [accession number AK112750], *Drosophila* CERK [accession number AAF52040], human CERKL [accession number AY357073], murine CERKL [accession numbers BY742285 and BC046474], *C. elegans* CERK [accession number AAC67466], human DGKb [accession number Q9Y6T7], murine DGKb [accession number XP_147651], human DGKe [accession number NP_003638], murine DGKe [accession number NP_062378], *C. elegans* DGK1 [accession number NP_508190], *A. thaliana* DGK1 [accession number Q39017], *Drosophila* DGK2 [accession number Q09103], and *Drosophila* DGK1 [accession number Q01583])

NCBI BLAST, http://www.ncbi.nlm.nih.gov/BLAST/

NCBI Conserved Domain Database, http://www.ncbi.nih.gov/Structure/cdd/cdd.shtml

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for RP)

RetNet, http://www.sph.uth.tmc.edu/RetNet/

SNP Consortium, The, http://snp.cshl.org/

UCSC Genome Bioinformatics Web Site, http://genome.ucsc.edu/

## References

Acharya U, Patel S, Koundakjian E, Nagashima K, Han X, Acharya JK (2003) Modulating sphingolipid biosynthetic pathway rescues photoreceptor degeneration. Science 299: 1740–1743

Bayés M, Goldaracena B, Martínez-Mir A, Iragui-Madoz MI, Solans T, Chivelet P, Bussaglia E, Ramos-Arroyo MA, Baiget M, Vilageliu L, Balcells S, González-Duarte R, Grinberg D (1998) A new autosomal recessive retinitis pigmentosa locus maps on chromosome 2q31-q33. J Med Genet 35:141–145

Bayés M, Martínez-Mir A, Valverde D, del Rio E, Vilageliu L, Grinberg D, Balcells S, Ayuso C, Baiget M, Gonzàlez-Duarte R (1996) Autosomal recessive retinitis pigmentosa in Spain: evaluation of four genes and two loci involved in the disease. Clin Genet 50:380–387

Buccoliero R, Futerman AH (2003) The roles of ceramide and complex sphingolipids in neuronal cell function. Pharmacol Res 47:409–419

Collins JS, Schwartz CE (2002) Detecting polymorphisms and mutations in candidate genes. Am J Hum Genet 71:1251–1252

Dawkins JL, Hulme DJ, Brahmbhatt SB, Auer-Grumbach M, Nicholson GA (2001) Mutations in *SPTLC1*, encoding serine palmitoyltransferase, long chain base subunit-1, cause hereditary sensory neuropathy type I. Nat Genet 27:309–312

Frago LM, Cañón S, de la Rosa EJ, León Y, Varela-Nieto I (2003) Programmed cell death in the developing inner ear is balanced by nerve growth factor and insulin-like growth factor I. J Cell Sci 116:475–486

Frago LM, León Y, de la Rosa EJ, Gómez-Munoz A, Varela-Nieto I (1998) Nerve growth factor and ceramides modulate cell death in the early developing inner ear. J Cell Sci 111: 549–556

Hannun YA, Obeid LM (2002) The ceramide-centric universe of lipid-mediated cell regulation: stress encounters of the lipid kind. J Biol Chem 277:25847–25850

Hims MM, Diager SP, Inglehearn CF (2003) Retinitis pigmentosa: genes proteins and prospects. Dev Ophthalmol 37: 109–125

Luberto C, Kraveka JM, Hannun YA (2002) Ceramide regulation of apoptosis versus differentiation: a walk on a fine line. Lessons from neurobiology. Neurochem Res 27:609–617

Martínez-Mir A, Paloma E, Allikmets R, Ayuso C, del Rio T, Dean M, Vilageliu L, Gonzàlez-Duarte R, Balcells S (1998) Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene *ABCR*. Nat Genet 18:11–12

Paloma E, Hjelmqvist L, Bayés M, García-Sandoval B, Ayuso C, Balcells S, Gonzàlez-Duarte R (2000) Novel mutations in the *TULP1* gene causing autosomal recessive retinitis pigmentosa. Invest Ophthalmol Vis Sci 41:656–659

Paloma E, Martínez-Mir A, García-Sandoval B, Ayuso C, Vilageliu L, González-Duarte R, Balcells S (2002) Novel homozygous mutation in the alpha subunit of the rod cGMP gated channel (*CNGA1*) in two Spanish sibs affected with autosomal recessive retinitis pigmentosa. J Med Genet 39: E66

Pettus BJ, Bielawska A, Spiegel S, Roddy P, Hannun YA, Chalfant CE (2003) Ceramide kinase mediates cytokine and calcium ionophore-induced arachidonic acid release. J Biol Chem 278:38206–38213

Puranam KL, Guo WX, Qian WH, Nikbakht K, Boustany RM (1999) CLN3 defines a novel antiapoptotic pathway operative in neurodegeneration and mediated by ceramide. Mol Genet Metab 66:294–308

Spiegel S, Milstien S (2002) Sphingosine 1-phosphate, a key cell signaling molecule. J Biol Chem 277:25851–25854

——— (2003) Sphingosine-1-phosphate: an enigmatic signalling lipid. Nat Rev Mol Cell Biol 4:397–407

Sugiura M, Kono K, Liu H, Shimizugawa T, Minekura H, Spiegel S, Kohama T (2002) Ceramide kinase, a novel lipid kinase: molecular cloning and functional characterization. J Biol Chem 277:23294–23300

Sullivan LS, Daiger SP (1996) Inherited retinal degeneration: exceptional genetic and clinical heterogeneity. Mol Med Today 2:380–386

Thompson JD, Gibson TJ, Plewniha F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Wang Q, Chen Q, Zhao K, Wang L, Wang L, Traboulsi EI (2001) Update on the molecular genetics of retinitis pigmentosa. Ophthalmic Genet 22:133–154

# CAPÍTOL 3

## Caracterització de les variants produïdes pel processat alternatiu de transcrits del gen *CERKL*

Manuscrit al qual ha donat lloc aquest treball:

Characterization of the alternatively spliced isoforms of *CERKL*, a gene responsible for autosomal recessive retinitis pigmentosa.

En aquest darrer capítol de resultats, presentem, la caracterització inicial del gen *CERKL*, que comprèn la identificació del processat alternatiu de transcrits del gen a la retina, l'anàlisi del patró d'empalmament alternatiu en diferents teixits humans, la localització subcel·lular de la proteïna, i els assajos d'activitat enzimàtica.

A la retina humana, *CERKL* presenta diversos transcrits diferencials, quatre dels quals codifiquen proteïnes en pauta. Aquestes 4 isoformes de CERKL són la isoforma a, descrita inicialment, de 532 aminoàcids; la isoforma b, de 558 aminoàcids, que inclou un exó diferencial (E4b), entre E4 i E5; la isoforma c, de 419 aminoàcids, produïda per empalmament alternatiu entre l'exó E2 i l'exó E6; i la isoforma d, de 463 aminoàcids, produïda per empalmament alternatiu entre l'exó E3 i l'exó E6. Diversos teixits humans, tant adults com fetals, expressen alguna de les isoformes de CERKL, però, únicament a la retina, són presents les quatre.

La proteïna CERKL mostra un patró de localització subcel·lular complex, que podríem diseccionar en dos patrons bàsics: un patró nuclear, en el qual també detectem CERKL fora del nucli, i un patró citosòlic, en què no s'immunodetecta la proteïna al nucli. Ambdós són compatibles amb la topologia descrita per al metabolisme dels esfingolípids. Aquesta localització dinàmica de CERKL podria reflectir canvis als diversos *pools* de ceramida, que dependrien de l'estat cel·lular.

Finalment, també presentem en aquest capítol els resultats preliminars dels assajos d'activitat enzimàtica, que no ens permeten demostrar que la proteïna CERKL tingui activitat ceramida quinasa.

# Characterization of the alternatively spliced isoforms of *CERKL*, a gene responsible for autosomal recessive retinitis pigmentosa

Miquel Tuson, Roser Gonzàlez-Duarte and Gemma Marfany

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona.
Av. Diagonal 645, E-08028 Barcelona (Spain)

Sphingolipid metabolites are important regulators of cell growth, differentiation and death. Thus, the enzymes that control their relative levels have been the focus of intense scrutiny, particularly, lipid kinases, which phosphorylate sphingoid-bases and produce bioactive lipids with antagonistic functions to the unphosphorylated forms. Recently, we identified CERKL as one gene responsible for autosomal recessive retinitis pigmentosa. Comparison of the encoded protein revealed a putative novel lipid kinase sharing high similarity with ceramide kinases. Here we report *CERKL* in-frame alternatively spliced variants in the human retina, which encode four isoforms of 532, 558, 419 and 463 amino acids, also present in other human tissues. Over-expressed recombinant CERKL isoforms did not phosphorylate ceramide neither diacylglycerol and showed a complex subcellular pattern of localization.

## 1. Introduction

Sphingolipids (sphingoid-based lipids, SLs) were described more than a century ago by JLW Thudichum and were named –due to their enigmatic nature– after the Sphinx of the Greek mythology [1,2]. Although originally considered as structural and inert components of eukaryotic cell membranes, in the last decade SLs have been increasingly involved in crucial cellular functions as the regulation of proliferation, differentiation, survival, and death. In this context, the main products of the SL metabolism –ceramide, ceramide-1-phosphate, sphingosine and sphingosine-1-phosphate– have been characterized as lipid second messengers of a diversity of signalling pathways. In particular, ceramide —located at the core of the SL metabolism as a precursor of other bioactive and complex SLs— has been unveiled as a key player of stress-induced apoptosis. Therefore, a variety of extra- and intra-cellular stimuli —cytokines, cytotoxic agents and stress signals— cause ceramide accumulation through changes in SL metabolism; i.e. via sphingomyelin hydrolysis or *de novo* synthesis. The ceramide thus generated plays an important role in the promotion of programmed cell death. Hence, the relative levels of ceramide and other SLs act as finely tuned sensors of the cell status. Accordingly, SL enzymes have been considered rheostats or cell switches that integrate multiple stress stimuli [1], whose disruption –via genetic mutations– trigger multiple and complex phenotypic effects, among them, severe neurodegenerative disorders.

Corresponding author:
Gemma Marfany. Departament de Genètica, Facultat de Biologia, Universitat de Barcelona.
Av. Diagonal 645, E-08028 Barcelona (Spain).
Phone: +34934021502. Fax: +34934034420.
e-mail: gmarfany@ub.edu

Lately, the cloning of genes involved in ceramide metabolism has provided very valuable clues. Clearance of ceramide follows several pathways: (*a*) synthesis of sphingomyelin (SM) by SM synthase; (*b*) cleavage by ceramidases, thus producing sphingosine; (*c*) synthesis of glycosphingolipids by glucosylceramide synthase; or (*d*) ceramide phosphorylation by ceramide kinase, resulting in ceramide-1-phosphate. Again, although ceramide kinase activity had been documented more than a decade ago [3], research on the biological functions of ceramide-1-phosphate was hindered until the cloning of a gene encoding ceramide kinase [4].

Very recently, we identified a novel human ceramide kinase-like gene, *CERKL*, and described a nonsense mutation responsible for autosomal recessive retinitis pigmentosa in two Spanish families [5,6]. RP is an inherited retinal neurodegenerative disorder characterized by the progressive attrition of photoreceptor cells through apoptosis. Hence, our finding opened new scenarios for the contribution of SL metabolic alterations to retinal pathogenesis.

We here describe the isolation of four alternatively spliced variants of *CERKL* in the human retina, examine the presence of these variants in several human tissues and analyze their enzymatic activity and subcellular localization.

## 2. Materials and methods

### *2.1. Cloning of CERKL splicing variants in the human retina*

A PCR strategy on the Marathon-Ready™ human retina cDNA (BD Biosciences), using two specific primers that flanked exons 2 and 13, was devised and performed under the following conditions: 94°C for 30s, 58°C for 30s, and 72°C for 2 min for 35 cycles. The forward primer (5'-CTGTTAAACAGCAGAGAAGTGGTAC-3') annealed to exon 2, while the reverse primer (5'-AT*GGATCC*TTACTT

TGGAATCATTTCTTCCATG-3') annealed to exon 13 and introduced a *Bam*HI site (in italics) for subsequent subcloning. The reaction mixture (50 µl) contained 10 µM of each primer, 2 µM of dNTPs, 1.5 mM $MgCl_2$ and 1 U of *Pfu* Turbo polymerase (Stratagene). The PCR bands detected after electrophoresis onto a 1.5% agarose gel were cloned into the *Eco*RV site of pBluescriptII KS and verified by sequencing. The full-length cDNAs of the *CERKL* isoforms were reconstituted by digesting the partial clones with *Eco*RI and *Bam*HI and subcloning the corresponding fragments into a pBluescriptII KS vector that harboured the 5' end of *CERKL* (the full exon 1 and the stretch of exon 2 before the *Eco*RI site, obtained from the IMAGE clone 3870103).

### *2.2. RT-PCR analysis of the presence of CERKL splicing variants in human tissues*

The splicing pattern of *CERKL* was analyzed in human retina cDNA and a panel of first-strand cDNAs of several human adult and fetal tissues, using specifically designed pairs of primers to amplify and identify each isoform, as follows: 1) to detect splicing variants a and b (forward primer: 5'-GTAACAATAATGGAATATGAAG GG-3'; reverse primer: 5'-ATCCTGCTGGTATTAAGCC AAG-3'); 2) splicing variant c (forward primer: 5'-CAGTTCAAGAAAATATTGGCAGGATC-3'; reverse primer: 5'-CCTTAACAACAGCAAAATCTCTCCG-3'); 3) splicing variant d (forward primer: 5'-GGAATAAAAACT GATGTAACAAGATC-3'; reverse primer: 5'-CCTTAACAACAGCAAAATCTCTCCG -3'). Two-step PCR conditions were performed as follows: 35 cycles of 94°C for 30s and 58°C for 30s. The reaction mixture (25 µl) contained 10 µM of each primer, 2 µM of dNTPs, 1.5 mM $MgCl_2$ and 1 U *Taq* polymerase (Promega). The final concentration of the Marathon-Ready™ human retina cDNA (BD Biosciences) and of the first-strand cDNAs from the tissue panel (BD Biosciences) was 0.01 ng/µl and 0.1 ng/µl, respectively.

### *2.3. Overexpression of CERKL isoforms in cultured cells and preparation of protein cell lysates*

Construction of the expression vectors containing the full-length open reading frame of the four *CERKL* isoform was carried out by PCR amplification of each variant using a forward primer (5'-CTGTTAAACAGCAG AGAAGTGGTAC-3') located upstream of the *Eco*RI site in exon 2, and a reverse primer (5'-CG*GATATC*AAGCGT AATCTGGAACATCGTATGGGTACTTTGGAATCAT TTCTTCC -3') located in exon 13, which also encoded a carboxy-terminal HA epitope tag (underlined) and an *Eco*RV site for subsequent cloning. The PCR products were digested with *Eco*RI and *Eco*RV and cloned into a pcDNA3 (Invitrogen Life Technologies) that contained exons 1 and 2 of *CERKL* (previously cloned between *Hind*III and *Eco*RI sites). The integrity of all constructs was verified by sequencing.

Human embryonic kidney (HEK293) cells were seeded at 3 | $10^6$ cells/dish in 10-cm cell culture dishes one day before transfection. Cells were tranfected with 24 µg/dish using either pcDNA3 vector alone or each pcDNA3–CERKL construct, and 60 µl/dish Lipofectamine™2000 (Invitrogen Life Technologies), following the manufacturer's instructions. After 2 days, the cells were harvested and lysed by sonication in lysis buffer (20 mM MOPS pH 7.2, 2 mM EGTA, 1 mM dithiothreitol, 10% glycerol, and complete™protease inhibitor, Roche Diagnostics), as described elsewhere [4].

### *2.4. Immunoblotting*

Protein cell lysates (50 µg) from each transfection (pcDNA3 vector alone or each isoform pcDNA3–CERKL expression construct) were resolved on 10% SDS-PAGE under denaturing conditions and transferred to 0.45-µm polyvinylidene fluoride (PVDF) microporous membrane (Immobilon-P, Millipore). After blocking for 1 h with 5% non-fat milk in phosphate-buffered saline, 0.1% Tween 20, membranes were incubated for 1 h with an anti-HA monoclonal primary antibody (1:1000) in 0.5% non-fat milk, phosphate-buffered saline, 0.1% Tween 20 (MTP). After extensive washes in MTP, membranes were incubated for 1 h with a horseradish peroxidase-conjugated anti-mouse IgG antibody (1:3000, Sigma) and the bands were visualized using the ECL Western blotting detection system (Amersham Biosciences).

### *2.5. Kinase activity assay*

Diacylglycerol (DAG) and ceramide kinase activities were assayed as described [7]. Briefly, lysates from HEK293 cells overexpressing each *CERKL* isoform were added (250 µg total protein) to reactions (100 µl final volume) that contained 880 µM lipids (ceramide, sphingosine, both from bovine brain sphingomyelin, or 1,2-dioleoyl-sn-glycerol, Sigma), 1 mM cardiolipin, 1.5% ß-octylglucoside, 0.2 mM diethylenetriamine-pentaacetic acid, 20 mM MOPS pH 7.2, 50 mM NaCl, 1 mM dithiothreitol, 2 mM EGTA, and 3 mM $CaCl_2$ for brain ceramides (or 3.3 mM $MgCl_2$ for reactions with DAG as substrate). Reactions were started with the addition of 10 µl 5 mM $MgCl_2$, 5 mM |-$^{32}$P-ATP (2–5 µCi/reaction), incubated at 30ºC for 30 min, and stopped with 250 µl of chloroform, 250 µl of methanol and 125 µl HCl 2.4 M. After vortexing, the phases were separated by centrifugation for 5 min. The organic phase was removed and dried in a speed-vac. Once dried, lipids were resuspended in 25 µl chloroform:methanol (95:5, v/v), spotted onto silica gel plates and separated by thin-layer chromatography with chloroform:acetone:metha-nol:acetic acid:water (10:4:3:2:1, v/v). Radiolabelled lipids were visualized by autoradiography.

### *2.6. Purification of recombinant proteins*

To express the four *CERKL* isoforms as glutathione-S-transferase (GST) fusion proteins, each variant was subcloned in-frame into pGEX-4T-1 (Amersham Biosciences), between the *Xho*I and *Not*I sites. The constructs were transformed into BL21 Codon Plus *E.coli* strain (Stratagene) to avoid premature protein truncation, as the codon usage of the *CERKL* human gene is strongly biased. Production of the recombinant proteins was carried out in 500 ml cultures, and expression was induced by the addition of 0.05 mM IPTG. The GST-recombinant protein bound to the glutathione-sepharose beads was recovered by batch elution with reduced gluthatione following the manufacturer's instructions (Amersham Biosciences).

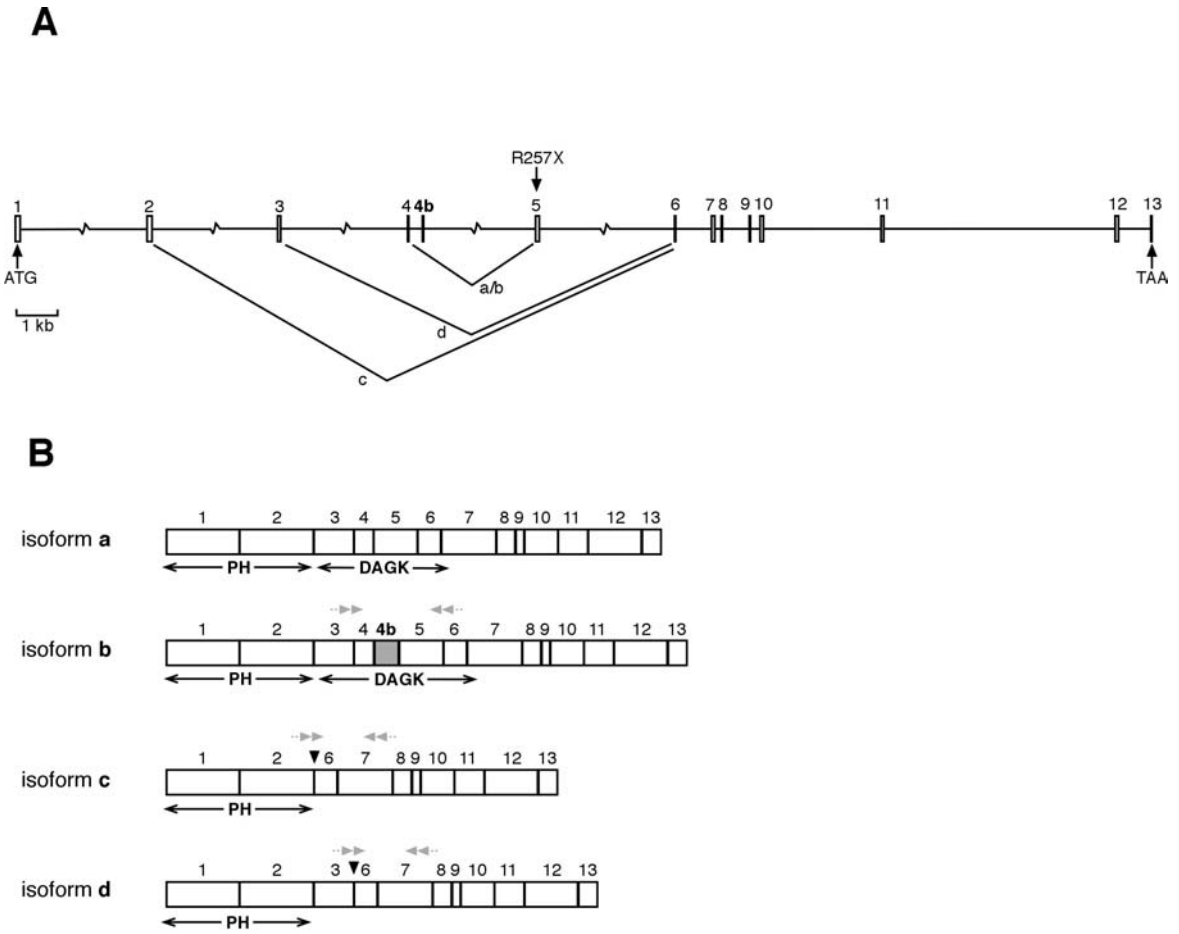



**Figure 1 | *CERKL* isoform structure**. (**A**) Scheme depicting the genomic exon-intron structure of the *CERKL* gene, with the initiation and stop codons. The position of the nonsense mutation (R257X) identified in the RP families as well as the splicing junctions resulting in isoforms a, b, c and d are indicated. (**B**) Mature *CERKL* mRNA isoforms, showing the relative position of the encoded DAG kinase and putative PH domains. Note that exon 4b (isoform b, in grey) is located well within the DAG kinase domain, while isoforms c and d lack this domain (in-frame exon junction in isoforms c and d are indicated by a black triangle). Grey small double arrows indicate the position of the primers used for specific isoform RT-PCR analyses (see Material and Methods). The figure is drawn to scale except for the introns depicted with a broken line. For the sake of clarity, disrupted (not in frame) isoforms e (fusing exons 2-5) and f (exons 3-5) are not included.

### *2.7. Lipid-protein overlay assay*

Binding of CERKL isoforms to sphingolipids was assessed using SphingoStrips™ (Echelon Biosciences), which consist of several nitrocellulose-immobilized sphingolipids (100 pmol/spot). The membranes were equilibrated in TBST [10mM Tris-HCl, pH 8.0, 150mM NaCl and 0.1% (v/v) Tween-20] for 5 min, followed by blocking with 3% Bovine Serum Albumin in TBST (blocking solution) for 1 h at room temperature. The membrane was then incubated overnight at 4ºC in blocking solution containing 0.5 μg/ml of the GST-fusion protein on a rocking platform. The membranes were washed twice for 15 min in TBST with gentle agitation, and then incubated for 1 h with and anti-GST monoclonal antibody (1:1000, Santa Cruz Biotechnology). After washing as before, membranes were incubated for 1 h with a horseradish peroxidase-conjugated anti-mouse IgG antibody (1:3000, Sigma). Finally, the membranes were washed 6 times over 1h in TBST, and developed with the ECL Western blotting detection system (Amersham Biosciences).

### *2.8. Immunolocalization of CERKL isoforms and confocal laser scanning microscopy of transiently transfected COS-7 cells*

COS-7 (African green monkey kidney) cells were grown in DMEM containing 10% fetal bovine serum, 4 mM L-glutamine, 100 U/ml of penicillin and 100 μg/ml streptomycin (Invitrogen Life Technologies). Cells were grown on coverslips in 12-well plates. Transient transfections were performed using Lipofectamine™2000 (Invitrogen Life Technologies) according to the manufacturer's protocol. Twenty-four hours or 48 h post-transfection, cells were rinsed with 100 mM PBS and fixed in 3% paraformaldehyde and 2% sucrose in 0.1 M phos-

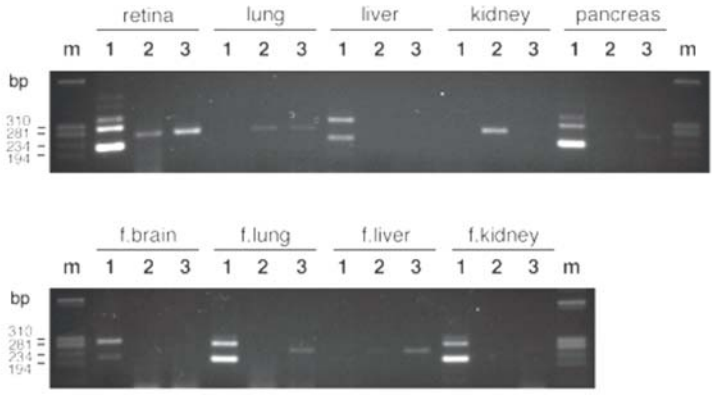



**Figure 2 | RT-PCR analysis of CERKL isoforms in several human tissues**. Lane 1 for each tissue shows amplification of isoforms a (fast migrating band) and b (slow migrating band); lane 2, amplified product of isoform c, and lane 3, isoform d. "f." stands out for fetal tissue. Marker band sizes are indicated in base pairs (bp). The additional faint bands in lanes number 1 correspond to heteroduplexes of the PCR products.

phate buffer at 4ºC for 30 min, then washed and permeabilized with 0.1% Triton X-100/20 mM glycine/10 mM PBS for 10 min. After permeabilization, cells were rinsed and blocked in 1% BSA/20 mM glycine/10 mM PBS. Then, cells were incubated with an anti-HA polyclonal antibody (1:250) and specific anti-calnexin (1:25), anti-GM130 (1:250) or anti-EEA1 (1:1000) monoclonal primary antibodies (all from BD Biosciences) at 37ºC for 1 hour. To label mitochondria, reduced MitoTracker Orange (Molecular Probes) was added to the cell culture medium at a final concentration of 500 nM for 45 min before the fixation procedure and subsequent immunocytochemistry. Upon washing, cells were incubated with AlexaFluor 488-conjugated anti-rabbit and AlexaFluor 546-conjugated anti-mouse (1:300, Molecular Probes) secondary antibodies. When required, slides were counter-stained with DAPI (Sigma) nuclear blue dye (1:200 in PBS) for 15 min. All preparations were mounted in Vectashield medium for fluorescence (Vector Laboratories) and analyzed by confocal laser scanning microscopy with Olympus Fluoview 500, and Leica TSC NT and TSC SP2 laser scanning microscopes. *In silico* predictions for CERKL subcellular localization were performed using PSORT (http://bioweb.pasteur.fr/seqanal/interfaces/psort2.html), ProSLP v2.0 (http://proslp.kisti.re.kr), ESLPred (http://www.intech.res.in/raghava/eslpred) and SubLoc v1.0 (http://www.bioinfo.tsinghua.edu.cn/SubLoc).

## 3. Results

### *3.1. CERKL transcripts are alternatively spliced in the human retina*

We attempted to clone the full-length *CERKL* cDNA on retina cDNA using primers spanning the ATG (forward) and the STOP (reverse) codons [6]. As the first exon is extremely GC-rich, optimized PCR conditions yielded minimal amplifications. However, detection of several faint bands suggested alternatively spliced variants. Therefore, we devised a two-step strategy of RT-PCR and cloning (see Material and Methods) to characterise the full-length isoforms. The RT-PCR analysis from exon 2 to 13 allowed us to identify four in-frame alternatively spliced variants of the *CERKL* gene in the human retina (Fig. 1): 1) one corresponded to the already reported variant [6], spanning 13 exons and encoding a 532 aa protein, which we have renamed CERKL isoform a (GenBank accession number AY357073); 2) a 14-exon isoform, which included an additional exon (E4b) between exons 4 and 5 and encoded a 558 aa protein (CERKL isoform b, GenBank accession number AY690329); 3) a splicing variant devoid of exons 3-4-4b-5, which encoded a 419 aa protein (CERKL isoform c, GenBank accession number AY690330); and 4) a splicing variant that skipped exons 4-4b-5 and encoded a 463 aa protein (CERKL isoform d, GenBank accession number AY690331). Two additional not in-frame splicing variants were detected, named isoforms e and f, which skipped exons 3 to 5 and 4 to 5, respectively (GenBank accession numbers AY690332 and AY690333). In these variants, the exon fusions produced a frameshift leading to a premature truncation of the encoded proteins. The longest isoforms (a and b) contained the full diacylglycerol (DAG) kinase and the putative N-terminal Pleckstrin Homology (PH) domains, whereas the short forms (c and d) skipped most of the DAG kinase domain, although the putative PH domain (encoded in exons 1 and 2) was preserved (Fig. 1).

### *3.2. CERKL complex alternative splicing pattern in human tissues*

Our previous data on CERKL expression in a variety of human adult and fetal tissues [6] prompted us to analyse CERKL alternatively spliced products in tissues other than the retina. To this end, we used a panel of first strand cDNAs from several human tissues. In accordance to our previous data, neither CERKL isoform could be found in adult skeletal and heart muscle. However, some of the four in-frame splicing variants were detected in other adult and fetal tissues (Fig. 2). Isoforms a and b (with an intact kinase domain) were present in adult liver and pancreas, as well as in fetal brain, lung and kidney. One or both short isoforms (c and d) were detected in adult lung, kidney, pancreas, fetal lung and fetal liver. Notably, retina was the only tissue in which all isoforms were detected under the tested conditions, although after a saturating number of PCR cycles basal levels of most isoforms could be also observed in other tissues (data not shown).

### *3.3. Is CERKL a ceramide kinase?*

Given the structural similarities of CERKL to the reported ceramide kinases, a series of experiments were conducted to assess ceramide kinase activity. Overexpression of each CERKL isoform, HA-tagged at the

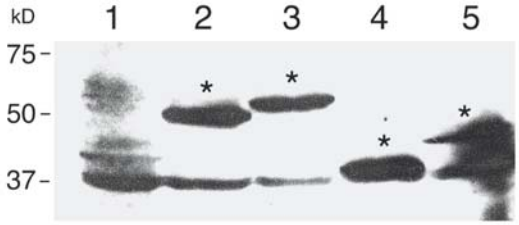



**Figure 3 | Transiently transfected HEK293 cells express high levels of the HA-tagged CERKL isoforms**. Lane 1- HEK293 cells transfected with empty pcDNA3.1 (negative control); lanes 2, 3, 4 and 5- HEK293 cells transfected with either pcDNA3.1-CERKL isoform a, b, c or d, respectively. CERKL proteins are shown with an asterisk.

C-terminus for immunodetection, was achieved by transient transfections on cultured HEK293 cells, whose negligible endogenous ceramide kinase levels make them suitable for this type of analysis [4]. Protein lysates for each isoform were obtained 48 h post-transfection, quantified and kept frozen at –20ºC, until assayed. The expression of each recombinant protein was verified by Western blot of the lysates with a monoclonal anti-HA antibody before the assay (Fig. 3).

The protocol for ceramide kinase activity was adapted from [7] with minor modifications. No kinase activity could be detected for any isoform at several protein lysate concentrations, with either ceramide from bovine brain sphingomyelin, or 1,2-dioleoyl-sn-glycerol (data not shown). However, positive controls to verify the integrity of the protein lysate, with sphingosine as substrate, detected endogenous sphingosine kinase activity (data not shown).

The absence of kinase activity prompted us to test the ability of CERKL isoforms to bind sphingolipids by protein-lipid overlay. We used the CERKL isoform a (532 aa, containing the full DAG kinase domain) either as a pure preparation of recombinant GST-CERKL fusion protein (in *E.coli*) or as total fresh protein lysate from HEK293 cultured cells transiently transfected with the HA-tagged construct. After incubating the protein extracts with the lipid-membrane strips, the corresponding primary (monoclonal anti-GST or monoclonal anti-HA) and secondary antibodies were added to identify protein-lipid binding. Remarkably, no positive binding signals were produced, neither with ceramide, sphingosine, sphingosine-1-phosphate, nor other more complex lipids, such as gangliosides and sphingomyelin.

### *3.4. Subcellular CERKL localization is highly dynamic*

Ceramide and other sphingoid lipids are highly dynamic membrane components. Proteins involved in sphingolipid metabolism appear to be mainly localised in subcellular compartments of the secretory pathway, from the ER-Golgi network to the plasma membrane. In this context, we aimed to determine CERKL subcellular localization. Besides, assays with the four isoforms were performed as the specific domain architecture displayed by each variant could account for subtle nuances in function or compartment localization.

Interestingly, neither a unique localization pattern nor a restricted subcellular localization could be ascribed to any CERKL isoform, as all variants showed similar cell-specific patterns. Two main protein localizations emerged consistently in all preparations (Fig. 4A–D). One of them showed CERKL finely distributed through the cytosol compartment and the nucleus, either distinctly co-localizing to, or clearly excluded from, the nucleoli (Fig. 4A, B and D). The other highlighted perinuclear CERKL distribution, while the nucleus and the rest of cytosol were practically devoid of signal (Fig. 4C and D). No clear-cut co-localization with reference markers of subcellular compartments, such as the ER (calnexin) or the Golgi matrix (GM130), could be observed in any case (Fig. 4E), as the strongest marker signal did not correlate to the highest CERKL accumulation. To further assess CERKL association to other membraneous compartments, we assayed co-localization to endosomes (EEA1) or mitochondria (labelled by MitoTracker staining) (Fig. 4E). Again, no precise colocalization to any of the markers of the above compartments was observed. Overall, CERKL localized adjacent to, but not overlapping with, the ER, Golgi and mitochondria compartments (Fig. 4E).

Similar CERKL subcellular localizations were obtained at 24 h or 48 h post-transfection, although a shift towards the perinuclear accumulation pattern was clearly observed. These results point to a highly dynamic localization of CERKL, which may relocate according to the stimuli acting upon a cell.

## 4. Discussion

As more functional information is being gathered for human genes, there is mounting evidence that new regulatory layers of gene function rely on alternatively spliced isoforms, many with tissular relevance and that, when disturbed, lead to severe genetic diseases (reviewed in [8]). We here report the characterization of four in-frame alternatively spliced isoforms of *CERKL*, a gene responsible for an autosomal recessive form of retinitis pigmentosa (RP26). In our case, isoform multiplicity increases the complexity of the functional analysis of an as yet uncharacterised lipid kinase. For instance, comparison of isoforms a and b reveals an additional exon (E4b), which introduces 26 extra amino acids in-frame within the DAG kinase catalytic domain, immediately upstream of the phosphate-donor binding site (GGDGS). The position of this exon in the presumptive catalytic domain makes it a good candidate to modulate the catalytic, substrate- or partner- binding abilities of the protein, as has been recently postulated for alternatively-spliced exons within encoded functional domains after an *in silico* genomic survey [9].

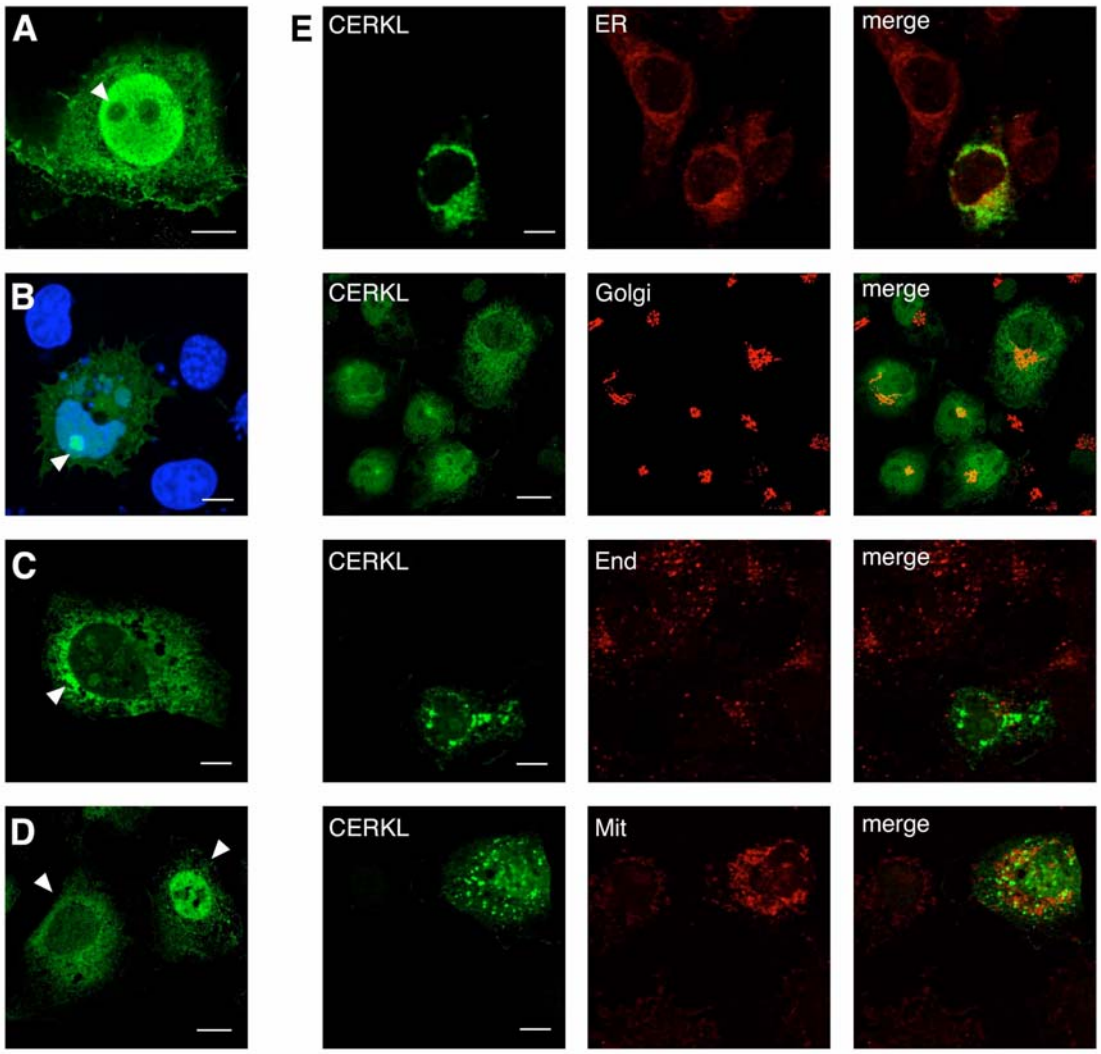



**Figure 4 | CERKL shows a dynamic subcellular localization**. (**A**) Some cells show uniform distribution of CERKL in the cytosol and the nucleus, with clear exclusion from the nucleoli, (**B**) a similar pattern but with strong accumulation in the nucleoli (nuclei, counter-stained with DAPI, appear in blue), (**C**) in other cells CERKL is absent from the nucleus and instead, accumulates in clusters, preferentially in the perinuclear region, (**D**) the two patterns are cell-specific and can be observed in the same field. (**E**) CERKL does not colocalize with but is associated to reference markers of the membraneous subcellular compartments (calnexin for ER, GM130 for Golgi, EEA1 for endosomes and MitoTracker for mitochondria). ER (endoplasmic reticulum), End (endosomes), Mit (mitochondria). Arrows highlight the relevant CERKL localizations. Scale bar corresponds to 10 μm.

Interestingly, the other two in-frame isoforms, c and d, lack the exons encoding the DAG kinase domain, which should severely compromise the catalytic function while preserving other segments involved in substrate recognition. It is remarkable that the reported RP mutation (R257X) lies within the *CERKL* exon 5, skipped in the short variants (isoforms c and d). Thus, these two variants are most probably expressed in the RP26 family patients. In this context, the physiological relevance of the CERKL alternative isoforms and their contribution to retinal disorders are yet to be determined, as retinal degeneration is the only phenotypic trait associated to the R257X nonsense mutation, even though CERKL is expressed in many adult and fetal tissues.

From sequence comparisons and phylogenetic analysis, CERKL unambiguously clusters within the ceramide kinase subfamily of lipid kinases [6]. However, no experimental evidence could be gathered to support kinase activity on ceramide and diacylglycerol. It may well be that, although not yet identified for other lipid kinases, a partner is required for substrate recognition or catalysis.

The subcellular localization of CERKL, common to all the isoforms, reveals a complex and highly dynamic pattern, which may reflect changes in the intracellular sphingolipid pools. The *in silico* predictions of subcellular localization on the full-length protein do not define any sorting signal domain, although specific domains of the protein could account for different protein localizations, as it is the case for the ceramide transfer protein [10,11]. In this respect, the first *CERKL* exon encodes a string of basic residues compatible with a nuclear localization signal. Whether the nuclear localization is an intrinsic property of the CERKL protein or depends on the binding of other partners or translocators remains to be elucidated. Indeed,

intranuclear localization of CERKL may be related to the regulation of the chromatin-associated ceramide and sphingomyelin pools, in accordance to the intranuclear distribution of other key enzymes of the ceramide metabolism, such as SMase [12] and ceramidase [13]. In addition, the consistently observed close proximity, but with no overlap, to the ER, Golgi and mitochondria subcellular compartments, is compatible with the membrane-associated activity reported for lipid kinases. It has been claimed that compartmentalization of ceramide function requires transfer between different membraneous organelles [10,14] and, on the other hand, ceramide kinase activity has also been detected in the cytosol, not associated to membranes [7]. Together these observations suggest that SL metabolic enzymes translocate between different compartments. In this scenario, the observed variable CERKL intracellular distribution may not be so surprising, particularly when considering that similar protein localization dynamics –which shift depending on the cellular state or stimuli– are now being reported for other proteins involved in lipid metabolism [10,15].

Until recently, most enzymes of the sphingolipid metabolism were approached through classical biochemical purification and characterization. Nonetheless, many interesting contributions to the field are now emerging from the reverse genetics approach, particularly since the availability of full genomic data from model organisms, from yeast to human [4,16]. The challenge to ascribe function to the newly identified sequences may offer unexpected insights onto the bioactive lipids research, such as the recently described multiple lipid kinase [17].

Remarkably, and contrary to other lipid enzymes, which are spread throughout all the eukaryotes, CERKL may be an exclusive innovation in the vertebrate lineage. Highly homologous sequences are being found in chick and fishes but no evidence whatsoever has been found in fully sequenced genomes of model invertebrates, such as *Drosophila* and *C. elegans*, neither in basal eukaryotes, such as yeast (Tuson, unpublished results). In this context, new enzymes with novel functions may add further refinement to the already complex metabolism of bioactive lipids.

## Aknowledgements

We are grateful to Yolanda León and Isabel Varela-Nieto for invaluable help in the ceramide kinase activity assays, Esther Pomares for technical help, Neus Cols for helpful comments concerning the recombinant protein production, and the Serveis Científico-Tècnics de la Universitat de Barcelona for the Sequencing and Confocal microscopy facilities. We thank Robin Rycroft and the Servei d'Assessorament Lingüístic for revising the English. This study was funded by Fundaluce (2004), Ayuda para la Investigación (ONCE, 2004) and grant BMC2003-05211 (Ministerio de Ciencia y Tecnología) to R.G.-D.

## References

[1] Spiegel, S. and Milstien, S. (2003) Nat Rev Mol Cell Biol 4, 397-407.
[2] Yang, J., Yu, Y., Sun, S. and Duerksen-Hughes, P.J. (2004) Cell Biochem Biophys 40, 323-50.
[3] Bajjalieh, S.M., Martin, T.F. and Floor, E. (1989) J Biol Chem 264, 14354-60.
[4] Sugiura, M., Kono, K., Liu, H., Shimizugawa, T., Minekura, H., Spiegel, S. and Kohama, T. (2002) J Biol Chem 277, 23294-300.
[5] Bayés, M. et al. (1998) J Med Genet 35, 141-5.
[6] Tuson, M., Marfany, G. and Gonzalez-Duarte, R. (2004) Am J Hum Genet 74, 128-38.
[7] Bajjalieh, S. and Batchelor, R. (2000) Methods Enzymol 311, 207-15.
[8] Garcia-Blanco, M.A., Baraniak, A.P. and Lasda, E.L. (2004) Nat Biotechnol 22, 535-46.
[9] Wen, F., Li, F., Xia, H., Lu, X., Zhang, X. and Li, Y. (2004) Trends Genet 20, 232-6.
[10] Hanada, K., Kumagai, K., Yasuda, S., Miura, Y., Kawano, M., Fukasawa, M. and Nishijima, M. (2003) Nature 426, 803-9.
[11] Munro, S. (2003) Nature 426, 775-6.
[12] Albi, E., Lazzarini, R. and Magni, M.V. (2003) FEBS Lett 549, 152-6.
[13] Tsugane, K., Tamiya-Koizumi, K., Nagino, M., Nimura, Y. and Yoshida, S. (1999) J Hepatol 31, 8-17.
[14] Hannun, Y.A. and Luberto, C. (2004) Curr Biol 14, R163-5.
[15] Irvine, R. (2004) Curr Biol 14, R308-10.
[16] Obeid, L.M., Okamoto, Y. and Mao, C. (2002) Biochim Biophys Acta 1585, 163-71.
[17] Waggoner, D.W., Johnson, L.B., Mann, P.C., Morris, V., Guastella, J. and Bajjalieh, S.M. (2004) J Biol Chem.