CLONING AND FUNCTIONAL CHARACTERIZATION OF NOVEL GENES EXPRESSED PREFERENTIALLY IN THE HUMAN RETINA

DISSERTATION ZUR ERLANGUNG DES

NATURWISSENSCHAFTLICHEN DOKTORGRADES

DER BAYERISCHEN JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

VORGELEGT VON

JELENA STOJIC

AUS BELGRAD

SERBIEN UND MONTENEGRO

WÜRZBURG, 2005

Eingereicht am: 21.02.2005.

Bei der Fakultät für Biologie

Mitglieder der Promotionskommission:

Vorsitzender: Prof. Dr. Ulrich Scheer

Gutachter: Prof. Dr. Bernhard Weber

Gutachter: Prof. Dr. Roy Gross

Tag des Promotionskolloquiums: 08.06.2005.

Doktorurkunde ausgehändigt am: ………………….

Erklärung gemäß §4, Absatz 3 der Promotionsordnung für die Fakultät für Biologie der Universität Würzburg:

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig durchgeführt und verfasst habe.

Andere Quellen als die angegebenen Hilfsmittel und Quellen wurden nicht verwendet.

Die Dissertation wurde weder in gleicher noch in ähnlicher Form in einem anderen Prüfungsverfahren vorgelegt.

Es wurde zuvor kein anderer akademischer Grad erworben.

Die vorliegende Arbeit wurde am Institut für Humangenetik der Universität Würzburg unter der Leitung von Prof. Bernhard H.F. Weber angefertigt.

Würzburg, 2005

Jelena Stojic

ACKNOWLEDGMENTS

TABLE OF CONTENTS

1 SUMMARY

The human retina is a multi-layered neuronal tissue specialized for the reception and processing of visual information. The retina is composed of a great diversity of neuronal cell types including rod and cone photoreceptors, bipolar cells, ganglion cells, amacrine cells, horizontal cells and Müller glia. In response to light, a coordinated series of molecular events, the so-called phototransduction cascade, is triggered in photoreceptor cells and the signals from the photoreceptors are further processed by the bipolar and ganglion cells to the higher centers of the brain.

The retina as highly complex system may be greatly susceptible to genetic defects which can lead to a wide range of disease phenotypes. Therefore, isolation and characterisation of the genes active in the human retina will facilitate our deeper understanding of retinal physiology and mechanisms underlying retinal degeneration and provide novel candidates for the retinal disease genes. To identify novel genes that are specifically or predominantly expressed in the human retina, a cDNA library enriched for retina specific transcripts was generated using suppression subtractive hybridization (SSH) technique. In total, 1113 clones were randomly isolated from the retina SSH cDNA library and partially sequenced. On the basis of BLASTN algorithm analysis these clones were classified into four categories including those with I) significant homology to known human genes (766/1113), II) significant homology to partial transcripts and hypothetical gene predictions (162/1113), III) no homology to known mRNAs (149/1113), and IV) vector sequences and clones derived from mitochondrial genes (36/1113). After correcting for redundancy, category I represented 234 known human genes and category II a total of 92 unknown transcripts.

Clones from category I, were selected for expression analysis by RT-PCR in a great number of human tissues. This resulted in the identification of 16 genes which were expressed exclusively in the retina, 13 which were highly expressed in the retina compared to other tissues, 12 genes which were specifically expressed in neuronal tissues and 48 ubiquitously expressed genes. Thus, our expression analysis resulted in the identification of 29 genes exclusively or abundantly transcribed in the human retina. Of those, retina specific genes L25, L33, L35, L37, L38 and L40 were selected for further analysis. To characterize the complete mRNA sequences of these transcripts a full-length human retina cDNA library was constructed.

The analysis of the L25 gene revealed three splicing variants of the ABCC5 gene, consequently named ABCC5_SV1 (SV1), ABCC5_SV2 (SV2) and ABCC5_SV3 (SV3). These isoforms comprise the first five exons of ABCC5 and additional novel exons named 5a, 5b and 5c, generated by differential exon usage. The determined lengths of the three transcripts are 2039 bp, 1962 bp, and 1887 bp in size, respectively. RT-PCR, real-time PCR and Northern blot analysis of ABCC5 as well as the isoforms SV1, SV2 and SV3 demonstrated high levels of expression for all transcripts in the retina compared to other tissues. Analysis of their nucleotide sequences revealed that inclusion of exon 5a in splicing variant SV1 produced a frame shift and premature termination codon (PTC). Our data show that this splice variant is the target of nonsense mediated mRNA decay (NMD). This was shown by inhibition of protein synthesis with antibiotics puromycin and anisomycin in human cell lines A-RPE 19 and Y79. Our analysis resulted in an increase of the PTC containing transcript and a decrease of the ABCC5 transcript. Conversely, the amount of both transcripts (SV1 and ABCC5) returned to pre-treatment levels after removal of the inhibitors. Together, our results suggest that alternative splicing of the ubiquitously expressed ABCC5 gene in addition to NMD is involved in retina-specific transcriptional regulation of the mRNA level of ABCC5. In contrast, additional experiments demonstrated that the levels of expression of SV2 and SV3 isoforms do not appear to influence ABCC5 transcription.

The analysis of cDNA clones derived from the L33 transcript lead to the identification of the full-length cDNA sequence of 1663 bp in size. The L33 gene consists of five exons and spans 13 Kb of genomic DNA on chromosome 15q23. Further expression analysis of this gene confirmed its predominant transcription in retina with lower levels of expression in several other human tissues. The longest putative ORFs in the transcript have the capacity to encode peptides ranging from 28 to 42 amino acids.

The analysis of the L35 gene resulted in the comprehensive identification of two isoforms, termed L35_SV1 and L35_SV2, with full-length sizes of 3759 bp and 3618 bp, respectively. L35 is a four exon gene located on chromosome 7p21 and spanning 13 Kb of genomic sequence. It is located in its entirety in the first intron of the MEOX2 gene, and is encoded by the opposite strand. Expression analysis of the L35 gene demonstrated that it is highly expressed in the retina, with low levels of expression in skeletal muscle and testis. The longest predicted ORF within this cDNA sequence encodes only 44 amino acids.

The L38 gene was identified as a five exon gene on chromosome 12q21, with 1164 bp in size. Virtual Northern blot analysis confirmed its preferential expression in the retina. The longest ORFs predicted encode peptides ranging in size from 27 to 34 amino acids. In

summary, our data from the analysis of the nucleotide sequences of the genes L33, L35 and L38 and their putative coding potential strongly suggest that these transcripts may not represent protein-coding genes but rather may function as RNA-encoding transcripts involved in transcriptional regulation of gene expression.

The full-length transcript of L37 is 419 bp in size. The L37 gene consists of two exons separated by approximately 30 Kb of intervening sequence. This gene is located within the 36 Kb long first intron of ATP2C1, with opposite orientation. Virtual Northern blot confirmed its exclusive expression in the retina. Analysis of the coding potential of this gene predicts two putative proteins consisting of 33 and 35 amino acid residues, respectively, with no homology to any known human protein sequences or any known domains.

The L40 gene transcribes a full-length cDNA sequence with 1.037 bp in size. This gene spans approximately 156 Kb of genomic DNA and is divided in seven exons. Three 5'-terminal exons of L40 (exons 2-4) overlap to 3'-terminal exons (exons 5-7) of the Zona pellucida-binding protein (ZPBP) gene. Expression analysis by RT-PCR, with forward primers located in the respective unique exons of the two genes and with reverse primers located in the first shared exon showed high levels of expression of L40 in the retina with minor expression in RPE, whereas expression of ZPBP was specific for the testis. Analysis of the L40 nucleotide sequence indicated that its largest open reading frame is 333 bp coding for 111 amino acid residues by exons 2-5. The L40 overlaps ZPBP between coding sequences and both genes share a common stretch of 100 amino acids. This study revealed that two major mRNA species of the ZPBP and L40 genes sharing three exons but differing in their 5'-untranslated regions, can be generated in a tissue-specific manner through the use of tissue-specific promoters.

Several of these cloned genes were selected for additional genotyping of single nucleotide polymorphisms (SNPs) in order to construct their SNP maps which are going to be used for future association studies of complex disease AMD. Thus, identification of novel retinal genes and their functional characterization will further our elucidation of retinal physiology in general and in the diseased state in particular, by providing candidate retinal disease genes.

2   ZUSAMMENFASSUNG

Die menschliche Retina ist ein vielschichtiges neuronales Gewebe, das auf die Aufnahme und Verarbeitung visueller Informationen spezialisiert ist. Sie besteht aus einer großen Vielfalt neuronaler Zelltypen, inklusive der Stäbchen und Zapfen (Photorezeptoren), Bipolar- und Ganglionzellen, Amakrinen Zellen sowie Horizontalzellen und Müllerglia. Als Antwort auf einen Lichteinfall wird eine koordinierte Serie von molekularen Ereignissen, die so genannte Phototransduktionskaskade, in den Photorezeptoren ausgelöst, die von den Photorezeptoren kommenden Signale von den Bipolar- und Ganglionzellen weiterverarbeitet und an höhere Hirnzentren gesandt.

Als äußerst komplexes System kann die Retina hochanfällig für genetische Defekte sein, die zu einem breiten Spektrum an Krankheitsphänotypen führen können. Deshalb wird die Isolation und Charakterisierung von in der Retina aktiven Genen unser tieferes Verständnis für die Physiologie der Netzhaut, sowie den einer Retinadegeneration zu Grunde liegenden Mechanismen, erleichtern. Um neue Gene, die spezifisch oder überwiegend in der Netzhaut exprimiert werden, identifizieren zu können wurde, unter Nutzung der subtractive hybridization technique (SSH) eine cDNA Bibliothek, angereichert mit retinaspezifischen Genen, generiert. Insgesamt wurden 1113 Klone zufällig aus der Retina SSH cDNA Bibliothek ausgewählt und teilweise sequenziert. Auf der Basis einer BLASTN Algorithmus Analyse wurden diese Klone in 4 Kategorien eingeteilt, wobei I) Klone mit signifikanten Homologien zu bekannten menschlichen Genen (766/1113) beinhaltet, II) Klone mit signifikanten Homologien zu partiellen Transkripten und hypothetischen Genvorhersagen (162/1113), III) Klone ohne Homologie zu bekannten mRNAs (149/1113) und IV) Vektorsequenzen und Klone, erhalten von mitochondrialen Genen (36/1113). Nach Redundanzkorrektur repräsentierte Kategorie I 234 bekannte menschliche Gene und Kategorie II insgesamt 92 unbekannte Transkripte.

Klone der Kategorie I wurden für eine Expressionsanalyse mittels RT-PCR in einer Vielzahl menschlicher Gewebe ausgewählt. Dies führte zur Identifikation von 16 exklusiv in der Retina exprimierten Genen, weiteren 13 mit, im Vergleich zu anderen Geweben, hoher Expression in der Netzhaut, 12 Genen, die spezifisch in neuronalen Geweben exprimiert werden, sowie 48 ubiquitär exprimierten Genen. Somit resultierte unsere Expressionsanalyse in der Identifikation von 29 exklusiv oder reichlich in der menschlichen Retina transkribierten Genen. Aus diesen wurden die retinaspezifischen Gene L25, L33, L35, L38 und L40 für

weitere Analysen ausgewählt. Um die kompletten mRNA Sequenzen dieser Transkripte zu charakterisieren wurde eine full-length human retina cDNA Bibliothek konstruiert.

Die Analyse des L25 Genes erbrachte 3 Splicevarianten des ABCC5 Genes, die konsequent mit ABCC5_SV1 (SV1), ABCC5_SV2 (SV2) und ABCC5_SV3 (SV3) benannt wurden. Diese Isoformen bestehen aus den ersten 5 Exonen von ABCC5 und zusätzlichen neuen 5a, 5b und 5c genannten Exonen, die durch unterschiedliche Exonnutzung entstehen. Die ermittelten Längen der drei Transkripte sind 2039 bp, 1962 bp, und 1887 bp. RT-PCR, real time PCR und Northern Blot Analyse des ABCC5 Genes sowie der Isoformen SV1, SV2 und SV3, konnten hohe Expressionslevel all dieser Transkripte in der Retina im Vergleich zu anderen Geweben zeigen. Die Analyse ihrer Nukleotidsequenzen enthüllte, dass der Einschluss von Exon 5a in Splicevariante SV1 eine Leserahmenverschiebung sowie die Entstehung eines premature termination Kodons (PTC) bewirkt. Unsere Daten zeigen, dass diese Splicevariante Ziel eines nonsense mediated mRNA decay (NMD) ist. Dies wurde durch Hemmung der Proteinsynthese mittels Gabe der Antibiotika Puromycin und Anisomycin in den humanen Zelllinien ARPE-19 und Y79 gezeigt. Es kam bei diesem Versuch zu einem Anstieg des PTC enthaltenden Transkripts und einem Abfall des ABCC5 Transkripts. Umgekehrt ging die Menge beider Transkripte (SV1 und ABCC5) nach Wegnahme der Inhibitoren auf die Ausgangswerte vor Antibiotikagabe zurück. Zusammengefasst weisen unsere Ergebnisse darauf hin, dass alternatives Splicing des ubiqitär exprimierten ABCC5 Genes zusätzlich zu NMD in der retinaspezifischen Transkriptionsregulation der mRNA Level des ABCC5 Genes involviert ist. Im Gegensatz dazu zeigten zusätzliche Experimente, dass die Expressionslevel der SV2 und SV3 Isoformen die ABCC5 Transkription scheinbar nicht beeinflussen.

Die Analyse von aus dem L33 Transkript erhaltenen cDNA Klonen führte zur Identifikation einer full-length cDNA Sequenz von 1663 bp Größe. Das L33 Gen besteht aus 5 Exonen und überspannt 13 Kb an genomischer DNA auf Chromosom 15q23. Weitere Expressionsanalysen dieses Genes bestätigten dessen dominante Expression in der Netzhaut gegenüber geringeren Expressionsleveln in einigen anderen menschlichen Geweben. Die längsten vermeintlichen ORFs dieses Transkripts haben die Fähigkeit, Peptide von 28 bis 48 Aminosäuren Länge zu generieren.

Die Analyse des L35 Gens resultierte in der umfassenden Identifizierung zweier Isoformen, benannt L35_SV1 und L35_SV2, mit einer Gesamtlänge von 3759 bp bzw. 3618 bp. L35 ist ein aus vier Exonen bestehendes Gen, lokalisiert auf Chromosom 7p21, und überspannt 13 Kb an genomischer Sequenz. Es ist vollständig im ersten Intron des MEOX2

Gens lokalisiert und wird auf dem Gegenstrang kodiert. Eine Expressionsanalyse des L35 Gens zeigte, dass es äußerst stark in der Retina exprimiert wird, sowie in geringer Menge im Skelettmuskel und den Hoden. Der längste innerhalb dieser cDNA Sequenz vorhergesagte ORF kodiert lediglich 44 Aminosäuren.

L38 wurde als ein aus fünf Exonen bestehendes Gen auf Chromosom 12q21 mit einer Länge von 1164 bp identifiziert. Northern blot Analyse bestätigte seine bevorzugte Expression in der Retina. Die längsten vorhergesagten ORFs kodieren Peptide von 27 bis 34 Aminosäuren Länge. Zusammengefasst lassen unsere Analysedaten der Nukleotidsequenzen der Gene L33, L35 und L38 und ihre mutmaßlichen Kodierungspotentiale stark darauf schließen, dass diese Transkripte keine Protein-kodierenden Gene repräsentieren, sondern eher als RNA kodierende Transkripte fungieren, die in der Transkriptionsregulation der Genexpression involviert sind.

Das full-length Transkript von L37 ist 419 bp groß. Das L37 Gen besteht aus zwei Exonen, getrennt durch ungefähr 30 Kb intervenierender Sequenz. Dieses Gen ist in gegenläufiger Orientierung innerhalb der ersten 36 kb langen Introns von ATP2C1 lokalisiert. Virtuelle Northern Blot Analyse bestätigte seine alleinige Expression in der Netzhaut. Die Analyse des Kodierungspotentials dieses Gens sagt zwei vermeintliche Proteine vorher, bestehend aus 33 bzw. 35 Aminosäureresten ohne Homologie zu bekannten humanen Proteinsequenzen oder Proteindomänen.

Das L40 Gen transkribiert eine full-length cDNA mit 1.037 bp Länge. Dieses Gen überspannt ungefähr 156 Kb an genomischer DNA und ist in sieben Exone unterteilt. Drei 5´-terminale Exone des L40 (Exone 2- 4) überlappen mit 3´-terminalen Exonen (Exone 5- 7) des Zona pellucida-binding protein (ZPBP) Gens. Eine Expressionsanlyse mittels RT- PCR, bei der die forward Primer in den jeweiligen einzigartigen Exonen der zwei Gene lagen und die reversen Primer in den ersten gemeinsamen Exonen, zeigte hohe Expressionslevel des L40 in der Retina mit geringerer Expression im RPE, wohingegen die Expression des ZPBP spezifische für den Hoden war. Die Analyse der L40 Nukleotidsequenz ließ erkennen, dass sein größter offener Leserahmen von 333 bp 111 Aminosäurereste mittels der Exone 2-5 kodiert. L40 überlappt ZPBP zwischen den kodierenden Sequenzen und beide Gene teilen sich einen gemeinsamen Abschnitt von 100 Aminosäuren. Diese Studie deckte auf, dass zwei Haupt- mRNAs des ZPBP und L40 Gens, die sich drei Exone teilen, sich aber in ihren untranslatierten 5´- Bereichen unterscheiden, durch Nutzung gewebespezifischer Promotoren in gewebsspezifischer Weise generiert werden können.

Einige dieser klonierten Gene wurden zur zusätzliche Genotypisierungen ihrer single nucleotide Polymorphismen (SNPs) ausgewählt, mit dem Ziel SNP Karten zu erstellen, die für spätere Assoziationsstudien komplexer AMD Erkrankungen verwendet werden sollen. Auf diese Weise wird die Identifizierung neuer retinaler Gene sowie deren Funktionscharakterisierung die Aufklärung der Netzhautphysiologie im Allgemeinen und deren krankhafter Veränderungen im besonderen fördern, indem sie Kandidatengene für Retinaerkrankungen zu Verfügung stellt.

3   INTRODUCTION

3.1   Organization of the retina

The human retina is a multi-layered neuronal tissue specialized for the reception and processing of visual information. Vision begins with the conversion of light from the outside world into neural signals which can be processed within the retina and sent through the optic nerve to the brain higher centres. Light enters the eye by first passing through the cornea, is then focused by the lens, further passes through the vitreous humor and all intervening neural layers in the retina, and is finally absorbed by the pigments of the photoreceptor cells (Fig. 1). The population of photoreceptors responsible for reading and colour vision is mainly concentrated in the fovea (0.35 mm in diameter), which is located in the center of the macula (6 mm in diameter) (Tuo et al., 2004). As shown in Figure 1, this highly specific region of the macula is located in the central retina temporally to the optic disc.
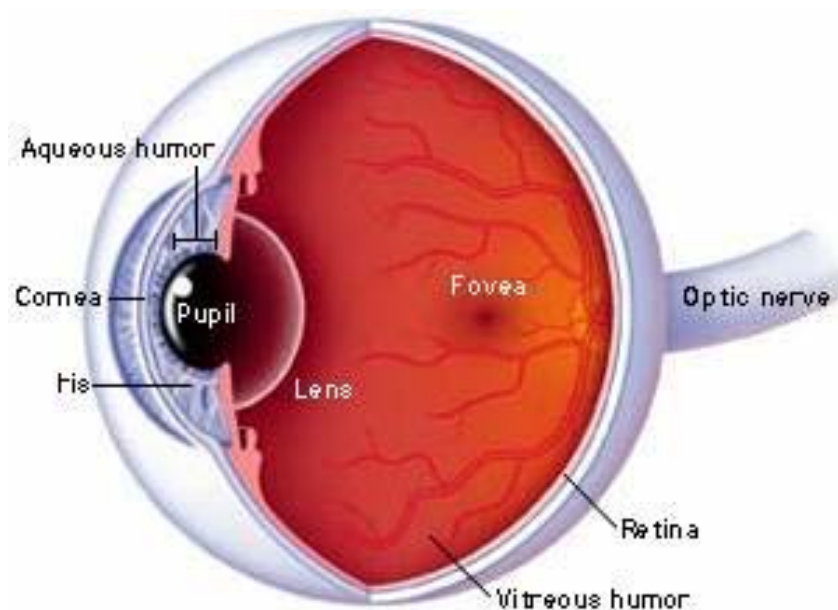
Figure 1. Schematic drawing of the human eye

(figure taken from http://www.brainconnection.com/topics/?main=anat/vision-anat)

The retina contains huge diversity of neuronal and non-neuronal cells, like many other central nervous system structures. Mammalian retinas have approximately 55 distinct cell types and each of them have different function (Masland, 2001). There are seven main classes of cells found in the vertebrate retina, including rod and cone photoreceptors, bipolar cells, ganglion cells, amacrine cells, horizontal cells and Müller glia (Fig. 2). For example, amacrine cells are the most diverse class within the retina with 29 different subtypes (Masland, 2001).
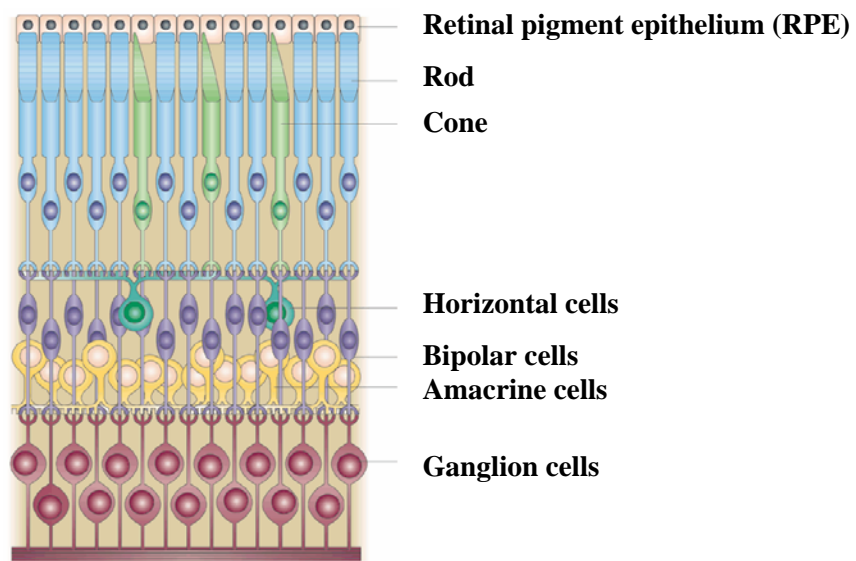
Figure 2. Structure of the adult mammalian retina. Main classes of cell types found in the vertebrate retina including rod and cone photoreceptors, bipolar cells, ganglion cells, amacrine cells and horizontal cells. There are many distinct subtypes within each class of neuron. For example, amacrine cells are the most diverse class of the retinal cells with at least 22 different subtypes in the rabbit retina (figure taken from Livesey and Cepko, 2001).

The cell bodies of the neuronal cells are organized in three different layers including inner nuclear, outer nuclear and ganglion cell layers (the basic organization of the retinal cell layers is shown in Figure 3.). Two synaptic layers, outer and inner plexiform layer, are interacting between three cellular layers.

The outer nuclear layer of the retina is composed of photoreceptor cell bodies (Fig. 3). When stimulated, rods and cones transmit their signals through bipolar cells, whose cell bodies are found in the inner nuclear layer, to the ganglion cells. From the ganglion cell layer, these signals make their way along the optic nerve to the brain. Horizontal and amacrine cells in the inner nuclear layer make lateral connections and modulate the direct signalling pathway

from photoreceptors to ganglion cells. Amacrine cells also make up a significant proportion of the cells in the ganglion cell layer and are referred to as displaced amacrine cells. Müller glial cell bodies lie in the centre of the inner nuclear layer and their processes span all three cellular layers of the retina. These specialized glial cells are believed to be involved in the protection and/or repair of retinal neurons.
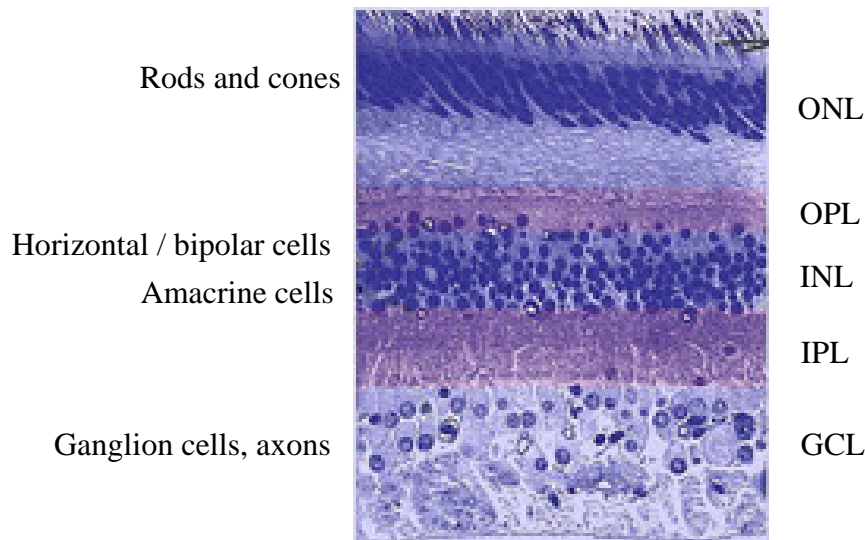
Figure 3. Vertical section through the human retina. Main layers of the retina including photoreceptors rods and cones, outer nuclear layer (ONL), outer plexiform layer (OPL), inner nuclear layer (INL), inner plexiform layer (IPL) and ganglion cell layer (GCL). The vitreous is at the bottom and the sclera is at the top. Figure taken from http://webvision.med.utah.edu/ sretina.html.

3.1.1 Photoreceptor cells

Most vertebrates have two types of photoreceptor cells, rods and cones, in the retina. Rods are highly sensitive to light and operate under dim lighting conditions while cones function in bright light and are less sensitive and responsible for colour vision. In most vertebrates, there is only one rod but 2-4 types of cone photoreceptor cells containing different cone pigments (Baylor, 1996). Only 5% of photoreceptor cells are cones and 95% are rods (Rattner et al., 1999). In primates, cones are found throughout the retina but are most concentrated within a small central region, the fovea (Fig. 1). A somewhat larger zone, called

the macula (Fig. 1), is centered in the fovea and also includes the immediately surrounding rod-rich retina. Both photoreceptors are highly specialized cells consisting of several morphologically and functionally distinct regions. The photoreceptor outer segment (Fig. 4), located adjacent to the retinal pigment epithelium (RPE) cell layer, is a specialized compartment uniquely designed for light reception and phototransduction. The outer segment is filled with flattened membrane sacs called discs and is constantly renewed by new synthesis and assembly at its base, and by shedding of older material from its tip. A thin, nonmotile cilium links the outer segment to the inner segment, a cellular compartment where subcellular organelles are located. This region further extends to the synaptic region where the electrical signal generated in the photoreceptor cells is transmitted to other neurons of the retina.
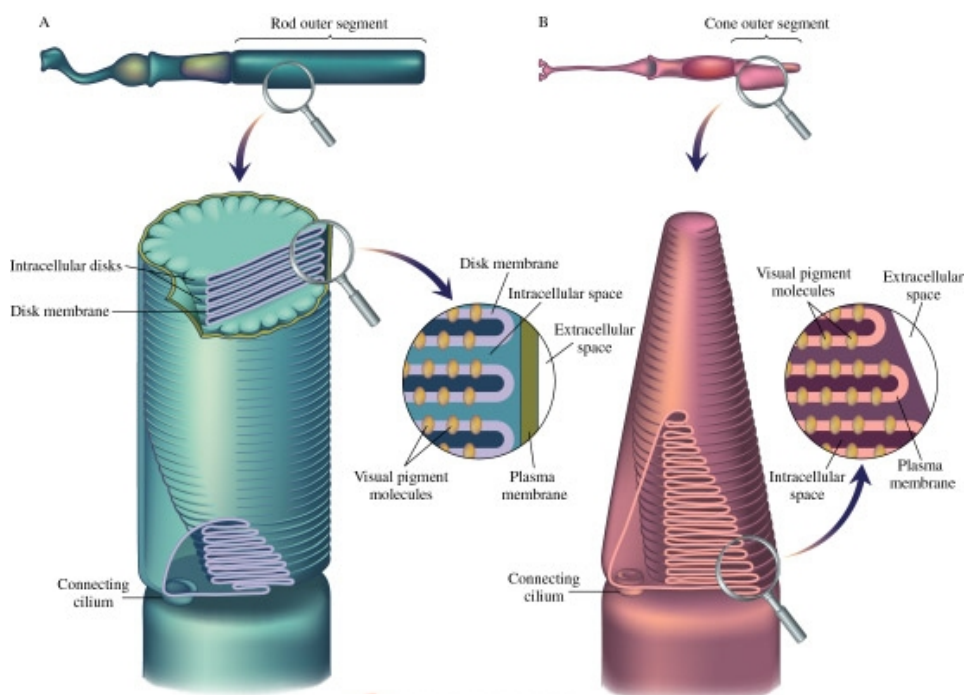
Figure 4. Structure of a rod and cone photoreceptor outer segment. A. In the rod cells, the outer segment appears as a cylindrical structure. **B.** The cone outer segment is generally shorter and often is tapered or conical. The outer segment of both photoreceptors consists of hundreds of flattened disc membranes assembled in an ordered axial array.

3.1.2 Phototransduction cascade

In response to light, a coordinated series of molecular events, the so-called phototransduction cascade, is triggered in photoreceptor cells. Photons excite pigment-containing proteins called rhodopsins, which then switch on the protein transducin by loading it with the small molecule guanosine triphosphate (GTP). When bound to GTP, transducin turns on a phosphodiesterase, an enzyme that breaks down cyclic guanosine monophosphate (cGMP). High concentrations of cGMP open specialized ion channels in the outer cell membrane. Thus, by reducing the concentration of cGMP, light changes the flow of ions across the membrane of photoreceptive neurons, resulting in hyperpolarization of the rods producing an electrical signal that is necessary for communicating with the brain (Arshavsky et al., 2002). In darkness Ca^{2+} and Na^+ enter the cell through the open channel, partially depolarizing the membrane. Once the concentration of cGMP decrease, membrane is hyperpolyrized by closing the channel and Na^+ is extruded by a pump in the inner segment, while Ca^{2+} is extruded by an exchanger who is driven by entry of Na^+ and efflux of K^+. Continued operation of the exchanger at the onset of light produces a fall in the intracellular concentration of Ca^{2+} (for reviews see Burns and Baylor, 2001; Arshavsky et al., 2002).

The cationic channel in the surface membrane is held open in darkness by the binding of cGMP, which is synthesized from the precursor GTP by guanylate cyclase. Once light-activated switch is on, the cells turn it off by limiting the amount of time that GTP-bound transducin can keep the phosphodiesterase enzyme active. Transducin can accomplish this task by converting - hydrolysing its bound GTP molecule into guanosine diphosphate, GDP (Burns and Baylor, 2001; Arshavsky et al., 2002; Blumer, 2004). Because transducin-bound GDP has a low affinity for phosphodiesterase, it releases the enzyme in an inactive form, allowing cGMP levels to rise again and return the flow of ions across the cell membrane to the 'dark' state. In this molecular cascade, then, the conversion of GTP to GDP by transducin is rate-limiting and defines the amount of time for which a photoreceptor responds to a light pulse.

Figure 5. Phototransduction in photoreceptor cells (see next page). In the rod class of photoreceptors, the pigment containing protein rhodopsin absorbs light (a) and activates transducin (b) by causing it to release GDP and bind GTP. GTP-bound transducin binds to and activates a phosphodiesterase (PDE), which converts cGMP to GMP (c). The concentration of cGMP decreases below what is required to open cGMP-gated ion channels, reducing the flow of ions across the cellular membrane. RGS9 bound to R9AP turns off the light-induced response by accelerating the rate of GTP hydrolysis by transducin, releasing phosphate, P(d). (picture taken from Blumer, 2004).
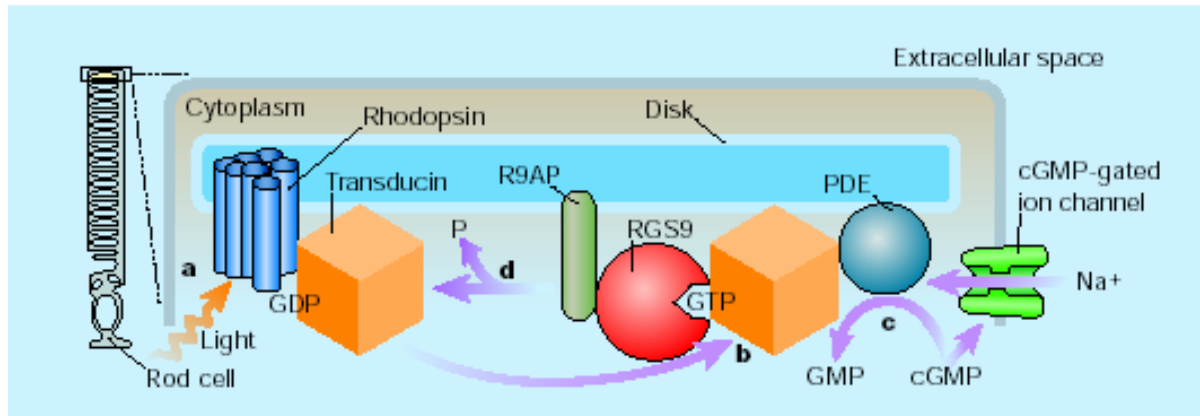
All vertebrate visual pigments consist of an apoprotein opsin, linked covalently to a chromophore 11-*cis* retinal (an aldehyde derivative of vitamin A). The opsin is different in different photopigments and its particular interaction with the retinal determines the wavelength that is absorbed. As a photon of light hits a molecule of vision pigment, the 11-*cis*-retinal chromophore undergoes isomerization to all-*trans*-retinal, after which all-*trans* retinal dissociates from opsin, to be replaced by a new molecule of 11-*cis* retinal (Fig. 6) (Arshavsky et al., 2002). This conversion, called the retinoid (or visual) cycle requires participation of the retinal pigment epithelium, RPE. In this cycle, the all-*trans* retinal chromophore dissociating from opsin is reduced to all-*trans* retinol (the corresponding alcohol derivative of vitamin A) within photoreceptors. All-*trans* retinol is then released into the extracellular space, where it is absorbed by cells of the pigment epithelium. There, all-*trans*-retinol is esterified to form all-*trans*-retinyl ester by the transfer of a fatty acid from phosphatidylcholine. The next reaction, perhaps the most important for the whole cycle, couples hydrolysis of the retinyl ester with all-*trans* to 11-*cis* isomerization of retinol. Finally, 11-*cis*-retinol is oxidized to 11-*cis*-retinal, which is released from the RPE and taken up by photoreceptors to regenerate opsin. An alternative retinoid cycle based on the interaction between Müller cells and cones is responsible for the regeneration of the cone opsins (Mata et al., 2002).

The flux through the visual cycle is extremely high. Looking at the blue sky on a sunny day produces 20,000 photoisomerizations per rod per second. Under these viewing conditions, the visual cycle in each RPE cell therefore processes approximately 1,000,000 chromophores per second (Rattner et al., 1999).
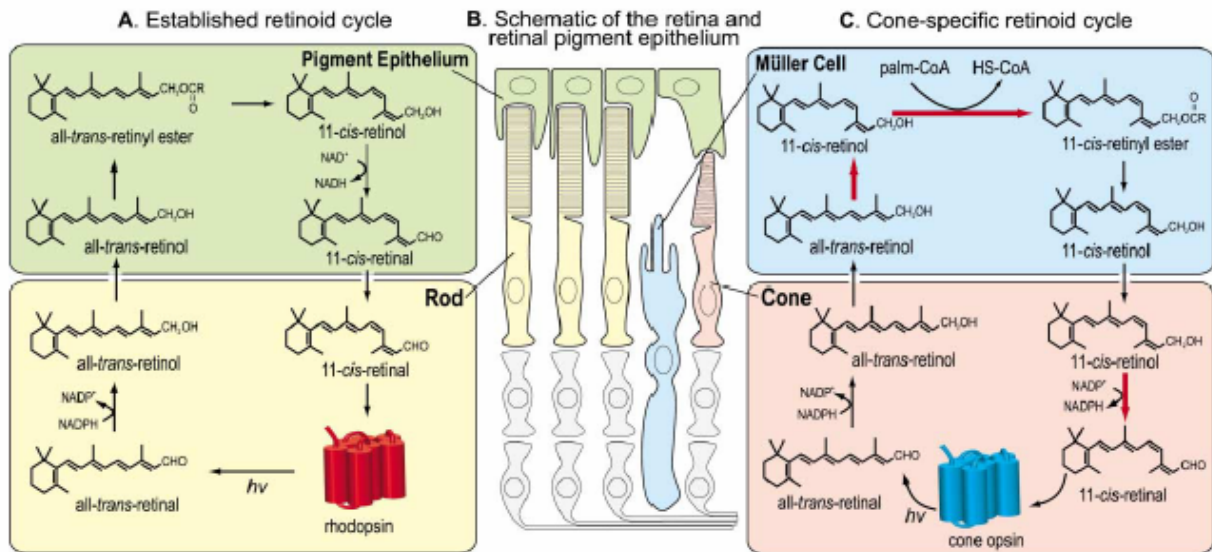
Figure 6. Two Retinoid Cycles in Vertebrate Retina (figure taken from Arshavsky et al., 2002)

The retina as highly complex system might be greatly susceptible to genetic defects leading to a wide range of retinal disease phenotypes. Hereditary eye diseases that cause the photoreceptor cells of the retina to degenerate, thus resulting in partial or total blindness, are affecting 1 in 3.000 people (Sullivan and Daiger, 1996). Hereditary retinal diseases are characterized by age of onset, severity and topographic pattern of visual loss, family history and ophthalmoscopic findings (Rattner et al., 1999). They can be caused by defects in single gene (monogenic), two genes (digenic) or they can be even more complex (Sullivan and Daiger, 1996). The monogenic forms can be autosomal dominant, autosomal recessive or X-linked, whereas complex or multifactorial forms are caused by defects in multiple genes or by combination of genetic and environmental factors (Sullivan and Daiger, 1996).

3.2   Age related macular degeneration

3.2.1   Disease phenotype

Age-related macular degeneration (AMD) is an example of the complex disease characterized by a progressive degeneration of the photoreceptors, the underlying retinal pigment epithelium, Bruch's membrane and possibly the choriocapillaries in the macular region (Fine et al., 2000; Hageman et al., 2001; Lutty et al., 1999). In industrialized countries,

AMD is a major cause of visual impairment in people aged 65 and older (Stöhr et al., 2000). Clinical and pathological features of AMD include drusen formation, hypo- and/or hyperpigmentation, loss of photoreceptors, choroidal neovascularization, subretinal fibrous and/or neovascular tissue in the macula, and alterations of Bruch's membrane and RPE cells (Tuo et al., 2004).

In the normal eye throughout life there is a regular turnover of the photoreceptor outer segments, which contain the photosensitive visual pigments. Material shed from the outer segments is engulfed and digested by the underlying retinal pigment epithelium and ultimately cleared by the capillaries of the choroid (the choriocapillaris). The RPE is separated from the choriocapillaris by a multilayered structure called Bruch membrane. It is known that this membrane thickens with age and that in the early stages of AMD there is accumulation of acellular debris in the membrane which contributes to the development of drusen, one of the characteristic clinical features of early AMD, and pigmentary abnormalities in the macular region (John et al., 2000). It is likely that this debris is derived from the cytoplasmic material being continually discharged into the inner portion of Bruch membrane from the RPE. Although small drusen are commonly observed in the aging macula, the presence of large drusen is considered diagnostic for AMD (Bird et al. 1995). Late stages of AMD are typically grouped in two forms: (1) a "dry" form associated with the loss of RPE and photoreceptors and diminished retinal function, known as "geographic atrophy" (GA); and (2) a "wet" exudative form, which is associated with the formation of choroidal neovascularization (CNV) (Fig. 7). The new vessels are fragile and often bleed leading to the formation of fibrovascular scars and irreversible visual loss. CNV is responsible for the majority of cases of severe loss of central vision. These alterations result in the loss of central visual activity (Ambati et al., 2003a). Although GA and CNV together account for only 10% of patients with AMD (Smith et al. 2001), they are responsible for most cases of associated blindness.
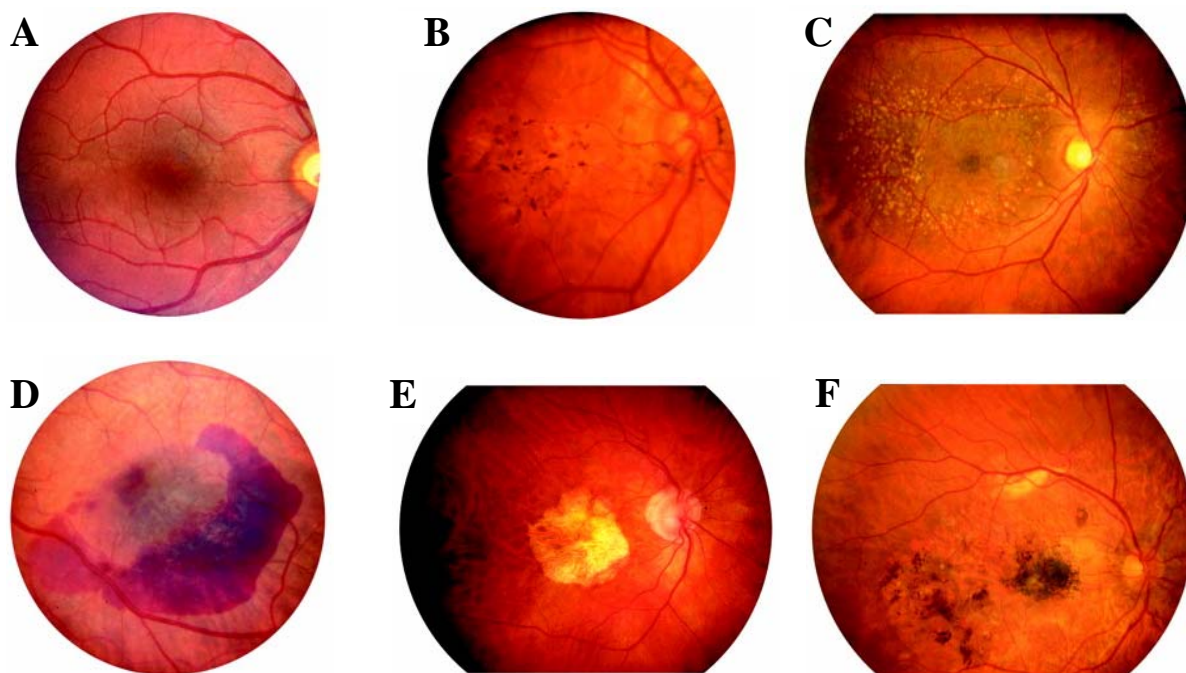
Figure 7. Range of ophthalmoscopic phenotypes consistent with the diagnosis of AMD. A. Normal fundus (for comparison); **B.** RPE hyperpigmentation; **C.** large and small drusen confined to the macula; **D.** subretinal hemorrhage secondary to a choroidal neovascular membrane; **E.** geographic atrophy; **F.** RPE hyperpigmentation and subretinal fibrosis secondary to a choroidal neovascular membrane (figure taken from Stone et al., 2001).

Laser photocoagulation targeting the choroidal feeding vessels, photodynamic therapy, surgical macular translocation, and antiangiogenesis agents are reported to reduce the risk of moderate to severe vision loss in some patients with the neovascular, or wet, form of the disease, but so far there is no proven treatment of advanced AMD to date (Ciardella et al., 2002; Fujii et al., 2003; Lim, 2002; McBee et al., 2003; Rechtman et al., 2002).

3.2.2 Risk factors in AMD

There are external, individual risk factors that contribute to AMD. For example, a number of lifestyle factors may influence disease manifestations (Evans, 2001). Recognized risk factors include smoking and diet (Husain et al., 2002; Hyman and Neborsky, 2002; Seddon et al., 2001). Oxidative stress may promote neovascularization via deposition of extracellular matrix along Bruch's membrane and increased levels of angiogenic factors in RPE cells (Mousa et al., 1999). Different opinions are held concerning the association

between AMD and hypertension (Delcourt et al., 2001; Hyman and Neborsky, 2002; Snow and Seddon, 1999). Among all potential environmental factors identified, the strongest risk factors are ageing, diet and smoking. Many investigators have speculated that dietary antioxidants such as carotenoids, zinc, and mineral cofactors confer protection against AMD but the evidence is not conclusive and further studies are needed (Frank, 1998; Frank et al., 1999).

Pathogenesis in AMD involves environmental factors as well as varying susceptibilities to these external factors based upon different genetic backgrounds (Guymer, 2001). Ethnicity, for example, is a component of genetic background. AMD is more prevalent in people with less pigmentation, although little difference in the occurrences of smaller-sized drusen in individuals with light and dark skin has been reported (Friedman et al., 1999). A possible explanation for this observed trend suggests an enhanced protective role for the melanin in RPE cells and choroidal melanocytes. There has been compelling evidence that heredity plays a role in AMD (Yoshida et al., 2000). The tendency for familial aggregation in AMD cases further supports a genetic context in disease development. Approximately 20% of AMD patients have a positive family history (de Jong et al., 1997; Klaver et al., 1998b; Meyers et al., 1995). Furthermore, there is higher prevalence of AMD among monozygotic twins as compared to their spouses and first-degree relatives Gottfredsdottir et al., 1999; Klaver et al., 1998b).

3.2.3 The genetics of age-related macular degeneration

There are several limitations in causative gene mapping by linkage studies in late-onset diseases such as AMD. First, many individuals in the affected families are young and have not yet developed the disease. Second, it is likely that many older family members have passed away making informative pedigree situations rarely available; and third, the heterogeneity of AMD requires the estimation of multiple genetic risks as this disease is thought to result from polygenic interactions rather than from single gene effects (Tuo et al., 2004). Furthermore, the variable presentation of AMD may lead to problems with the classification of affected individuals (Seddon et al., 2003). For these reasons, the study of hereditary monogenic retinal dystrophies that share clinical and pathological similarities with AMD may offer some insightful clues as to identify the AMD-related genes.

Several hereditary retinopathies have overlapping phenotypes with AMD, but the extent to which they are related to AMD remains unclear (Gorin et al., 1999). For example, similar to AMD Stargardt disease and Best vitelliform macular dystrophies commonly result in atrophy of the central retina. Drusen are a prominent feature of Doyne honeycomb retinal dystrophy (also known as autosomal dominant radial drusen, Malattia Leventinese) which is followed at later stages by the development of CNV. The latter feature is also shared by Sorsby fundus dystrophy, a rare late onset macular dystrophy. The genes responsible for these conditions are attractive candidates for AMD susceptibility genes.

VMD2 has been identified as the causal gene of dominant juvenile onset vitelliform macular dystrophy, commonly known as Best disease. Best disease affects the central retina and is characterized clinically by the classical abnormal electrooculogram and the presence of round or oval yellow subretinal macular deposits (Fig. 8A). The yellow material gradually reabsorbs over time, leaving an area of RPE atrophy and often subretinal fibrosis (Kramer et al., 2003; Michaelides et al., 2003; Petrukhin et al., 1998; Weber et al., 1994b). Affected individuals carry mutations in the VMD2 gene (Kramer et al., 2003) which has been localized to 11q13.2–13.3 (Stone et al. 1992, Weber et al., 1994b). The gene encodes a 585-amino-acid protein known as bestrophin, which is selectively expressed in the RPE of the eye (Marquardt et al., 1998; Petrukhin et al., 1998).

Sorsby fundus dystrophy (SFD) is an autosomal dominant disease clinically characterized by choroidal neovascularization and irregularities in the extracellular matrix of Bruch's membrane (Fig. 8 B and C). SFD leads to an onset of night blindness in the third or fourth decade of life and a loss of central vision from macular atrophy by the fifth (Gregory-Evans, 2000; Weber et al., 1994a). The SFD locus was mapped to 22q13-qter, containing the gene for tissue inhibitor of matrix metalloproteinases-3 (TIMP3). TIMP3 is known to play a pivotal role in the regulation of proteases such as the matrix metalloproteinases including the collagenases and gelatinases. Recently, TIMP3 was also shown to inhibit several members of the ADAM (a disintegrin and a metalloprotein domain) family. TIMP3 thereby greatly determines the extent of matrix degradation/regeneration during normal tissue remodeling and regulates the homeostasis of the extracellular matrix (Jomary et al., 1995). Reported point mutations in TIMP3 (Weber et al., 1994a; Felbor et al., 1995; Carrero-Valenzuela et al., 1996; Felbor et al., 1996; Felbor et al., 1997; Tabata et al., 1998; Langton et al. 2000; Jacobson et al., 2002) are predicted to disrupt the tertiary structure and consequently the functional properties of the mature protein.

Stargardt disease (STGD1) is the most frequent of the juvenile macular dystrophies. It is characterised by early onset and leads to progressive atrophy of the macular RPE and overlying photoreceptors (Fig. 8D). The condition is caused by mutations in the ABCA4 gene located in 1p21, a member of the ATP binding cassette (ABC) transporter superfamily (Allikmets et al., 1997). PROML1 in 4p15 has been identified as another causative gene for STGD (Maw et al., 2000) (autosomal dominant or recessive; you need to specify). The gene belongs to the prominin family of five-transmembrane domain proteins.

The disruption of EFEMP1 (EGF containing fibrillin-like extracellular matrix protein 1) at 2p16-21, which is encoding a protein homologous to a family of extracellular matrix glycoproteins known as fibulins, has been linked to Doyne honeycomb retinal dystrophy (DHRD), a dominant macular dystrophy. This disorder is characterised by a confluent drusen accumulation beneath the RPE (Fig. 8E) in early adult life (Stone et al., 1999).

Disruptions in the peripherin/RDS gene (Lith-Verhoeven et al., 2003a) on chromosome 6p21 have been identified in patients with pattern dystrophies characterized by a bilateral accumulation of lipofuscin-like material at the level of the RPE (Nichols et al., 1993a, b) (Fig. 8F). The peripherin/RDS protein is a membrane-associated glycoprotein restricted to photoreceptor outer-segment discs in a complex with ROM1 (Rod outer segment membrane protein 1). This protein may function as an adhesion molecule involved in the stabilization and maintenance of a compact arrangement of outer-segment discs (Travis et al., 1991). Peripherin has also been shown to interact with the GARP domain (glutamic acid- and proline-rich region) of the beta-subunit of rod cGMP-gated channels in a complex including the Na/Ca–K exchanger (Poetsch et al., 2001).

CTRP5 has recently been reported as the causal gene of late-onset retinal degeneration (L-ORD), an autosomal dominant disorder the retina (Hayward et al., 2003). Clinically, L-ORD is characterized by onset in the fifth to sixth decade of life with night blindness as the main clinical feature. Pathological features include thick punctuate yellow-white deposit between the basal lamina of the RPE and Bruch's membrane, progression to severe central and peripheral degeneration, and choroidal neovascularization and chorioretinal atrophy (Kuntz et al., 1996; Milam et al., 2000). CTRP5 is secreted by the RPE and is a potential constituent of Bruch's membrane. This protein most likely functions in the facilitation of basal RPE adhesion to Bruch's membrane by forming an extracellular hexagonal lattice.
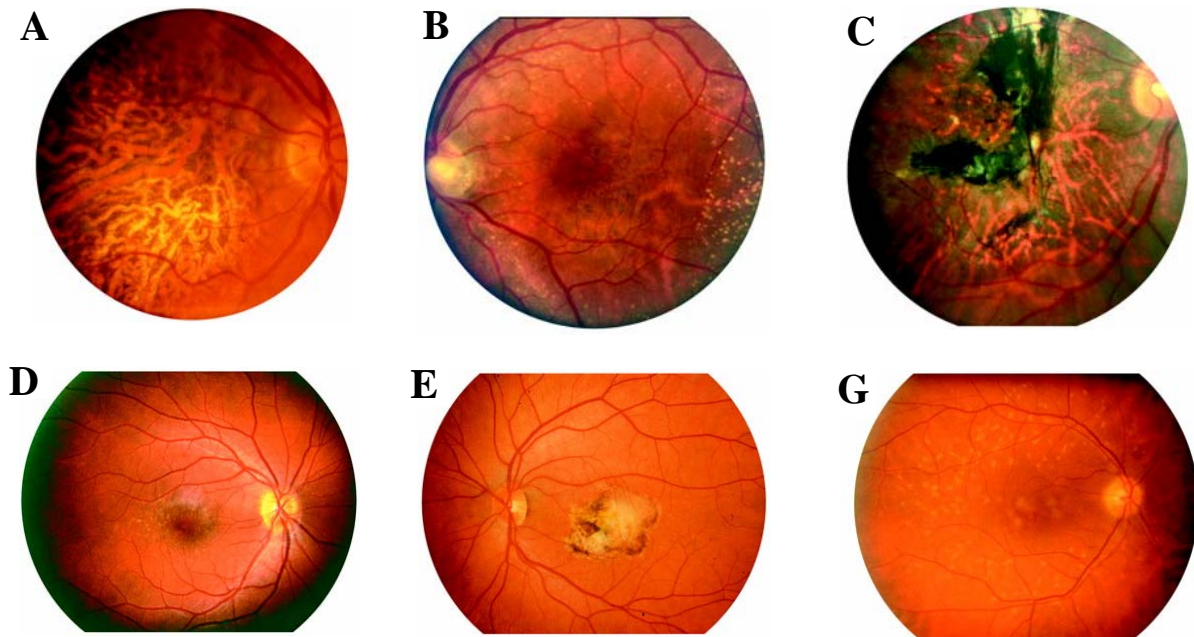
Figure 8. Ophthalmoscopic phenotypes of patients affected with molecularly confirmed early-onset heritable macular dystrophies. A. Best disease; **B.** Sorsby fundus dystrophy; **C.** Sorsby fundus dystrophy; **D.** Stargardt disease; **E.** Malattia Leventinese; **F.** *RDS*-associated pattern dystrophy (figure taken from Stone et al., 2001).

3.2.4 Search for the complex disease genes

Age-related macular degeneration has been recognized as a complex genetic disorder in which multiple genes may contribute to an individual's susceptibility for developing the condition. These factors are causing disease only when a threshold of susceptibility is reached and most of the susceptibility alleles are neither necessary nor sufficient to cause disease, but only confer a slight increase in risk (Johnson and Todd, 2000). Methods, based on associations between marker and disease, provide powerful tool for localization of disease loci. The single nucleotide polymorphisms (SNPs) became "markers of choice" for disease mapping (Johnson and Todd., 2000), allowing analyses for association between a phenotype and a functional variant, directly. Genetic association studies with a large sample size analyse weather a SNP is enriched in affected cases compared to unaffected controls from the same population (Taylor et al., 2001).

The SNPs are highly abundant, stable genetic markers in the human genome, and are estimated to occur at 1 out of every 1,000 bases (Sachidanandam et al., 2001; Venter et al., 2001). Depending on where a SNP occurs, it might have different consequences at the

phenotypic level. SNPs in the coding region of a gene may alter the function or structure of the encoded protein by amino acid substitution. Such a functional SNP is a necessary and sufficient cause of most of the known recessively or dominantly inherited monogenic disorders (Syvänen, 2001). These SNPs are routinely analysed for diagnostic purposes. However, restricting the search to cDNA sequences may fail to detect a predisposing variant located in a non-coding regulatory region of the gene, which may affect its expression (Johnson and Todd, 2000).

The reason for the current interest in large-scale SNP characterisation is the hope that they could be used as markers to identify genes that predispose individuals to common, multifactorial disorders by using linkage disequilibrium (LD) mapping (Risch & Merikangas, 1996; Schork et al., 1998). For example, such studies analyse SNP alleles in population-based analyses to identify loci that are associated with a particular disease or phenotype. With the advent of new molecular tools, in particular high-density SNP-typing, it has been argued that, while linkage analysis may retain some role, genome-wide association studies with SNPs offer a superior strategy for unravelling genetic complexity (Baron, 2001).

3.2.4.1 Genotyping single nucleotide polymorphisms (SNPs)

As a result of the efforts of the SNP Consortium - a collaboration of 14 major pharmaceutical companies and the Wellcome Trust, as well as members of the Human Genome Project (Sachidanandam et al., 2001) there were almost 2 million SNPs in public databases by the end of year 2001, and perhaps twice that number of SNPs in commercial databases, such as that of Celera Genomics (Venter et al., 2001). Genotyping a collection of SNPs that occur randomly distributed throughout the whole genome would be a preferred approach to detect genomic regions in which the frequencies of the SNP alleles may differ between patients and controls. Due to inefficient SNP genotyping capabilities and the immense costs involved, such endeavours are still infeasible at present but may be in reach in the not too far future.

Following the completion of the genetic, physical, and SNP maps, the next step is the generation of "haplotype" maps of the human genome. Such "haplotype" maps consist of a high density of SNPs defining the relatively small number of ancestral haplotypes (blocks of tightly correlated genetic variants) in each region of the human genome. Knowledge of these haplotypes will allow comprehensive and efficient testing of the association of human genes

with disease. Currently, the SNPs are now being verified experimentally and their distribution between and within populations are being assessed (Marth et al., 2001; Katsanis et al., 2001). The hope that SNPs will allow to pinpoint genes that underlie complex disease, is the driving force for establishing the technology for large-scale analysis of SNPs.

In summary, from the available data it appears that AMD is caused by environmental factors triggering disease in genetically susceptible subjects. Identification of AMD's genetic causes will allow characterizations of the disease phenotypes and better understanding of the pathogenesis of this disease. If a genetic predisposition is known at an early age, behavioural modifications may prevent disease or decrease disease severity later in life. Current evidence shows that it is unlikely that a single gene variant is responsible for the development of AMD, at least not in the common forms of the disease. Perhaps focusing on the dynamic interaction between multiple genes, gene-environmental components and their comprehensive effects will provide new insights in the search for the genetic risks, pathogenesis and treatment of AMD.

3.3 Goal of the thesis

The highly specialized function of the retina requires a large number of specifically expressed genes involved in performance and control of phototransduction process. These genes might be highly susceptible to the genetic defects. Therefore, isolation and characterization of novel retinal genes should further our understanding of the tissue structure and physiology, should provide novel candidates for the disease genes and contribute to understanding the pathogenesis of the retinal diseases.

The goal of this project was identification of the novel genes that are specifically or predominantly expressed in the human retina. For this purpose cDNA library enriched for differentially expressed retina specific transcripts of both high and low abundance was generated using suppression subtractive hybridization (SSH) technique (Diatchenko et al., 1996).

At the beginning of this project 1113 clones from the retina SSH cDNA library were sequenced and obtained sequences were subjected to homology searches against public databases. Expression profiling of the clones that derived from the transcripts which sequence didn't show homology to known human genes were performed by RT-PCR, standard and virtual Northern blot and real-time PCR. In this way, the transcripts with abundant or

exclusive expression in the human retina were identified and selected for additional analysis. Further efforts were directed towards identification of the full-length cDNA sequence of the several genes of interest (for this purpose full-length retinal cDNA library was constructed and screened for the homologue clones) and their functional characterization.

Identified genes with retina restricted expression could be candidates for the retina disease genes. Therefore, a few of them were subjected to the single nucleotide polymorphism (SNP) analysis and generated SNP maps will be used for the future association studies of the AMD as complex retinal disorder.

4 MATERIALS AND METHODS

4.1 Bioinformatics

Computational analyses of the nucleotide and protein sequences investigated in this doctoral thesis were performed using a number of bioinformatics tools.

4.1.1 Computational analysis of nucleotide sequences

Nucleotide sequence homology searches in public databases were performed using the BLASTN program at National Center for Biotechnology Information - NCBI (http://www.ncbi.nlm.nih.gov/), whereas BLAT searches were performed at the Human Genome Browser at the University of California, Santa Cruz (USCS Genome Browser at http://genome.ucsc.edu/). Repeats in the nucleotide sequences were identified using the Repeat Masker Web Server (http://repeatmasker.genome.washington.edu/). Putative ORFs within cDNA sequences were predicted using ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). ClustalW program, provided by the BCM Search Launcher (http://www.ebi.ac.uk/clustalw/) was used to generate multiple alignments of nucleotide sequences and visualization of the sequence alignments was done using BOXSHADE algorithms (http://www.ch.embnet.org/software/BOX_form.html) which shade the aligned sequences.

Information about partial and full-length cDNA sequences investigated in the course of this project was acquired from GeneCards (http://bioinfo.weizmann.ac.il/cards/index.shtml) and LocusLink (http://www.ncbi.nlm.nih.gov/projects/LocusLink/).

4.1.2 Computational analysis of amino acid sequences

The BLASTP program, available at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/), was used for amino acid sequence homology searches in public databases. The analyses of putative protein sequences for structural and functional motifs were carried out using different tools available at ExPASy Web server (http://www.expasy.org/) including TMHMM (http://www.cbs.dtu.dk/services/TMHMM) and SWISS-PROT (http://www.expasy.org/sprot/).

4.2 RNA Methods

4.2.1 RNA sources

Human retina and RPE were dissected from eyes obtained from the University Eye Clinic in Würzburg (eyes were donated for the purpose of cornea transplantion, post mortem, approval of local ethics committee were given) and were used for isolation of RNA. Commercially obtained mice (Harlan & Winkelmann, Borchen, Germany) were used as sources of murine total retina and eye RNA. After dissection of the eyes the isolated tissues were immediately subjected to RNA extraction or were frozen in liquid nitrogen and stored at -80°C. Total RNA from human brain, fetal brain, spinal cord, bone marrow, retina, heart, kidney, liver, fetal liver, lung, trachea, spleen, thymus, skeletal muscle, colon, testis, uterus, placenta, adrenal gland, salivary gland and bladder was purchased from Clontech (Heidelberg, Germany) and Ambion (Austin, TX, USA) and stored at -80°C.

4.2.2 RNA isolation

4.2.2.1 Total RNA

Total RNA from fresh / frozen tissue or cell culture was isolated with the Rneasy® Mini Kit (Qiagen, Hilden, Germany). Samples were first lysed and homogenized in the presence of the denaturing guanidine isothiocyanate (GITC) – containing buffer, which immediately inactivates RNases to ensure isolation of intact RNA. Homogenization was done on ice while subsequent steps were performed at room temperature. Briefly, 75% ethanol was added to the lysate to provide appropriate binding conditions and the sample was loaded on an Rneasy® mini column where the total RNA binds to the silica-gel based membrane. Subsequent washing steps were performed according to the manufacturer's protocol. Total RNA was eluted from the column with 40 µl RNase free water and stored at -80°C.

The quality and quantity of the purified RNA was analysed by loading 3 µl aliquot on a 1.2 % agarose-formaldehyde gel. Intact eukaryotic RNAs reveal a sharp 28S band that is approximately twice as intense as the 18S rRNA band. Partially degraded RNA is characterized by a smear, lacks the sharp rRNA bands, or is not exhibiting the 2:1 ratio whereas completely degraded RNA will appear as a very low molecular weight smear.

4.2.2.2 Cytoplasmic RNA

Cytoplasmic RNA was extracted from 30 mg of fresh tissue. In order to keep nuclear membrane intact a teflon homogenizer was used (Das et al., 2001). Cell membranes were lysed with a precooled lysis buffer (50mM Tris-Cl, pH 8.0; 140mM NaCl; 1.5mM $MgCl_2$; 0.5% nonident; 1000U/ml RNase inhibitor and 1mM DTT). Cell lysis was done on ice and following centrifugation, for pelleting the nuclei at 6.000g for 2 minutes, at 4°C. In order to pellet the intact nuclei the supernatant was then mixed with precooled 75% ethanol and all subsequent steps were performed at room temperature with Rneasy® Mini Kit (Qiagen, Hilden, Germany), following the manufacturers instructions. RNA was eluted from the column with 40µl RNase free water and stored at -80°C. The purity of the cytoplasmic RNA was tested by PCR amplification of the XIST gene using XISTF and XISTR primers (Table 19, Appendix).

4.2.2.3 Poly(A)$^+$ RNA

Poly(A)$^+$ RNA was purified using the Oligotex^{TM} Kit (Qiagen, Hilden, Germany). Following this procedure poly(A)$^+$ containing RNA is hybridized to a dT oligomer which is covalently linked to the solid phase matrix of the provided polystyrene-latex Oligotex particles. Total RNA diluted in a volume of 100µl was mixed with 10µl Oligotex suspension preheated at 37°C. Afterwards, the secondary structure of the RNA was disrupted by heating the sample to 70°C for 3 minutes. An incubation step at room temperature for 10 minutes was followed by centrifugation at 14.000 rpm for 2 minutes. In order to obtain high-salt concentration required for hybridization, the Oligotex/RNA pellet was resuspended in 400µl of the provided binding buffer and the sample was transferred to the spin column. Subsequently, poly(A)$^+$ mRNA was released by lowering the ionic strength and destabilizing the dT:A hybrids with 30µl eluting buffer which was preheated on 75°C.

4.2.3   Quantification of RNA

The concentration of RNA was determined by spectrophotometer measurement of the absorbance at 260 nm (A_{260}). An absorbance of 1 unit at 260 nm corresponds to 40 µg of

RNA per ml of water. Therefore, 100 μl of RNA was diluted in RNase free water (1:50 dilution) and the concentration of the sample was calculated according to the formula:

Concentration (μg/μl) = 40μg X A_{260} X dilution factor

Total yield (μg)   = concentration X volume of sample (μl)

4.2.4 First-strand cDNA synthesis

First-strand cDNA synthesis was carried out with reverse transcriptase Superscript II^{TM} (Invitrogen, Karlsruhe, Germany) under RNase free conditions. In the first step, 1-2μg poly(A)$^{+}$ RNA were mixed with 10 pmol 3'-RACE AP primer (Table 19, Appendix), 1μl dNTPs (10mM of dATP, dTTP, dGTP, dCTP) and RNase free water. RNA was denatured at 70°C for 10 min and the mixture was immediately transferred to ice. In the second step, 4μl of 5 X First strand buffer (250mM Tris-HCl, pH 8.3; 30 mM $MgCl_2$; 375mM KCl), 2 μl of 0.1M DTT and 1μl of SuperscriptII^{TM} (200 U/μl) were added. The reverse transcription was done at 42°C for 50 min in 20 μl final volume. Afterwards the mixture was incubated at 70°C for 15 min, in order to inactivate the enzyme. Synthesized cDNA was stored at -20°C. PCR amplification of the GUSB and G3PDH genes was performed using GUS B3/ GUS B5 or G3PDH F/ G3PDH R primers in order to test the quality of synthesized cDNAs (Table 19, Appendix).

4.2.5 Northern blot
4.2.5.1 Agarose gel electrophoresis

An aliquot of 10 μl of total RNA was electrophoretically separated on an 1.2% denaturing agarose gel. The gel was prepared by dissolving 1.2 g agarose in 87 ml DEPC treated water, cooling down to 60°C and adding 10 ml 10 X MOPS running buffer (0.2M MOPS, 50mM NaOAc, 10mM Na_2EDTA – pH 7.0) and 3ml 37% formaldehyde. Previous to loading, RNA was denatured at 65°C for 10 min in the presence of 1 vol of RNA loading buffer (1 X MOPS, 50% deionised formamide, 18.5% formaldehyde, 0.04% bromphenol blue, 10 μg ethidiumbromide, 5% glycerin, 0.1 mM EDTA). An RNA size marker (Promega, Mannheim, Germany) was run in parallel with RNA samples at 60-80V for 3-4 hrs. Gel was

examined on an UV transilluminator to visualize the RNA and photographed with a ruler so that the size of the transcript could be determined after blotting and hybridization.

4.2.5.2 Transfer to Nylon membrane

Total RNA was transferred to the Nylon Hybond N$^+$ membrane with the VacuGene^{TM} (Amersham Bioscience, Freiburg, Germany). Prior to vacuum blotting, the gel was rinsed for 10 min in 20 X SSC (3M NaCl, 0.3M NaCitrate, pH 7.0) and the nylon membrane was moistured with 2 X SSC. Vacuum transfer was performed at -60 mbar for 3-4 hrs and gel was kept permanently covered with 20 X SSC. Afterwards, RNA was immobilized by UV crosslinking and the membrane was stored between two sheets of Whattman paper at -20°C.

4.2.5.3 Radioactive random-prime probe labelling

DNA fragments were labelled by random-priming with [α^{32}P] dCTP (Hartmann, Braunschweig, Germany). Probes were generated by PCR amplification, purified from agarose gel by using Nucleospin$^®$ Extract Kit (Macherey-Nagel, Düren, Germany) and 20-50 ng were denatured at 100°C for 5 min in 50 μl final volume. Subsequently, 10μl 5 X OLB-buffer (250mM Tris-HCl, pH 8.0; 25mM MgCl$_2$; 50mM β-Mercaptoethanol; 2mM of each dATP, dTTP, dGTP; 1M Hepes, pH 6.6; 50 U (A$_{260}$) pd(N)6), 2 mg BSA, 4U Klenow fragment and 3μg [α^{32}P] dCTP (3000 Ci/mmol) were added and the mixture was incubated at 37°C for 3 hrs or overnight at RT. The radioactive labelled DNA probe was then separated from unincorporated [α^{32}P] dCTP using a G-25 Sephadex column (Amersham Biosciences, Freiburg, Germany) and denatured for 5 min at 100°C prior to hybridization.

4.2.5.4 Membrane hybridization

The RNA containing membrane was prepared for hybridization in 25 ml prewarmed Church buffer (0.5 M Na$_2$HPO$_4$, pH 7.2; 1 mM EDTA, pH 8.0; 1% BSA; 7% SDS) by rotating at 60°C for 1-2 hrs. The purified and denatured radioactive DNA probe was then added to the buffer and incubated overnight at 60°C, with rotation. Afterwards the incubation

solution was removed and the filter was washed once with SSPE/0.1% SDS buffer (20 X SSPE: 3 M NaCl, 200mM NaH_2PO_4, 20 mM Na_2EDTA) at 60°C. The increasing stringency of the washing solution was obtained by decreasing concentration of the SSPE starting from 2 X SSPE to 0.5 X SSPE. Membrane was rinsed by shaking in 50 ml prewarmed SSPE/SDS buffer at 60°C and background radiation was checked with Geiger counter in between. The membrane was then exposed to an X-ray film (RETINA, Fotochemische Werke, Berlin, Germany) at -80°C or to a Storage Phosphor screen (Amerscham Bioscience, Freiburg, Germany), for a minimum of 3 days.

4.3   DNA Methods

4.3.1   Phenol / chlorophorm extraction

For genomic DNA purification, 1 volume of phenol:chlorophorm:isoamylalcohol (25:24:1) mixture was added to the aqueous sample, mixed by continuous gentle inversion for 1 to 2 min and centrifuged on 14.000 rpm for 5 min. The upper aqueous phase was transferred to a fresh tube and 1 volume chlorophorm:isoamylalcohol (24:1) was added to the DNA solution, mixed by continuous gentle inversion for 1 to 2 min and again centrifuged at 14.000 rpm for 5 min. The upper layer was mixed with 1/10 volume 3 M sodium acetate (pH 4.8), 2.5 volume 95% ethanol and 100 μg glycogen which is important for precipitation of low concentrated DNA, centrifuged at 14.000 rpm for 20 min. The pellet was then washed with 80% ethanol, air dried for 10 min and resuspended with dH_2O.

4.3.2   Polymerase chain reaction (PCR)

4.3.2.1 Standard PCR

Primer sequences were designed using the OLIGO version 2.0 tool (Rychlik and Rhoads 1989) and Primer3 software (available online at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Primers with length ranging from 18 to 22 bp and having a G/C content of 45-55% were chosen as optimal, avoiding self-complementarity within a primer and formation of primer dimers. The lyophilized oligonucleotides synthesized by

MWG Biotech (Ebersberg, Germany) and Sigma-Aldrich Chemie GmbH (Münich, Germany) were dissolved in ddH2O to a final concentration of 100 pmol/µl and stored at -20°C.

PCR optimization was done at a temperature 3°C below the primers' melting temperature (T_m) which was calculated according to the Wallace formula [(A/T) x 2 + (G/C) x 4] (McConlogue et al., 1988). Amplification of complex genomic DNA templates was done using a "touch down" PCR method in order to avoid generation of nonspecific PCR products. Specifically, annealing took place at 6°C above the calculated T_m. During the following six cycles, temperature was gradually reduced by 3°C until the calculated T_m was reached, which was followed by additional 26 cycles of amplification, on obtained T_m temperature.

Amplification of genomic DNA or cDNA was carried out on the Thermocycler T3 (Biometra, Goettingen, Germany). Template DNA was mixed together with 1 X PCR buffer (50 mM KCl; 10 mM Tris-HCl, pH 8.3; 1.0 or 1.5 mM $MgCl_2$; 0.01% gelatine), 100 µM of each dNTPs (Peqlab, Erlangen, Germany), 0.4 µM of both forward (F) and reverse (R) primer, 1 U Taq DNA polymerase and formamide (used optionally at a final concentration of 4%) in 25 µl final volume. PCR consisted of an initial denaturation step at 94°C for 5 min, followed by 33 cycles of denaturation at 94°C for 30 sec, annealing at the optimal temperature for 30 sec and an elongation step at 72°C. Elongation time was 30 sec for templates up to 1 Kb long template and for larger PCR products an additional 1 min per Kb was added. Reaction was terminated with a synthesis step at 72°C for 5 min.

4.3.2.2 Quantitative Real-time PCR (qRT-PCR)

Real-time quantitative PCR amplification reactions were carried out on a BIORAD iCycler iQ™ Detection System (Münich, Germany) using SYBR® Green I as a fluorescent reporter (Sigma-Aldrich Chemie GmbH, Münich, Germany). Quantification is achieved by measuring the increase in fluorescence at the start of the exponential phase of PCR.

Primers, specific for each transcript, were designed to amplify 75 bp to 160 bp long PCR fragments. For chosen sequences, formation of primer dimers was not predicted with NetPrimer software (http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html) (PREMIER Biosoft International, Palo Alto, USA). Either the forward or the reverse primer was designed to anneal to two flanking exons and oligonucleotides were synthetized by MWG Biotech (Ebersberg, Germany). In order to obtain maximum efficiency and specificity, PCR conditions were optimized using the gradient function of the iCycler IQ using annealing

temperature ranging from 57 to 66°C and different $MgCl_2$ concentrations (1.5 mM, 2 mM or 3 mM). Primer sequences and optimized conditions are listed in Table 22). The iCycle iQ software automatically chose threshold limits and threshold cycles (C_T value), at which the fluorescence rises above background. This was reported for all samples and used for all subsequent quantifications. In order to determine the efficiency of each PCR reaction, a standard curve was generated using serial dilutions of retina cDNA as template. The slope (log of the concentration of a dilution series of the template DNA plotted versus C_T value), automatically calculated by iCycle iQ system software and used to estimate the efficiency by the formula $E = 10^{(-1/slope)}$.

One to two μg of total RNA were used for the first strand cDNA synthesis in a 20 μl reverse transcription reaction mixture as described (see RNA methods 4.4). Template DNA was diluted 1:5, 1 X PCR buffer (50 mM KCl; 10 mM Tris-HCl, pH 8.3; 1.5, 2.0 or 3.0 mM $MgCl_2$), 100 μM of each dNTPs, 0.5 μl Taq DNA polymerase, 10 nM FITC (Bio-Rad, Münich, Germany), 0.5 X SYBR® Green I and 0.2 μM of each primer in a 25 μl final volume. The PCR consisted of an initial activation step at 95°C for 1.5 min, followed by 40 cycles of denaturation at 95°C for 30 sec, annealing for 30 sec at the optimized temperature and measurement of the fluorescence optimized temperature for 10 sec. Subsequently, a melting curve analysis was carried out. It consisted of a 1 min incubation at 95°C, followed by the reduction of the temperature to the annealing temperature and an increasing of the temperature in 0.5°C steps that were held for 10 sec. All qRT-PCR reactions were done in triplicates for each template.

Amplification of the expected PCR product was confirmed by 2% agarose gel electrophoresis (see 4.3.3). The Ct values were exported to a Microsoft Excel worksheet, and converted according to a model published by Vandesompele et al. (2002). The amount of each transcript was calculated by normalization to housekeeping genes beta-2-microglobulin (B2M), TATA box binding protein (TBP), succinate dehydrogenase complex, subunit A (SDHA) and hypoxantine phosphoribosyl-transferase I (HPRTI) (primers listed in Table 22, Appendix).

4.3.3   Agarose gel electrophoresis

To prepare a 1% agarose gel, 1 g ultra pure agarose (Roth, Karlsruhe, Germany) was dissolved in 100 ml 1 X TBE buffer (10 X TBE: 89 mM Tris-HCl; 89 mM borate acid; 2 mM

Na$_2$EDTA, pH 8.3) or 1 X TAE buffer (50 X TAE: 2 M Tris-acetate; 0.05 M EDTA, pH 8.3). After cooling to 60°C, ethidium bromide (EtBr) was added to a final concentration of 0.5 µg/ml and the mixture was poured into an appropriate gel tray. DNA samples were mixed with 6 X loading buffer (0.25% bromophenol blue, 40% sucrose) prior to loading and gel was run in 1 X TBE or TAE buffer at 100-120 V. As a reference the 1 Kb$^+$ DNA ladder (Invitrogen, Karlsruhe, Germany) was size fractionated by electrophoresis on the same gel so that fragment size could be determined by comparison. After electrophoresis, fluorescent ethidium bromide-stained samples were visualized on the UV illuminator. The gel concentration varied, depending on the fragment size being analysed. For fragments shorter than 300 bp a 2.5% gel was used whereas products longer than 3 Kb were analysed on a 0.7% gel.

4.3.4   DNA purification
4.3.4.1 Purification of PCR products

The PCR products (see DNA methods 4.3.1) that were longer that 100 bp were extracted from the agarose gel using the NucleoSpin$^®$ Extract Kit (Macherey-Nagel, Düren, Germany) whereas those that were approximately 100 bp long were purified with NucleoTrap$^®$ Kit (Macherey-Nagel, Düren, Germany), following the manufacturer's protocol. Gel slices with DNA samples were melted in dissolving buffer at 50°C. In general, in the presence of chaotropic salts contained within the buffers provided, DNA binds to the silica membrane of the NucleoSpin$^®$ column or to the specially activated silica matrix of suspension particles of the NucleoTrap$^®$ suspension. All contaminating salts and soluble macromolecular components were washed with an ethanol-containing buffer and DNA was eluted under low ionic strength conditions with 40 µl H$_2$O and stored at -20°C.

4.3.4.2 Purification of plasmid DNA

Plasmid DNA purification was carried out using the NucleoSpin® Plasmid Kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol. In general, plasmid DNA was released from the host cells with SDS / alkaline lysis buffer. Subsequently, a neutralizing buffer was added to the resulting lysate in order to obtain appropriate

conditions for adhesion of the plasmid DNA to the silica membrane of the NucleoSpin® column. SDS precipitate and cell debris was pelleted by centrifugation and the supernatant containing plasmid DNA was loaded onto the column. Salt, metabolites and soluble macromolecular cellular components were washed with an ethanolic buffer and pure plasmid DNA was eluted under low ionic strength conditions with 40 µl H_2O and stored at -20°C. Quality of purified samples was analysed by agarose gel electrophoresis (see 4.3.3).

4.3.4.3 Restriction enzyme digestion

Restriction enzymes EcoRI, XhoI and NotI were purchased from New England Biolabs (NEB, Frankfurt, Germany) and one unit of enzyme per microgram of DNA was used. For a 100-200 µl digestion reaction, the mix typically included 2 to 3 µl of enzyme, DNA sample, 1 X appropriate NEB buffer and 1 X BSA (if required), according to the instructions given by NEB. The reaction was incubated for 3 to 5 hours at 37°C to insure complete digestion and the sample was analysed by agarose gel electrophoresis. Vectors were dephosphorylated with 2 µl shrimp alkaline phosphatase (USB, Cleveland, USA) to prevent religation.

4.3.4.4 Ligation and transformation

For the cloning of DNA fragments several commercially available vectors were used including pGEM (Promega, Mannheim, Germany), pGEX-4T-3 (Amersham, Freiburg, Germany), pMal-cxh (NEB, Frankfurt, Germany) and pCEP4 (Invitrogen, Carlsbad, CA). PCR products were amplified with primers containing restriction site sequences at their 5' ends, respectively. Two distinct cutting sites were used to ensure directional insertion of inserts. The cleavage with two different restriction enzymes was performed sequential or simultaneously by double digestions. For the ligation of PCR products into pGEM, 3 µl of DNA, 1 µl vector, 5 µl 2 X Rapid T4 DNA ligation buffer and 1 µl T4 DNA ligase (NEB, Frankfurt, Germany) were incubated at room temperature for 6 hrs. Pre-digested DNA fragments with appropriate restriction enzymes were cloned into pGEX-4T-3, pMal-cxh and pCEP4 vectors by incubating 10 µl reaction containing 1-3 µl DNA, 3-5 µl plasmid, 1 µl T4 DNA ligation buffer and 1 µl T4 DNA ligase at 14°C overnight. The recombinant plasmid

was then transformed into E.coli XL1-blue electrocompetent cells by electroporation using Gene Pulser Electropotation system (Bio-Rad, Muenchen, Germany)..

4.3.5 Sequencing

The sequencing reactions were done using the ABI PRISM® BigDye™ Terminator Kit (Applied Biosystems, Weiterstadt, Germany) and run on a single-capillary ABI PRISM® 310 genetic analyser (Applied Biosystems, Weiterstadt, Germany). In a 10 µl final volume, 0.5 - 3 µl PCR products were treated with 5 units of exonuclease I (USB, Cleveland, USA) and 1 unit of shrimp alkaline phosphatase (USB, Cleveland, USA), prior to sequencing. The reaction was incubated at 37°C for 15 min, terminating the reaction by enzyme inactivation at 80°C for an additional 15 min. For the 10 µl sequencing reaction, 5 µl of PCR product was mixed with 1 µl PCR of the sequencing primer (10 pmol) and 2 µl BigDye mix. The 25 cycle sequencing protocol consisted of denaturation at 95°C for 30 seconds, followed by annealing at the primer T_m temperature for 30 sec and elongation at 60°C for 30 sec to 3 min. Plasmid DNA was sequenced directly, using a 2 µl aliquot. Amplified PCR products were precipitated using 1/10 volume 3 M sodium acetate (pH 4.8) and 2.5 volume 95% ethanol. Analysis and sequence editing were done using Chromas software (Technelysium Pty Ltd, Helensvale, Australia).

4.3.6 Virtual Northern blot analysis

To obtain full-length first-strand cDNAs synthesis, 2 to 4 µg of total RNA from bladder, heart, lung, skeletal muscle, retina and RPE were used together with SMART III oligonuclotide (see cDNA library, 4.4.2.1), which is contained in the SMART™ cDNA Library Construction Kit (Clontech, Heidelberg, Germany). Double stranded (ds) cDNA synthesis was performed in a 25 µl reaction mix containing 1 X PCR buffer (10 X PCR buffer: 1.2 M Tris-HCl, pH 8.0; 100 mM KCl; 60 mM $(NH4)_2SO_4$; 1% Triton X-100; 0.01% BSA), 100 mM each dNTP mix, 200 nM 5' SMART primer and 200 nM CDS III/3'-PCR primer (Table 20, Appendix), 1 mM $MgCl_2$ and 10 U KOD HiFi Taq polymerase (Merck Biosciences GmbH, Schwalbach, Germany). The skeletal muscle ds cDNA was subjected to 19 cycles of amplification and all other samples were amplified in a 22 cycle PCR reaction.

Five hundred nanogram of amplified cDNA products were electrophoresed on 0.7% TBE agarose gels and blotted with 10 X SSC (1.5 M NaCl, 150 mM NaCitrate - pH 7.0) onto Nylon Hybond N$^+$ membranes (Amersham Biosciences, Freiburg, Germany). Afterwards, cDNA was UV-crosslinked and the membranes were hybridized with radiolabeled DNA probes (see RNA methods, 4.6). After washing, filters were exposed to a storage phosphor screen (Amersham Biosciences, Freiburg, Germany) and scanned after seven days with a Typhoon^{TM} 9200 Imager (Amersham Biosciences, Freiburg, Germany) or exposed to X - ray RETINA films (Fotochemische Werke, Berlin, Germany) at -80°C for 3 days.

4.4   cDNA libraries

4.4.1   Generation of a suppression subtracted cDNA library from the human retina

An adult retina SSH library was constructed in pCR$^®$2.1 TOPO TA cloning vector (Invitrogen, Karlsruhe, Germany) using the PCR-Selected^{TM} cDNA Subtraction Kit (Clontech, Heidelberg, Germany). This work was performed by Dr. Andrea Gehrig at the Institute of Human Genetics in Würzburg (for further details see PhD Thesis of Gehrig A, University of Würzburg, Germany; 2003) and the SSH library was kindly provided for further analysis.

4.4.1.1 Sequencing of the clones from the library

Sequencing was carried out with ABI 310 automated sequencer (as described above) or with a Beckman CEQ 2000 sequencer using the Dye Terminator Cycle Sequencing Chemistry provided by the Quick Start Kit (Beckman Coulter, Fullerton, USA). Colonies deriving from the transformed cells that were grown on LB/ampicilin plates at 37°C overnight were randomly picked, transfered to 100 µl LB medium with ampicilin and grown again overnight. In order to check if they are containing inserts, test PCR using M13F and M13R primers (Table 19, Appendix) was performed and products of positive clones were sequenced with either PCR-f or PCR-r or both primers (Table 19, Appendix).

4.4.2 Full-length cDNA library

4.4.2.1 Construction

Full-length mouse eye and human retina cDNA libraries were constructed in λTriplEx2 vector using the commercially available SMART^{TM} cDNA Library Construction Kit (Clontech, Heidelberg, Germany). Cytoplasmic RNA was purified from 240 mg pooled fresh mice eyes obtained from five animals and 140 mg fresh human retina using the Rneasy® Mini Kit (see RNA methods 4.2.2) with concentrations as determined by spectrophotometry (see RNA methods 4.3). Subsequently, poly(A)$^+$ RNA was extracted from 56.8 μg mouse and 44.7 μg human total cytoplasmic RNA, eluted with 100 μl RNase free water from Oligotex^{TM} suspension (see RNA methods 4.2.3).

4.4.2.1.1 First strand cDNA synthesis

First strand cDNA was synthesised from poly(A)$^+$ RNA, which was separated from total RNA by oligo(dT) selection (see Poly(A)$^+$ RNA isolation 2.2.3), using SMART cDNA Construction Kit (Clontech, Heidelberg, Germany). First strand cDNA synthesis was performed with denatured 0.5 μg poly(A)$^+$ RNA at 72°C for 2 min together with 1 μl SMART III oligonuclotide and 1 μl CDS III/3' PCR primer (Table 19, Appendix) in a 5 μl final volume, following the manufacturers protocol. Subsequently, reverse transcription was carried out by adding 2 μl 5 X First strand buffer, 1 μl 20 mM DTT, 1 μl dNTPs mix (10 mM each) and 1μl SuperscriptII^{TM} (200 U/μl) on 42°C for 1 hr.

During first strand synthesis, reverse transcriptase terminal transferase activity adds a few additional nucleotides, primarily deoxycytidine, to the 3' end of generated cDNA. The SMART III (Switching Mechanism At 5' end of RNA Transcript) oligonucleotide used in this amplification system, has an oligo G sequence at it's end and base-pairs with the deoxycytidine stretch creating an extended template (Fig. 9). Enzyme then switches templates and continues replicating to the end of the oligonucleoide.
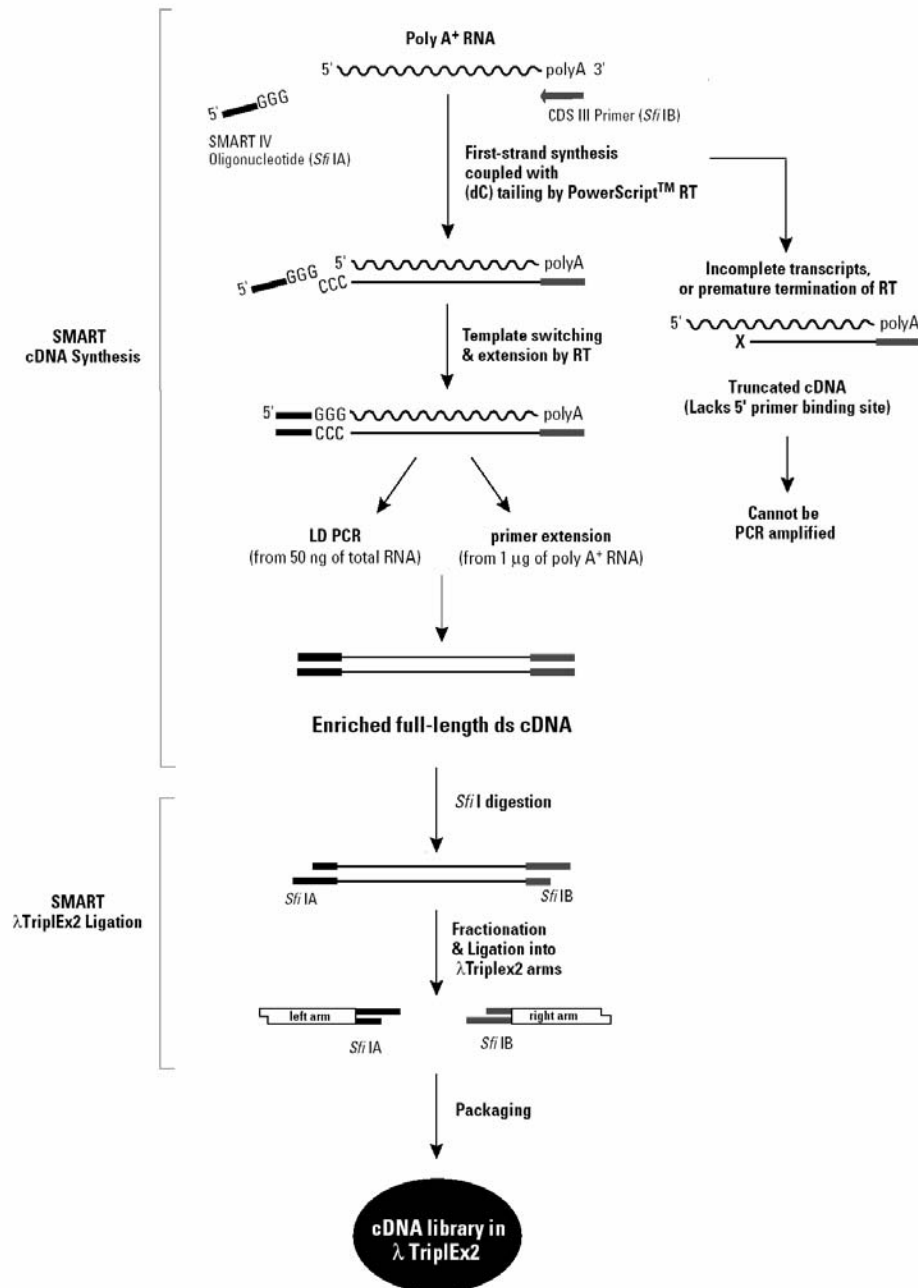
Fig. 9. Overall strategy for full-length cDNA library construction. The right side of the flow chart shows the fate of incomplete transcripts caused by RNA degradation or premature termination of reverse transcription (figure taken from SMART cDNA Library Construction Kit User Manual; PT3000-1; Clontech, Heidelberg, Germany).

4.4.2.1.2 Double strand cDNA synthesis and digestion

In the following step of second strand synthesis by long-distance PCR, only those single stranded cDNAs with complete 5' end of the transcripts were selected through the 5' primer binding site and in this way, clones containing the full-length of the transcript should exclusively be synthesized. Double-stranded cDNA was synthesised using the Advantage 2 PCR Kit (Clontech, Heidelberg, Germany). By mixing 2 µl first strand cDNA with 2 µl 50 X dNTP mix (50 mM each), 2 µl 5' PCR primer (Table 19, Appendix), 2 µl CDS III/3' PCR primer, 10 µl 10 X Advantage 2 PCR buffer and 2 µl 50 X Advantage 2 Polymerase Mix in a 100 µl final volume. Only those single stranded cDNAs having proper primer binding site at the 5' end amplified in the 22 cycle LD PCR. As a control a parallel cDNA synthesis of the provided human placenta poly(A)$^+$ RNA was done. Yield and size distribution of double stranded (ds) cDNAs was determined by analysing 5 µl of the sample on a 1.1% agarose gel. The ds cDNA should appear as 0.1 to 4 Kb smear with some bright bands corresponding to the abundant tissue specific mRNAs. The quality of synthesized double stranded cDNA was tested by amplification of the GUSB gene using the GUS B3/ GUS B5 primers (Table 19, Appendix**)**.

In order to inactivate the DNA polymerase, 270 µl of PCR product (~10-15 µg ds cDNA) were mixed with 100 µg/ml of Proteinase K (20 mg/ml,stock concentration) and incubated on 45°C for 20 min. The DNA was then extracted with phenol/chlorophorm and ethanol precipitated (see DNA methods, 4.3.1). The sample was then digested with 200 units *Sfi* I restriction enzyme by incubating the mixture of 79 µl ds cDNA, 10 µl 10 X Sfi I buffer and 2 µg BSA for 2 hrs on 50°C.

4.4.2.1.3 Size selection of ds cDNA

Prior to ligation into the vector, ds cDNA was size selected by separation on a 0.6% TAE agarose gel without EtBr. A DNA size marker (1 Kb$^+$ DNA ladder; Gibco, Eggenstein, Germany) was run in parallel with the DNA samples at 90V for one and a half hour. Subsequently, the gel lane with the DNA size marker was cut out, stained with EtBr (Applichem, Darmstadt, Germany) for 15 min and used to select the 850 bp – 3.0 Kb fraction of ds cDNA to be purified. The ds cDNA fractions were excised from the gel and purified by using Nucleospin$^®$ Extract Kit (see DNA methods, 4.3.4.1). Samples were eluted from the

columns with 100 µl dH$_2$O and concentrated by centrifugation on a SpeedVac without heating for 2 hrs.

4.4.2.1.4 Ligation to λTriplEx2 vector and packaging

In the next step, ds cDNA was cloned into the *Sfi* I A & B-digested arms of the λTriplEx2 vector by incubating 2 µl dsDNA, 1 µl vector (0.5 µg/µl), 0.5 µl 10 X DNA ligation buffer (500 mM Tris-HCl, pH 7.8; 100 mM MgCl$_2$; 100 mM DTT; 0.5 mg/ml BSA), 0.5 µl ATP (10 mM) and 200 U T4 DNA ligase at 16°C over night, in a 5 µl final volume reaction.

The vectors were incorporated into the bacteriophage using λ phage packaging extracts (Gibco, Eggenstein, Germany) and following the manufacturer's protocol. Extracts were quickly thawed and 4 µl of the ligation reaction containing approximately 0.1-1.0 µg ligated DNA, were added and mixed by pipetting without introducing air bubbles. The tube was incubated at room temperature for 2 hrs. Afterwards, 500 µl SM buffer (5.8 g/l mM NaCl; 2 g/l MgSO4 X 7H$_2$O; 1 M Tris-HCl, pH 7.5; 2% gelatine) and 20 µl chlorophorm were added, gently mixed, and briefly centrifuged to sediment the debris. The supernatant containing the phages was stored at 4°C.

4.4.2.2 Determination of library titre

The titre of the library was determined by λ phage transduction of the host *Escherichia coli* XL1 blue cells. Host cells were grown at 37°C overnight in 15 ml LB/MgSO$_4$/maltose medium (20 g/l peptone, 10 g/l yeast extract, 20 g/l NaCl, 20 mM MgSO$_4$ and 0.2% maltose). The culture was centrifuged at 5.000 rpm for 5 min and the pellet was resuspended in 7.5 ml of sterile 10 mM MgSO$_4$. In order to determine the titre of unamplified library, 600 µl of the host cells were mixed with 10 µl serial dilutions (1:1 to 1:10^4) of library aliquots and incubated for 20 min at 37°C allowing phage infection. Afterwards, 10 ml of top agar (20 g/l peptone, 10 g/l yeast extract, 20 g/l NaCl and 7.2 g/l agar) at 50°C was added to each aliquot, mixed by gentle inversion and poured onto prewarmed Petri dishes (Ø 135 mm) containing LB medium (20 g/l peptone, 10 g/l yeast extract, 20 g/l NaCl and 15 g/l agar). Plates were incubated overnight at 37°C and the next day the number of plaque forming units (pfu) was counted and used for the calculation of the titre (pfu/ml):

$$\text{pfu/ml} \ = \ \frac{\text{number of plaques X dilution factor X } 10^3 \ \mu l/ml}{\mu l \text{ of diluted phage plated}}$$

In order to determine the percentage of clones containing inserts, a test PCR with 1 μl λ dilution buffer containing phages (see 4.2.5) using λTriplEx2 5' and λTriplEx2 3' primers (Table 19, Appendix) was done.

4.4.2.3 Library amplification

In the next step, the library was amplified by plating the primary library onto six Bio-Assay dishes (NUNC, Wiesbaden, Germany). For this purpose 2.1 ml of the host cells were mixed with 83 μl of the unamplified library and incubated (see cDNA libraries 4.4.2.2). After infection, 37 ml top agar was added and plates were incubated on 37°C for 15 to 18 hrs until plaques became confluent. One plate was then washed with precooled 30 ml 1 X λ dilution buffer (10 X λ buffer: 1 M NaCl; 0.1 M MgSO$_4$ X 7H$_2$O; 0.35 M Tris-HCl, pH 7.5) by shaking the plate for 2 hrs at 4°C. The lysate was then collected, transferred to another plate and the same procedure was repeated until all plates were washed. Washing was repeated in the same way with additional 10 ml 1 X λ dilution buffer and end volume of the amplified library lysate was approximately 35 ml since agar plates socked some of the buffer. Subsequently, 700 μl chlorophorm was added, cell debris was precipitated by brief centrifugation and the amplified library was stored at 4°C. The titre was determined by plating out serial dilutions (1: 10^3 to 1:10^8) of library aliquots (see cDNA libraries 4.4.2.2). For the long term storage, DMSO was added to a final concentration of 7% and 1 ml aliquot and tubes was stored at -80°C.

4.4.2.4 Plating of phage library and preparation of replicas

One million clones of the amplified library were plated on six LB Bio-Assay dishes, 2.1 ml host XL1 blue cells were infected with 100 μl library aliquot per plate, as described in (see cDNA libraries 4.4.2.2). Plates were incubated at 37°C for 8 to 16 hrs in order to obtain well separated plaques and prevent cross contamination due to phage diffusion. Duplicates of

nylon membranes were used for making phage replicas from each plate. The Hybond N$^+$ membrane (Amersham Biosciences, Freiburg, Germany) was placed on the surface and unsymmetric needle holes were made in order to make possible the localization of a particular colony after the hybridization of the colony filter. Subsequently, the filter was incubated with denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 2 min and shortly with neutralizing solution (25 mM Na3PO4, pH 6.5) shortly and air dried. Plates and replicas were stored at 4°C.

In addition, three more retina cDNA libraries including DKFZ3 and DKFZ4 (constructed in Heidelberg, Germany) and C1F1 (constructed by Claudia Berger), which were cloned in the λTriplEx2 vector, were plated out following the explained procedure and used for the isolation of the clones containing the sequence of the genes investigated in this project (DKFZ and C1F1 are local laboratory IDs).

4.4.2.5 Isolation of positive clones from the phage library

To isolate phage clones containing an insert of interest, library lifts were screened by colony filter hybridization with radioactively labelled probe (see RNA methods 4.5.3 and 4.5.4). Positive plaques were visualized by autoradiography, excised from the plate and incubated in 350 μl 1 X λ dilution buffer for 3 to 4 hrs at 37°C. In order to obtain optimal concentration of plaques per plate, the phage containing buffer was diluted for second round of plating. Lift for the second screen was made using the Colony/Plaque Screen^{TM} membrane (DuPont, Bad Homburg, Germany), which was hybridized with the same radioactive probe used in the first screening. Plaques corresponding to a positive signal were transferred from the plate to 350 μl 1 X λ dilution buffer.

4.4.2.6 Conversion of λTriplEx2 to pTriplEx2

The conversion of a λTriplEx2 phage to a pTriplEx2 plasmid is performed by Cre recombinase-mediated site-specific recombination. *E.coli* BM25.8 cells were used as host since Cre recombinase is constitutively expressed in these bacteria. They were grown overnight at 31°C in 10 ml LB/MgSO$_4$ medium containing MgCl$_2$ at a 10 mM final concentration and 200 μl were mixed with 150 μl of the eluted positive plaque. After 30 min

incubation at 31°C, 400 μl of liquid LB medium were added and the reaction was incubated for an additional 1 hour at the same temperature with shaking. Subsequently, 10 μl of infected cell suspension were spread on an LB/ampicilin plate and single colonies from each clone were used to inoculate overnight cultures which were subsequently used for plasmid DNA isolation with Nucleospin® Extract Kit (see DNA methods, 4.3.4.2).

4.5   Mammalian cell lines

4.5.1   ARPE-19 cell line

ARPE-19 is a retinal pigment epithelial cell line, which grows in a monolayer. This cell line derived originally from the globe of a 19-year-old male eye donor obtained from the Lions Eye and Tissue Bank, Sacramento, CA, USA (Dunn et al., 1996). Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) / Ham's F12 (1:1 mixture) (Sigma-Algrich, München, Germany) supplemented with L-glutamine, 15 mM HEPES, 42 mM $NaHCO_3$, 100.000 units/l penicillin and 100 mg/l streptomycin (Sigma, Eggenstein, Germany) and 10% fetal calf serum, FCS (PAN Systems, Aidenbach, Germany). The ARPE-19 cell line was maintained in a 5% CO_2 environment at 37°C.

4.5.2   Y79 cell line

The Y79 cell line, which grows in a suspension, was derived from human retinoblastoma cells (Green et al., 1979). The cell line was obtained from the Westdeutsches Tumorzentrum Essen, Germany. Cells were grown in DMEM medium supplemented with 10% FCS, 100.000 units/l penicillin and 100 mg/l streptomycin. The Y79 cell line was maintained in a 5% CO_2 incubator at 37°C.

4.5.3 EBNA-293 cell line

4.5.3.1 Maintenance

The EBNA-293 cell line is composed of human embryonic kidney cells expressing Epstein-Barr virus (EBV) replicons, which include the genetic element oriP and a functional gene for Epstein-Barr nuclear antigen (EBNA-1) (Young et al., 1988). The cells, growing in monolayer, were cultured in DMEM medium containing L-glutamine, 4500 mg/l glucose and 110 mg/l sodium pyruvate, supplemented with 10% FCS, 100.000 units/l penicillin, 100 mg/l penicillin, 100 mg/l streptomycin and 250 mg/l geneticin (G418 sulphate, Calbiochem, San Diego, USA), on 5% CO_2. The EBNA-293 cell line was maintained in a 5% CO_2 incubator at 37°C.

4.5.3.2 Transfection

Transfection of the EBNA-293 cells was achieved by the calcium phosphate precipitation method (Chen et al., 1987). Cells were transformed with pCEP-1D4 plasmid constructs generated by inserting the respective insert into the multiple cloning site of the vector (primers listed in Table 21, Appendix). EBNA cells were seeded in a 10 cm Petri dish 24 hrs prior to transfection. Cells were transfected with 20 μl plasmid DNA that was mixed with water and 123 μl 1M $CaCl_2$ in 495 μl final volume. After 3 min of incubation, equal volume of 2 X BBS (50 mM N,N-bis[2-Hidroxyethyl]-2 aminoethane sulfonic acid, 280 mM NaCl, 1.4 mM Na_2HPO_4 X 2 H_2O, pH 6.9) was added by dropping, with permanent gentle vortexing. The mixture was incubated for 17 min at room temperature to allow formation of calcium phosphate - DNA complexes and afterwards added onto the cells drop by drop. EBNA cells were incubated with transfection mixture for 16 hrs under their normal growth conditions. Subsequently, cells were washed with 1 X PBS (137 mM NaCl; 0.3 mM KCl; 8.1 mM Na_2HPO_4; 6.7 mM KH_2PO_4, pH 7.4) containing 1 mM Na_2EDTA and fresh medium was added. After two days of growth, EBNA cells were harvested and used for the preparation of protein extracts.

4.5.3.3 Purification of recombinant proteins

EBNA-293 cells were solubilized in 1 X PBS with 2% Triton X-100, 20% glycerol and protease inhibitor, 1% SDS was added by dropping and vortexed for 30 min at 4°C. Afterwards, DNA, RNA and membrane fragments were precipitated by centrifugation at 14.000 rpm for 20 min at 4°C. Supernatant containing soluble proteins was transferred to a clean tube and used for Western blot analysis.

4.6   Nonsense mediated mRNA decay

4.6.1   Principle of the decay

Nonsense-mediated mRNA decay (NMD) is a molecular RNA surveillance system that detects and degrades mRNA transcripts containing premature termination codons (PTCs) whose translation could otherwise lead to potentially harmful truncated protein products (Wilkinson et al., 2002). In the case of mRNA that contains a premature termination codon more than 55 nt upstream of the last exon–exon junction, termination of translation occurs and a surveillance complex triggers decapping of the mRNA and its degradation (Fig. 10).

4.6.2   Inhibition of protein synthesis

Nonsense-mediated mRNA decay requires translation for nonsense codon recognition. It is inhibited by the antibiotics anisomycin, cycloheximide, emetine, pactamycin or puromycin, each of which binds to and inactivates translationally active ribosomes (Maquat 2004). In order to block the decay, protein synthesis was inhibited by puromycin (isolated from *Streptomyces albo-niger* causing premature chain termination by acting as an analogue of the 3' terminal end of aminoacil-tRNA) and by anisomycin (isolated from *Streptomyces griseolus*, inhibiting peptidyl transferase activity in ribosomes).

Puromycin and anisomycin (Gibco, Eggenstein, Germany) were used at concentration of 100 μg/ml (Lamba et al., 2003). The ARPE 19 and Y79 cell lines (Mammalian cell lines, 4.5.1 and 4.5.2) were grown in antibiotic-containing medium for 2 hrs. Subsequently, cells were washed with 1 X PBS and then grown in medium without antibiotic for an additional 4

hrs. Total RNA was extracted using the RNeasy® Mini Kit (RNA methods, 4.2.1) and the amount of ABCC5 isoform transcripts (ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3) were quantified with qRT-PCR (DNA methods, 4.3.2.2).
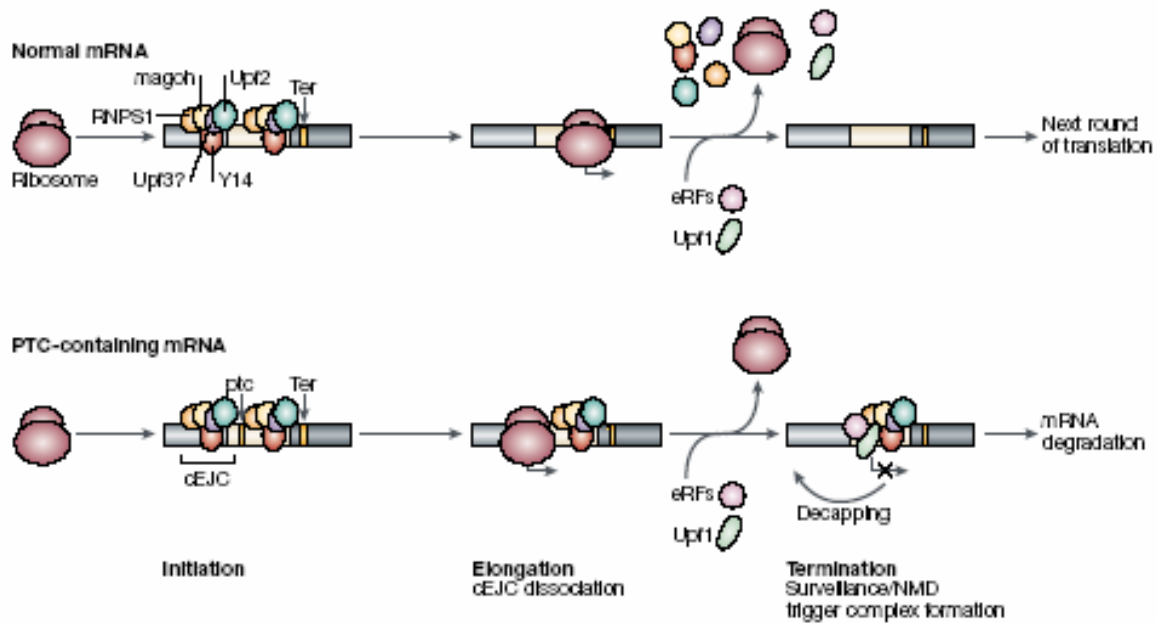
Figure 10. **Role of the exon–exon junction in nonsense-mediated mRNA decay.** Translating ribosome removes all the exon junction complexes (cEJCs) by the time of termination and translation proceeds to the next round. In the case of mRNA that contains PTC that is more than 55 nt upstream of the last exon–exon junction, translation termination occurs before all the EJCs are removed and surveillance complex triggers decapping of the mRNA and its degradation. (figure taken from Dreyfuss et al., 2002).

4.7   RNA interference

4.7.1   Principle of RNA interference

A conserved biological response to double-stranded RNA, known as RNA interference (RNAi) or post-transcriptional gene silencing, mediates resistance to both endogenous parasitic and exogenous pathogenic nucleic acids, and regulates the expression of protein-coding genes (Hannon 2002).

RNAi has been cultivated as a means to experimentally manipulate gene expression in mammalian cells. Introduction of double-stranded RNA into the cells results in targeted post-transcriptional gene silencing. The ability to specifically down-regulate the expression of the

targeted genes holds enormous scientific, commercial and therapeutic potential. RNAi can be induced in the cell using chemically synthesised individual double-stranded siRNAs duplexes.

4.7.2   siRNA design and sequences

The siRNA duplexes (HPP grade, 20nM) were synthesised by Xeragon (Qiagen, Hilden, Germany). The antisense siRNA was targeted against an $AA(N)_{19}$ sequence (N, any nucleotide), the GC content of the duplexes was kept within 40-60% range and poly G and poly C sequences were avoided. Synthetic 21mer siRNA oligonucleotide was created with a dTdT 3'overhangs.

siRNA.1 is targeting the ABCC5 isoforms ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3

Target DNA sequence		5'-AATTCAGCGTAGCTACCTCCA-3'
siRNA duplex sequence	Sense:	5' – UUCAGCGUAGCUACCUCCAd(TT) – 3'
	Antisense:	5'- UGGAGGUAGCUACGCUGAAd(TT) – 3'

siRNA.2 is targeting the ABCC5 isoforms ABCC5_SV1 and ABCC5_SV2

Target DNA sequence		5'- AATCTCTCGCCAAGAGTTCAG -3'
siRNA duplex sequence	Sense:	5' – UCUCUCGCCAAGAGUUCAG d(TT) – 3'
	Antisense:	5'- CUGAACUCUUGGCGAGAGA d(TT) – 3'

siRNA.3 is targeting the ABCC5_SV1 isoform of ABCC5

Target DNA sequence		5'- AACCAGTGACAGAAGCTGTTC -3'
siRNA duplex sequence	Sense:	5' – CCAGUGACAGAAGCUGUUCd(TT) – 3'
	Antisense:	5'- GAACAGCUUCUGUCACUGGd(TT) – 3'

Non-coding siRNA, without known homology to mammalian sequences (used for the control experiment)

Targeted gene		No known homology to mammalian sequences
Target DNA sequence		5'- AATTCTCCGAACGTGTCACGT -3'
siRNA duplex sequence	Sense:	5' – UUCUCCGAACGUGUCACGUd(TT) – 3'
	Antisense:	5'- ACGUGACACGUUCGGAGAAd(TT) – 3'

The lyophilized oligonucleotides were dissolved in the provided Suspension buffer (100 mM potassium acetate, 30 mM HEPES-KOH, 2 mM magnesium acetate, pH 7.4) and annealed to double stranded siRNA duplexes by heating at 90°C for 1 min and incubating at 37°C for additional 60 min. The dissolved siRNA was and stored at -20°C.

4.7.3 siRNA transfection

Human ARPE 19 cells were seeded in 6-well plates (1 X 10^6 cells/well) 24 hrs prior to transfection. Transfection of the cells with synthetic siRNA duplexes was carried out using TransMessenger[TM] Transfection Reagent (Qiagen, Hilden, Germany). Duplexes were used at different concentrations of 2.0, 4.0 or 8.0 μg/well. In the first step of TransMessenger-RNA complex formation, the siRNA was condensed by interacting with the provided Enhancer R reagent (1:2 ratio) in a denaturing buffer system and preincubated for 5 minutes at room temperature. In the second step, TransMessenger reagent (8 μl/well) was added to the condensed RNA and incubated for 10 min at room temperature for lipid complex formation. Subsequently, cell growth medium without serum and antibiotics (to avoid possible RNase contaminations), was added and the entire mixture was immediately dropped onto the cells which were then incubated for 3.5 hrs with the transfection mix under normal growth conditions. Afterwards, the transfection mix was removed, cells were washed with 1 X PBS and medium with serum and antibiotics was added. Total RNA was isolated from the cells at different posttransfectional time points with Rneasy® Mini Kit (RNA methods, 4.2.1) and the amount of ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3 transcripts were quantified by qRT-PCR (DNA methods, 4.3.2.2).

4.8 Polyclonal antibody production

4.8.1 Expression of recombinant fusion proteins

Recombinant Glutatione-S-Transferase (GST; Amersham, Freiburg, Germany) and Maltose-Binding-Protein (MBP; New England Biolabs, Frankfurt, Germany) fusion proteins were constructed by inserting the gene fragment of interest into the multiple cloning site of the pGEX-4T-3 and pMal-cxh vectors. Expression of recombinant proteins in BL21 *E.coli* strain was induced by the lactose analog isopropyl b-D thiogalactoside (IPTG) since it is

under control of the *tac* promoter. BL21 cells were transformed with 20 μg plasmid DNA and grown in LB medium with 0.3 mM IPTG for 3 hrs. After incubation, cells were precipitated by centrifugation for 15 min on 4.000 g at 4°C and pellet was lysed with sonication. GST fusion proteins were purified from the cell lysate for immunization in mice (strain Balb/c), whereas lysates containing MBP fusion proteins were used afterwards for the analysis of the immune response of the mice by immunoblotting.

4.8.2   Purification of GST fusion proteins

GST fusion proteins were purified by affinity chromatography using glutation sepharose 4B beads (Amersham, Freiburg, Germany). Cell lysate was applied to the column containing beads, unbound proteins were cleaned up using 1 X PBS and fusion proteins were then eluted with elution buffer (10 mM reduced glutathion, 50 mM Tris-HCl pH 8.0 and 3 mM Na_2EDTA) by incubation on room temperature for 10 min.

4.8.3   Immunization

GST fusions proteins coupled to SV2 peptide sequence (5'-NFQDGCILRSE-3', 198-208 aa of the L25SV2 protein, MW ~1.1 kDa; clone number: pGEX4T3_SV2_6) or SV3 peptide sequence (5'-PSFGDCSISAEVCGNRLHCTAILLSCFT-3', 198-225 aa of the L25SV3 protein, MW ~ 2.95 kDa; clone number: pGEX4T3_SV3_24) were used for the immunization. Immunization of the rabbit was performed by ImmunoGlobe (Himmelstadt, Germany). Immunization of the eight weeks old Balb/c female mice (Harlan Winkelmann, Germany) was carried out using the ImmunEasy Mouse Adjuvant (Qiagen, Hilgen, Germany), following the manufacturer's protocol. The mixture of 5-10 μg of fusion protein and 15 μg ImmunEasy Mouse Adjuvant was incubated for 5 min at room temperature. Mice were subcutaneously injected with protein suspension, and boosted two times at an interval of two weeks (Table 21, Appendix).

4.8.4   Analysis of the immune response

To determine if mice show positive immune response against the peptide of interest, MBP fusion proteins containing the same polypeptide were analysed by immunoblotting using the immune serum from the injected mice. In parallel, the recombinant SV2 full-length protein (peptide sequence 5'- MKDIDIGKEYIIPSPGYRSVRERTSTSGTHRDREDSKFRRT RPLECQDALETAARAEGLSLDASMHSQLRILDEEHPKGKYHHGLSALKPIRTTSKHQ HPVDNAGLFSCMTFSWLSSLARVAHKKGELSMEDVWSLSKHESSDVNCRRLERLWQ EELNEVGPDAASLRRVVWIFCRTRLILSIVCLMITQLAGFSG<u>PNFQDGCILRS</u> -3', aa of the L25SV2 protein, MW ~ 25 kDa) and SV3 full-length protein (peptide sequence 5'- METKDIDIGKEYIIPSPGYRSVRERTSTSGTHRDREDSKFRRTRPLECQDALETAARAE GLSLDASMETHSQLRILDEEHPKGKYHHGLSALKPIRTTSKHQHPVDNAGLFSCMET TFSWLSSLARVAHKKGELSMETEDVWSLSKHESSDVNCRRLERLWQEELNEVGPDA ASLRRVVWIFCRTRLILSIVCLMETITQLAGFSGP<u>PSFGDCSISAEVCGNRLHCTAILLS CFT</u> -3', aa of the L25SV3 protein, MW ~ 25 kDa) were fused to the 1D4 tag (amino acid sequence: 5'-TETSQVAPA-3'), clone numbers: pCEP4_SV2_4 and   pCEP4_SV3_3, respectively, and subjected to immunobloting in order to test if polyclonal antibodies generated against part of the protein (underlined amino acid sequence) reveals immunaffinity. Immune serum was separated from the blood cells with BD microtainers (Becton Dickinson, Heidelberg, Germany) by 2 min centrifugation on 10.000 rpm, using the renal vain blood samples.

4.8.5   Western blot analyses
4.8.5.1 SDS - polycrilamide gel electrophoresis (SDS-PAGE)

Proteins were electrophoretically separated by SDS-PAGE using the Mini-protean system (BioRad, München, Germany). The SDS-polyacrylamide gel was composed of two layers: stacking buffer (0.4% SDS, 50 mM Tris-HCl, pH 6.8) was used for upper gel and separating buffer (0.4% SDS, 150 mM Tris-HCl, pH 8.8) was used for lower gel preparation. The stacking gel ensures that proteins are entering the separating gel simultaneously. To activate polymerization, 10 µl TEMED and 30 µl 10% ammonium persulfate solution (APS) were added to the 10 ml gel mixture. Prior to loading, proteins were mixed with 4 X sample buffer, made from stock sample buffer (240 mM Tris-HCl, 240 mM SDS, 40% glycerol, 20%

β-mercaptoethanol, pH 6.8), and denatured on 90°C for 3 min. In general, protein molecules bind SDS during heating and negatively charged SDS-polypeptide complexes migrate along the gradient gel according to their size. A prestained protein molecular weight marker (BioRad, München, Germany) was run in parallel with 10-30 µl of the sample loaded per lane. The acrylamide gels were run for 1.5 hrs at a constant voltage of 120V in a 1 X SDS buffer (10 X SDS buffer: 30 mM Tris, 14% glycine, 1% SDS; pH 8.6). After electrophoresis, protein gels were subjected to immunoblotting.

4.8.5.2 Immunoblotting

Proteins were transferred to a Millipore Immobilon-P membrane (PVDF membrane, Eschborn) using the Trans-Blot system (BioRad, München, Germany). The membrane was activated with 100% methanol for 30 sec and transfered at a constant current of 40 mA for 2 hrs in transfer buffer (25 mM Tris, 192 mM glycine, 15% methanol). Afterwards, the membrane was blocked in 1 X PBS containing 2% non-fat dry milk for a minimum of 2 hours. Mice serum was diluted in 0.2% dry milk / 1 X PBS, blot was incubated for 1 hour at room temperature with shaking and washed 3 minutes in 1 X PBS / 0.05% Tween. Blot was then incubated with secondary goat anti-mouse antibody (Calbiochem, San Diego, USA), diluted in 0.2% dry milk / 1 X PBS, for a minimum of 30 min at room temperature with shaking and washed again 3 times in 1 X PBS / 0.05% Tween. Secondary anti-mouse antibodies conjugated to horseradish peroxidase were detected with chemiluminescence detection system (SuperSignal West Pico Chemiluminescent Substrate; Perbio, Bonn, Germany) and membrane was exposed to X - ray RETINA films (Fotochemische Werke, Berlin, Germany).

5   RESULTS

5.1   Evaluation of a cDNA library enriched for retinal transcripts by the suppression subtraction hybridization (SSH) technique

Efficient suppression subtraction hybridization decreases the concentration of (housekeeping) genes expressed ubiquitously, and conversely increases the concentration of genes expressed differentially in a given tissue (Diatchenko *et al.*, 1996). Human retina RNA was amplified by reverse transcription (RT) using oligoT primers, subtracted with cDNAs from two nonocular tissues, namely kidney and liver, and cloned into the pCRII vector (see Materials and Methods, 4.4.1.1.).

In total, 1113 clones randomly isolated from the retina SSH cDNA library were partially sequenced by Maren Weikert and Lynkeus BioTech Company (Lynkeus, Science Park Würzburg). The distribution of the insert length was between 150 to 850 bp. The sequences of cDNAs were analysed and compared with sequences deposited in the GenBank database. On the basis of BLASTN algorithm analysis, clones were classified in four categories including those with I) significant homology to known human genes (766/1113), II) significant homology to partial transcripts and hypothetical gene predictions (162/1113), III) no homology to known mRNAs (149/1113), and IV) vector sequences and clones derived from mitochondrial genes (36/1113) (Table 1). The extent of redundancy in clones originating from known genes, ESTs and novel sequences was also determined. After correcting for redundancy, category I represented 234 known human genes and category II a total of 92 unknown transcripts (Table 1).

Table 1. Bioinformatic analysis of 1113 clones from the retina SSH cDNA library

Category	Homology to	Number of clones	Number of genes
I	Known human gene	766	234
II	Partial transcripts and hypothetical gene predictions	162	92
III	Unknown human gene	149	unknown
IV	Vector sequences and clones derived from mitochondrial genes	36	6
Total number of genes		1113	>332

Alignment of clone sequences demonstrated that 96% of the inserts derived from known genes, unknown sequences or novel retina ESTs. As further shown in Fig. 11, 28% of cDNAs do not show identity to known GenBank sequences and therefore represent novel human retinal transcripts. Of these, 15% were classified as partial transcripts or hypothetical genes, whereas 13% of sequences derived from anonymous sequences, i.e. sequences reported by ESTs and genomic sequences with no annotation or assigned function. A total of 68% of analyzed clones showed identity to known GenBank entries. Minor fractions of inserts derived from mitochondrial genes (1%) or were vector sequences (3%). cDNAs representing known and mitochondrial genes as well as vector sequences were discarded from further analysis. All *in silico* analysis of the clones were performed on February 2002.

1113 clones from retina SSH cDNA library divided into subcategories

- Known human gene 766/1113
- Partial transcripts and hypothetical gene predictions 162/1113
- Unknown human gene 149/1113
- Mitochondrial genome 7/1113
- Vector sequences 29/1113
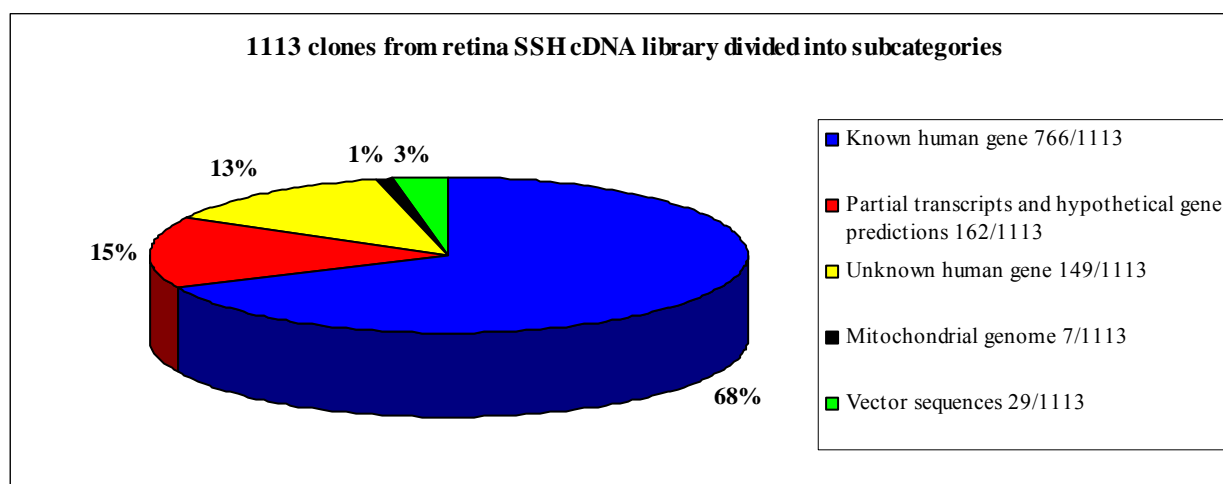
Fig. 11. BLASTN analysis of 1113 clones from retina SSH cDNA library.
Based on BLASTN homology search, 1113 clones were divided into subcategories including those that derived from known human gene (68%), partial transcripts and hypothetical genes (15%), unknown human gene (13), vector sequences (3%) and clones which derived from mitochondrial genes (1%).

5.1.2 Expression analysis

Clones from category II, representing a total of 92 unknown transcripts, were selected for expression analysis by RT-PCR which was performed using RNA from 9 human tissues including brain, heart, lung, retina, RPE, placenta, thymus, uterus and fibroblast (Fig. 12). For each gene, primers were designed based on the clone sequences (primers are listed in Appendix, Table 20).
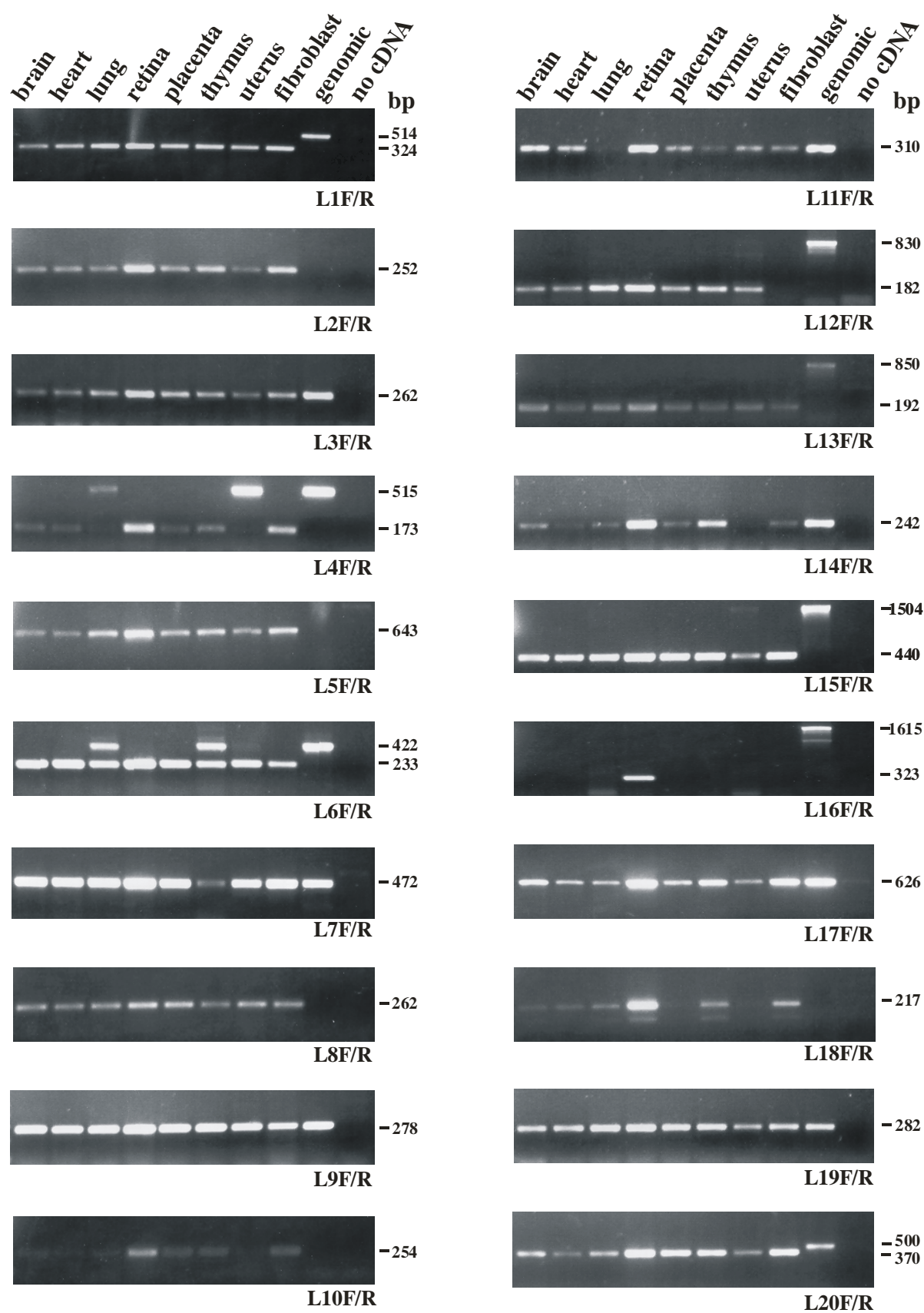
Fig.12. Expression analysis of 92 cDNAs isolated from retina SSH cDNA library, representing partial transcripts and hypothetical genes, by RT-PCR in a panel of 9 human tissues: 1) brain, 2) heart, 3) lung, 4) retina, 5) RPE, 6) placenta, 7) thymus, 8) uterus, as well as 9) fibroblast cDNA, 10) genomic DNA, 11) negative control. L represents internal laboratory ID.
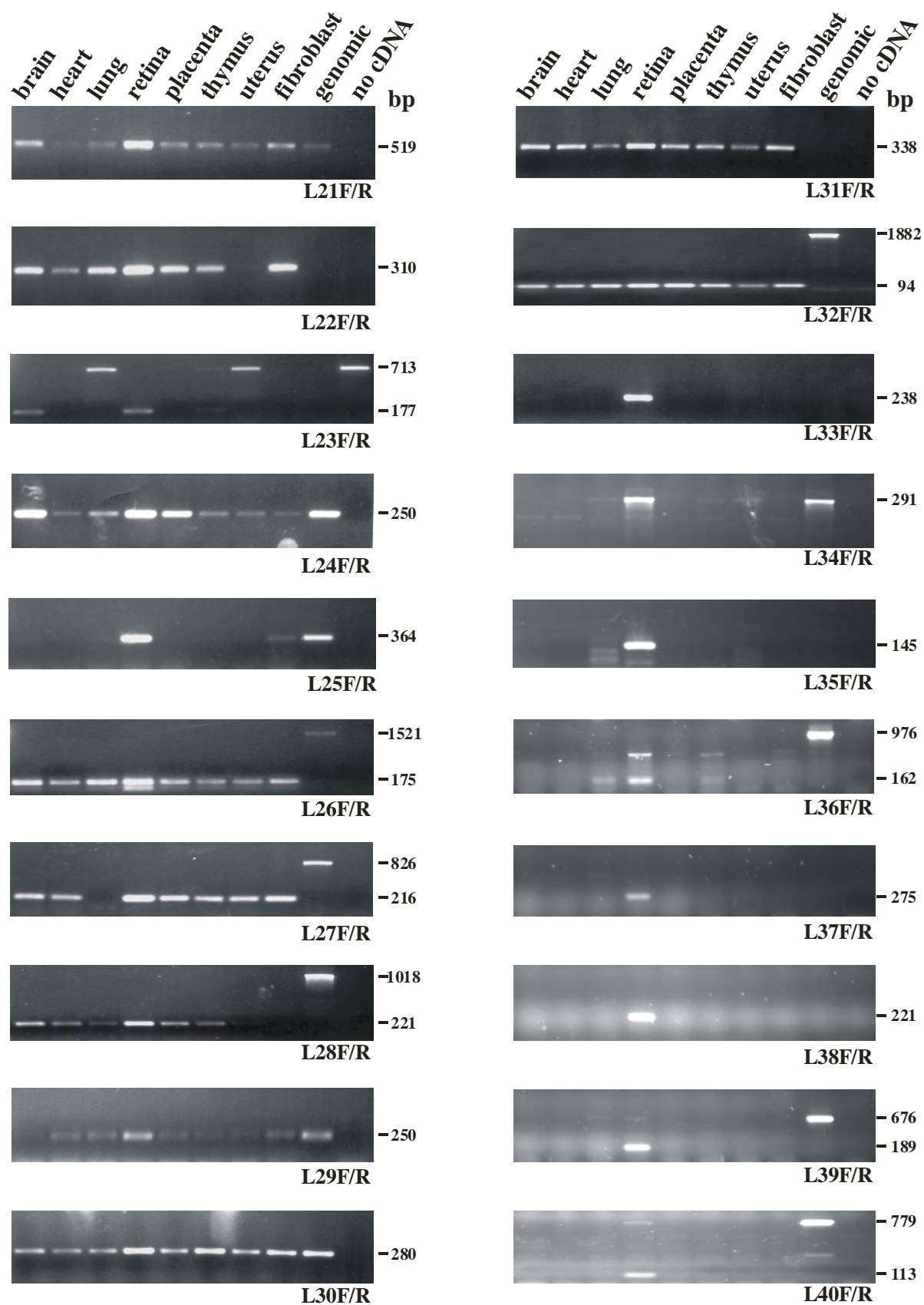
Continue Fig. 12. Expression analysis of 92 cDNAs isolated from retina SSH cDNA library, representing partial transcripts and hypothetical genes, by RT-PCR.

Continue Fig. 12. Expression analysis of 92 cDNAs isolated from retina SSH cDNA library, representing partial transcripts and hypothetical genes, by RT-PCR.
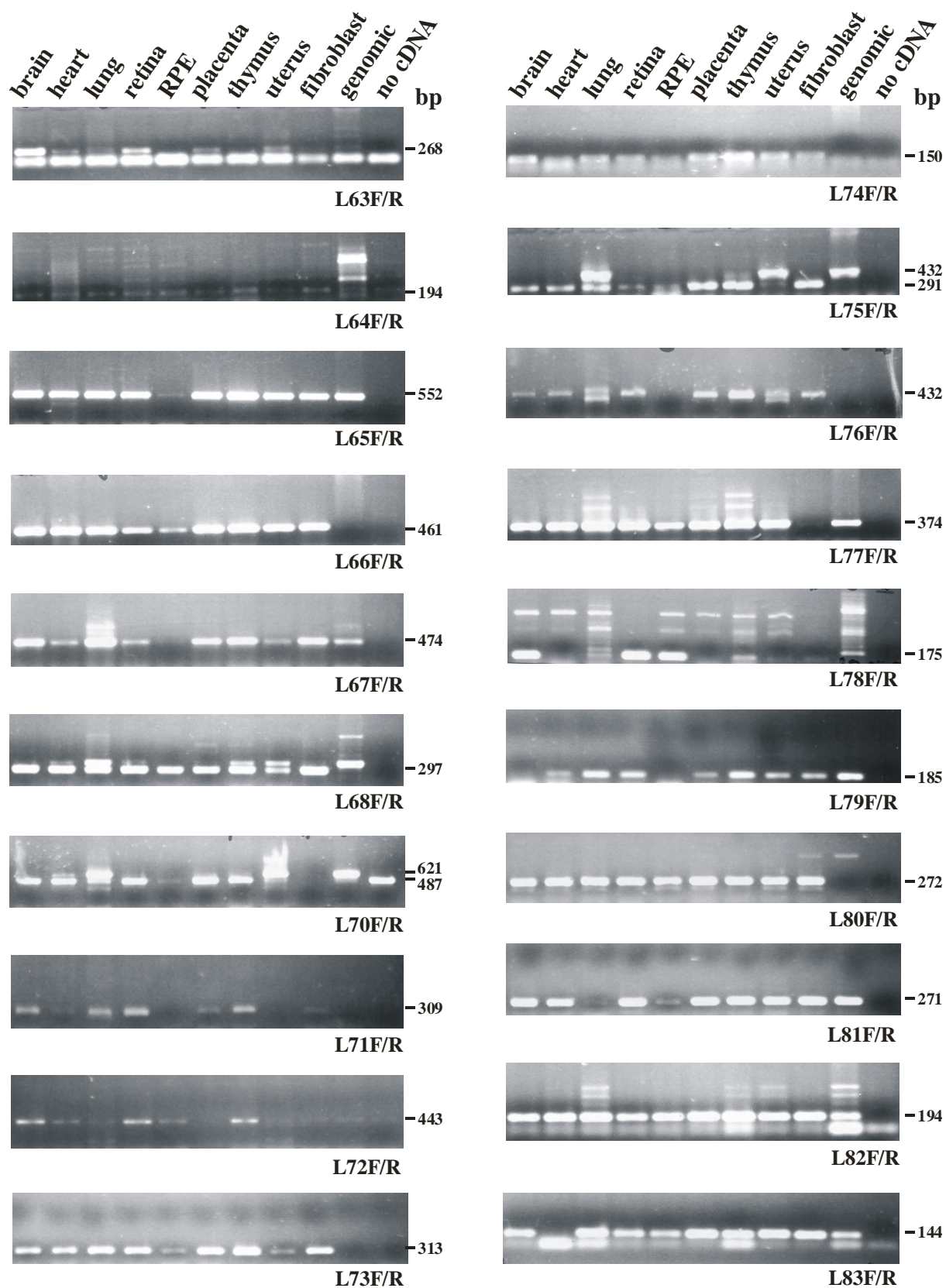
Continue Fig. 12. Expression analysis of 92 cDNAs isolated from retina SSH cDNA library, representing partial transcripts and hypothetical genes, by RT-PCR.

Continue Fig. 12. Expression analysis of 92 cDNAs isolated from retina SSH cDNA library, representing partial transcripts and hypothetical genes, by RT-PCR.

Based on the expression pattern, analysed transcripts were divided in four categories, including 1) retina-specific, 2) neuronal-specific, 3) abundant in retina and 4) ubiquitously expressed (Table 2). Representative examples of expression profiles, determined by RT-PCR analysis, are L33, L45, L18 and L9 genes, respectively (Fig. 12). For three genes (L50, L51 and L69) conditions for PCR amplification could not be optimized. Of 92 genes analysed, 16 were expressed exclusively in retina, 13 were expressed at considerably higher amounts in retina compared to the other tissues, and 12 were generally expressed in neuronal tissues (Table 2). Thus, the expression analysis resulted in the identification of 29 genes exclusively or abundantly transcribed in the human retina. These transcripts are of great potential interest for future investigations (Table 3). Of these, retina specific genes L25, L33, L35, L37, L38 and L40 were selected for further analysis in the course of this thesis.

Table 2. Evaluation of the expression analysis of the 92 genes from category II

Gene expression	Partial transcripts and hypothetical gene predictions
Retina-specific expression	16
Abundant expression in retina	13
Neuronal-specific expression	12
Ubiquitous expression	48
No data	3
Total number of genes	**92**

Table 3. Chromosomal location and expression pattern of 29 genes that are expressed specifically or abundantly in the human retina

Lab ID	Chromosomal localization	RT-PCR expression
L02	3p23	Abundant in retina
L04	1p22.3	Abundant in retina
L05	10q21.3	Abundant in retina
L10	14q32.2	Abundant in retina
L14	2q33.1	Abundant in retina
L16	2	Retina-specific
L18	2p15	Abundant in retina
L20	5	Abundant in retina
L21	12p12.1	Abundant in retina
L25	3q27.1	Retina
L27	2	Abundant in retina
L28	15q26.1	Abundant in retina
L30	4	Abundant in retina
L33	15q23	Retina-specific
L34	5q13.2	Retina-specific
L35	7p21.2	Retina-specific
L36	12p11.21	Abundant in retina
L37	3q22.1	Retina-specific
L38	12q21.2	Retina-specific
L39	4q21.22	Retina-specific
L40	7p21.2	Retina-specific
L47	1p21.2	Retina-specific
L48	11q12.3	Retina-specific
L48	11q12.3	Retina-specific
L50	11q23.3	Retina-specific
L54	9q33.1	Abundant in retina
L56	11q13.5	Retina-specific
L59	3p21.31	Retina-specific
L86	3p14.1	Retina-specific

5.2   Construction of a human retina full-length cDNA library

5.2.1   Optimization of the conditions for the full-length cDNA library construction

To identify the full-length mRNA sequences of the retina-specific genes a full-length human cDNA library was constructed. Full-length cDNA identification and characterisation still remains the most reliable way to determine the exon – intron organization of a gene. Pre-mRNA, transcribed in the nucleus, still contains exons and intervening sequences. After splicing, the intronic sequences are eliminated, and mature mRNA transcripts containing exonic sequences are transported to the cytoplasm. For cDNA library construction, the cytoplasmic RNA fraction was used. By removing a bulk of nuclear pre-mRNA, this modification significantly decreases the amount of intron-containing clone inserts.

Construction of a mouse retina cDNA library was first performed to optimize the conditions for subsequently generating a human retina cDNA library. Efficiency of cytoplasmic RNA isolation was tested using XISTF and XISTR primers (Table 19, Appendix) (Fig. 13). The XIST gene is unique among X-linked genes in being exclusively expressed from the inactive X chromosome in mammals, with the majority of the transcripts having nuclear localization (Das et al., 2001 and Brown et al., 1992). Therefore, efficiency of the cytoplasmic RNA separation from the nuclear RNA fraction, isolated from the female mouse retina, was tested by performing PCR using the XIST primers.
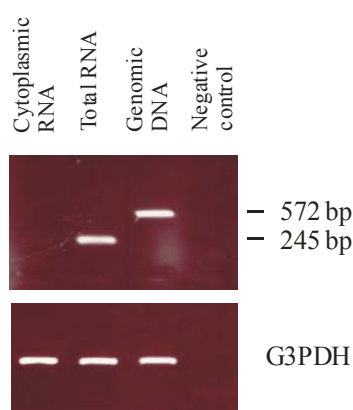
Fig. 13. Fractionation of the cytoplasmic RNA from the mouse retina.
RT-PCR analysis of the XIST gene in cytoplasmic and total RNA fraction using primers directed against the nuclear mouse XIST transcript (see Appendix, Table 19). Cytoplasmic RNA preparation does not contain any contaminating nuclear XIST transcript, as shown on the picture. Integrity of the generated cDNA was tested using primers which are annealing to the mouse G3PDH gene (G3PDHEx8F and G3PDHEx9R, Appendix; Table 19).

The mRNA was extracted from the cytoplasmic RNA (Fig 14A) and used for the first and second strand cDNA synthesis (see Materials and Methods, 4.4.2.1.1 and 4.4.2.1.2). As shown in Fig. 14B, the ds cDNA appeared as a 0.1 - 4 Kb smear, as expected, with several bright bands distinguishable against the background smear, corresponding to the abundant mRNAs for this tissue. In order to avoid generation of the non-specific PCR products by overcycling, the optimal number of cycles for the second strand cDNA synthesis was assessed as 22 thermal cycles (Fig. 14B and 14C).
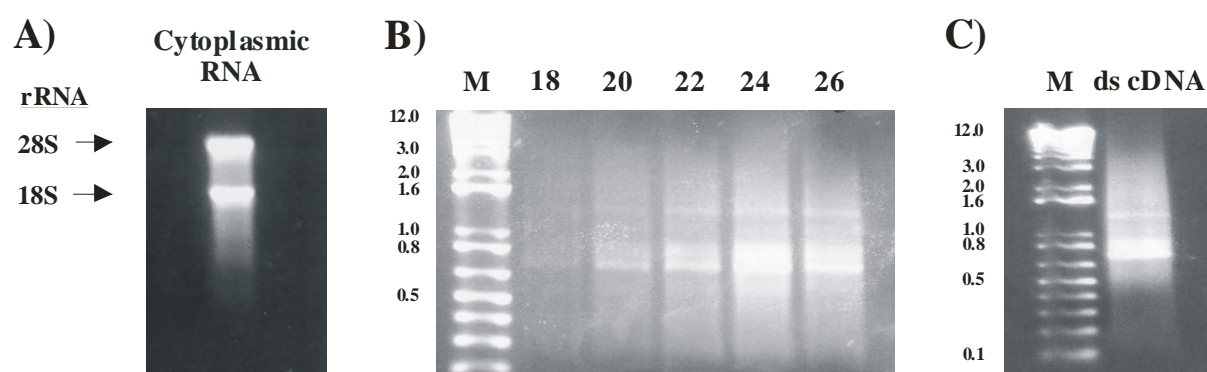
Fig. 14. Cytoplasmic RNA and synthesized ds cDNA. A. Formaldehyde/agarose gel analysis of the cytoplasmic RNA isolated from fresh mouse retina. Ten micrograms of RNA was fractionated on a 1% agarose/formaldehyde gel and visualised by staining with EtBr. Arrows indicate the positions of migration of the 28S and 18S rRNA, respectively. **B.** 6 µl (0.5 µg) of the mouse retina poly (A)$^{+}$ was used as a starting material in a first-strand cDNA synthesis. Afterwards, 2 µl of the first strand cDNA served as a template for the optimization of the number of cycles for the long distance PCR based second-strand synthesis, using the 18, 20, 22, 24 and 26 thermal cycles. **C.** Analysis of a 5 µl sample of the ds cDNA amplified with 22 cycles, which is thought to be optimal, alongside of the 1-Kb DNA size marker, on a 1% agarose/EtBr gel.

The 1 – 12 Kb ds cDNA fraction was separated from the low-molecular-weight cDNA fragments in order to avoid generation of the library having the abundance of small inserts and ligated into the λTriplEx2 vector. The packaged DNA produced more than 1 X 10^{5} plaques. From ~ 50.000 obtained independent clones, more than 84% were recombinant. After their amplification, titre of the final cDNA library was more than 2 X 10^{9} plaques. To determine the average size of the cloned cDNA, 36 clones were isolated from the amplified cDNA library. Subsequently, plasmid DNA was prepared from randomly picked clones and the length of the inserts was examined using agarose gel electrophoresis. The results are summarized in the Table 4. There were 11.1% of clones having inserts longer than 2Kb, 27.8% of cloned inserts which are 1-2 Kb in size, whereas the same percentage of clones had inserts below 1 Kb, indicating that size fractionation of the ds cDNA was not overly efficient.

Table 4. Size distribution of isolated clones from the mouse cDNAs library

Fragment size (bp)	Number of clones	Percentage of clones analyzed (%)
no insert	8	22.2
0 ~ 500	9	25.0
500 ~ 1000	5	13.9
1000 ~ 2000	10	27.8
> 2000	4	11.1
Total	36	100.0

Obtaining the complete 5' end of a transcript is the most difficult step in full-length cDNA cloning. To evaluate the efficacy of the cDNA cloning for the construction of the mouse retina cDNA library, 9 clones were selected for the sequencing of both ends using λTriplEx5' and λTriplEx3' primers (Table 19, Appendix). The 5' part was considered complete when the published initiation site for transcription was included within the sequenced clone insert. The results show that the percentage of the clones having homology to the coding sequence of the gene (CDs) for those with mRNA of <1Kb, 1 - 2 Kb, and >2 Kb were 100.0%, 80% and 66.7%, respectively (Table 5). When the initiation site of the translation was searched, results demonstrated that the likelihood for clone to contain the 5' end was mainly dependent by the size of mRNA since for mRNAs <1 Kb and between 1 and 2 Kb, the percentages of the inserts having initiation site of translation were 100 and 40.0, respectively. For genes with mRNAs longer than 2 Kb none of the clones contained the translation initiation ATG (Table 5).

Table 5. Coverage of the 5' end of known genes in selected mouse cDNA clones

mRNA size	Number of clones	CDs included (%)	5' end included (%)[*]
<1 Kb	3	3 (100.0)	3 (100.0)
1-2 Kb	5	4 (80.0)	2 (40.0)
>2 Kb	3	2 (66.7)	0
Total	11	9 (81.8)	5 (45.5)

*Defined as the clone containing published initiation site of translation.

5.2.2 Evaluation of the human retina full-length cDNA library

Human retina full-length cDNA library (local library ID: Jelena) was constructed using the same conditions optimized for the construction of the mouse cDNA library. The fraction of 850 bp – 3 Kb of synthetized ds cDNA (Fig. 15C) was selected for the ligation into the λTriplEx2 vector. The packaged DNA produced more than 1.3×10^6 plaques. From ~ 650.000 obtained independent clones, 68% were recombinant. After their amplification, titre of the final cDNA library was more than 5.6×10^9. The size of the cloned cDNA was analysed by sequencing of the 50 clones from the library. There were 22% clones containing inserts which are between 500 bp and 1 Kb in size, and 16% having inserts longer than 1 Kb (Table 6).
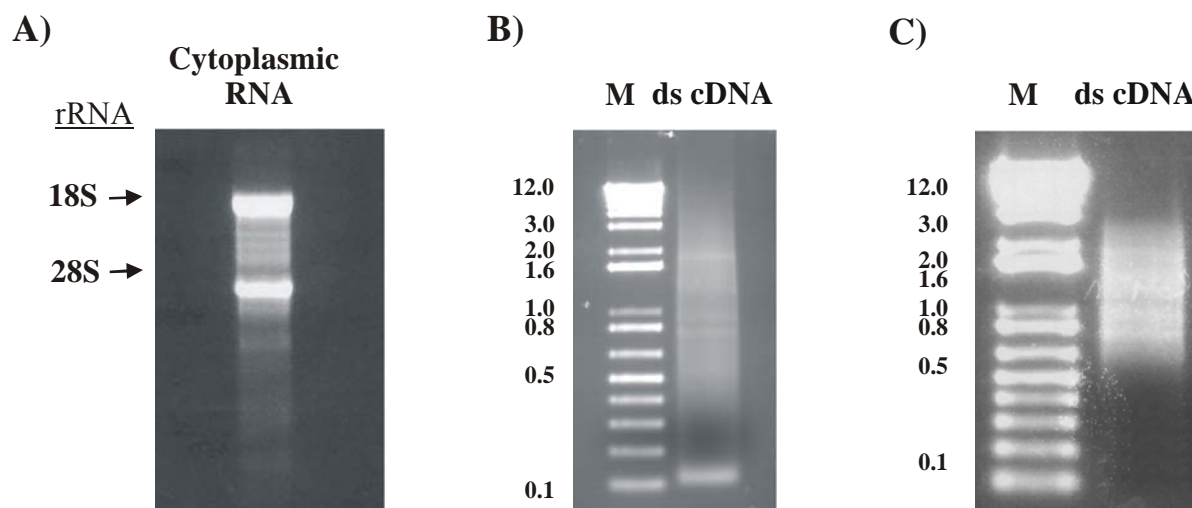
Fig. 15. Cytoplasmic RNA and synthesized ds cDNA.
A. Formaldehyde/agarose gel analysis of the 10 µg cytoplasmic RNA isolated from fresh human retina. RNA was visualised by staining with EtBr. Arrows indicate the positions of migration of the 28S and 18S rRNA. **B.** Analysis of the 5 µl sample of the ds cDNA amplified with 23 thermal cycles, which were determined as optimal, alongside of 1-Kb DNA size marker, on a 1% agarose/EtBr gel. For this purpose 6 µl (~0.4 µg) of the human retina poly(A)$^+$ was used as a starting material in a first-strand cDNA synthesis and 2 µl of the first strand cDNA served as a template for the second-strand synthesis. **C.** Analysis of the 5 µl sample of the size selected 850 bp – 3 Kb ds cDNA fraction on a 1% agarose/EtBr gel.

Table 6. Size distribution of isolated clones from the human retina full length cDNAs library

Fragment size (bp)	Number of clones	Percentage of clones characterized (%)
no insert	19	38
0 ~ 500	12	24
500 ~ 1000	11	22
1000 ~ 2000	6	12
> 2000	2	4
Total	50	100

To examine the representation of full-length cDNA clones in the human retina cDNA library, 17 clones were selected for further analysis. They were sequenced from both ends with λTriplEx5' and λTriplEx3' primers (Table 19, Appendix). The percentage of the clones having homology to a known open reading frame (ORF) of a gene for those with mRNAs <1 Kb, between 1 and 2 Kb, and >2 Kb were 100.0%, 100.0% and 33.3%, respectively (Table 7). The results show that the likelihood for a clone to obtain the 5' end was again determined by the size of the corresponding mRNA. For mRNAs of <1Kb and 1 - 2 Kb, the percentages of the inserts containing a translation initiating first ATG were 85.7 and 100.0, respectively, whereas for genes with mRNA longer then 2 Kb none of the clones reached initiation site of translation (Table 7).

Table 7. Coverage of the 5' end of known genes by human cDNA clones

mRNA size	Number of clones	CDs included (%)	5' end reached (%)[*]
<1 Kb	7	7 (100)	6 (85.7)
1-2 Kb	4	4 (100)	4 (100)
>2 Kb	6	2 (33.3)	0
Total	17	13 (76.5)	10 (58.8)

*Defined as the clone containing published initiation site of translation.

5.3 Cloning and characterization of retina specific genes

5.3.1 Cloning and functional characterization of three novel ABCC5 splicing variants

5.3.1.1 Identification of the full-length cDNA sequences

Sequence analysis of the clones from the human retinal cDNA library constructed by suppression subtraction hybridization identified two clones (6D2F and A06-Ret7) that derive from an identical locus on chromosome 3q27. The genome sequence comparison also revealed that these sequences map within intron 5 of the ABCC5 gene (Fig. 16A). Additional BLASTN homology searches with a consensus sequence assembled from the two clones led to the identification of AL524917, an mRNA which contains the first five known 5' exons of ABCC5 plus two exons located in the fifth intron of the reference sequence for ABCC5. Clone BF954562 which contains two exons of ABCC5 plus one exon also mapped to the same intron (Fig. 16A). Since the AL524917 sequence terminates 119 bp upstream from clone 6D2F it suggests that the retina sequences could derive from a novel ABCC5 variant. To investigate this, retina cDNA was amplified using primers L25F2 and L25R which anneal to the first exon of ABCC5 and to the novel retina sequence within intron 5, respectively (Fig. 16A). A 1700 bp long PCR product could be amplified (Fig. 17A).

To obtain the full-length sequence of the gene, a radiolabeled 364 bp PCR product was amplified with primers L25F and L25R and used to screen the human retina full-length cDNA library (local library ID: Jelena), which was cloned in the λTrplEx2 vector. After two rounds of screening, three positive clones (L25_1/2, L25_5/1 and L25_6/2) were isolated from approximately 1×10^6 recombinant phages. Insert clones ranging from 700 to 850 bp were partially or completely sequenced. All clones overlapped the novel retina transcript (Fig. 16A). In addition, clone L25_6/2 revealed a potential polyadenilation signal (AATAAA) located 21 bp upstream from the start of the poly (A) tail and a 30 bp long poly (A) tail.

To further analyze the novel splice variants, retina cDNA was amplified with primer L25F1, which anneals to the fifth exon of ABCC5 and L25rR3 which was designed from the novel exon sequences located in intron 5. The PCR reaction not only amplified the expected 203 and 275 bp products corresponding to BF954562 and AL524917, respectively, but also a novel 353 bp product (Fig. 16B). The latter amplicon resulted from the use of an alternative splice acceptor site generating an exon termed 5a. The shorter exon isoform is referred to as exon 5b whereas the last novel exon was named exon 5c.

To determine whether the alternatively spliced exons may also constitute additional internal exons of the ABCC5 transcript, PCR amplification with primers F1 and R1, which

binds to the sixth exon of the ABCC5, was carried out. Only a 267 bp product was amplified, indicating that there are no additional exons as part of the full-length ABCC5 transcript and that the novel exons encode an alternative 3' end of the gene (Fig. 16B). In agreement with this is the cDNA sequence BC050744, made public at a later time point. This cDNA contains the complete 5c exon and a poly (A) tail (Fig.16A).
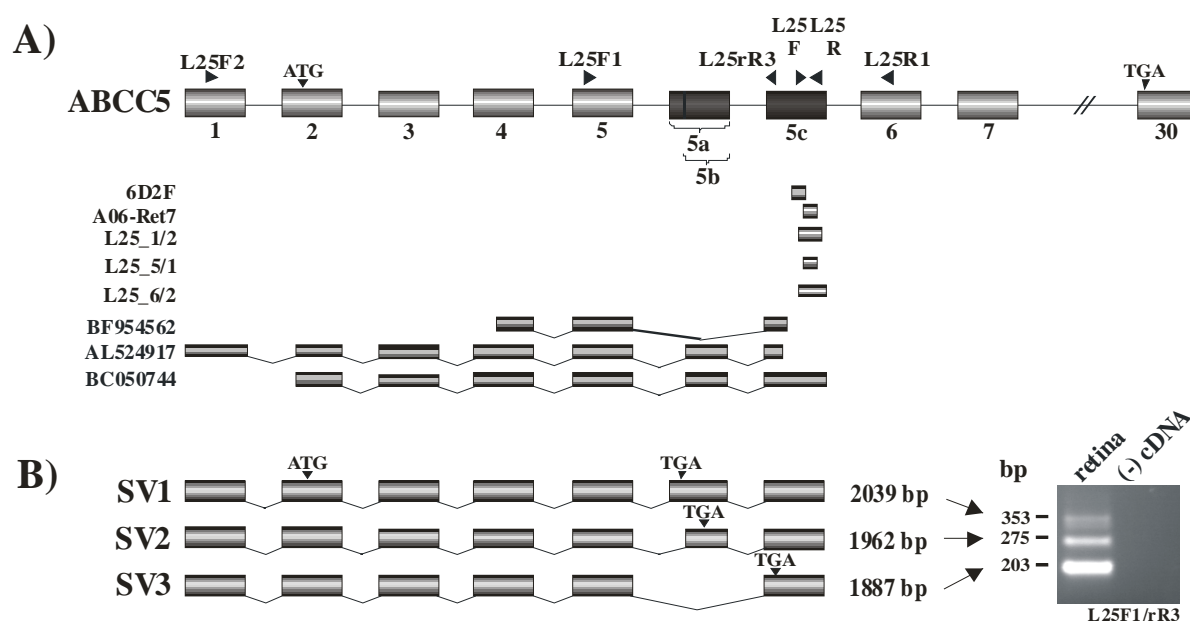
Fig. 16. Genomic organization of the ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3 isoforms.
A. Identification of the novel exons 5a, 5b and 5c (represented by the shaded boxes), located between exons 5 and 6 of the ABCC5 transporter, was based on the analysis of the 6D2F and A06-Ret7 clone sequences from the human retina cDNA library constructed by suppression subtraction technique, those isolated from the human retinal full-length cDNA library (local library ID: Jelena) (L25_1/2, L25_5/1 and L25_6/3) and known human cDNAs (Acc. No. BF954562, AL524917 and BC050744). Exons and introns are not represented in proportional sizes. The names and positions of the primers used in this study are shown above the schematic representation of the ABCC5 gene. **B.** Sequences of the ABCC5_SV1 (SV1), ABCC5_SV2 (SV2) and ABCC5_SV3 (SV3) isoforms which were assembled by alignment of the clone sequences and three PCR products (amplified with L25F1 and L25rR3 primers) to the corresponding human draft sequence. Calculated full-length sequences of the isoforms and relative locations of stop codons (TGA) are indicated.

5.3.1.2 Genomic structure

Three novel alternative exons of ABCC5 transporter named 5a, 5b and 5c and comprise 152 bp, 75 bp and 1.142 bp, respectively. They have splice donor and acceptor sequences that correspond to the consensus splice junction sequences (Table 8) and conform to the GT-AG rule (Burset et al, 2000). The score for the donor splice site is 0,803 for both 5a and 5b exons, and scores for the acceptor splice sites for 5a, 5b and 5c exons are 4.07, 10.666 and 5.16 respectively, which is within the expected range for true donor (0-11) and acceptor site (0-20) (spreadsheet compiled by Dr. Christian Sauer, 2001; Institute for Human Genetics, Wuerzburg, Germany).

In conclusion, analyses resulted in the identification of three splicing variants of ABCC5 gene, named ABCC5_SV1 (GeneBank accession number AY754874), ABCC5_SV2 (GeneBank accession number AY754875) and ABCC5_SV3 (GeneBank accession number AY754876). All three splicing variants share the first five exons with ABCC5, but contain different 3' ends derived from alternative splicing of exons located in intron 5 of ABCC5. As a result their mRNAs are 2039 bp, 1962 bp, and 1887 bp long, respectively.

5.3.1.3 In-silico analysis of cDNA sequences ABCC5_SV1, ABCC5_SV2, ABCC5_SV3

Analysis of the open reading frames of the novel isoforms led to the discovery that the stop codon of SV3 is located in the last exon, whereas those of SV1 and SV2 are located upstream of the last exon-intron junction. The SV2 termination codon is located only 39 bp upstream of the last exon/exon junction, but the SV1 stop codon is found more than 100 bp upstream of the last exon/exon junction (Fig. 16B).

Table 8.
Exon-intron boundaries of the human ABCC5 gene

Exon		3' Splice acceptor site[a]		5' Splice donor site[a]		Intron	
No.	Size (bp)	Sequence	Score[b]	Sequence	Score[b]	No.	Size (bp)
1	60			..CAG**gt**aggt..	1.894	1	3.382
2	184	..gaatttttaatttc**ag**G..	6.629	..CCG**gt**cagt..	4.773	2	28.880
3	158	..ctatctgctggaac**ag**T..	10.655	..CAA**gt**aaga..	3.140	3	498
4	156	..ggttttctctcctt**ag**A..	4.766	..AAG**gt**aggc..	3.041	4	654
5	148	..gtgtgcccttgttaa**g**A..	10.527	..CCA**gt**aagt..	3.517	5	4.762
5a	152	..ctctctccacctgc**ag**C..	4.070	..CAG**gt**aagg..	0.803	5	409
5b	75	..agaagctgttcttt**ag**A..	10.666	..CAG**gt**aagg..	0.803		409
5c	1.142	..cctgctgtcctttc**ag**C..	5.160				
6	234	..gtgtctctccctgt**ag**G..	4.096	..GAG**gt**gagt..	0.944	6	149
7	174	..tctttgtatcttata**g**C..	6.558	..ATG**gt**gagt..	1.783	7	576
8	148	..tcgcttttcctttc**ag**A..	2.577	..AAA**gt**gagt..	2.384	8	3.075
9	149	..tgctttttattcata**g**A..	5.030	..CAG**gt**cagg..	4.142	9	878
10	108	..cttctcctattttc**ag**G..	2.779	..AAG**gt**aagt..	0.221	10	5.597
11	357	..tcacaactcctatc**ag**A..	9.244	..GAG**gt**aagt..	0.865	11	3.796
12	72	..tgtgtgccttgtcc**ag**G..	6.334	..CAG**gt**aaga..	1.057	12	2.193
13	125	..gtgtctctctccct**ag**A..	4.571	..AAG**gc**aagg..	6.406	13	119
14	73	..tctttgtgtcctcc**ag**A..	3.885	..GAG**gt**acag..	6.034	14	1.596
15	204	..ttctgggctcttgt**ag**A..	7.324	..CAG**gt**atgg..	3.131	15	1.730
16	144	..ctgcctttgccttc**ag**T..	4.647	..GAG**gt**aaga..	1.921	16	1.675
17	103	..ttttaaatattatta**g**A..	11.973	..AAG**gt**aatg..	3.913	17	6.462
18	185	..tttttttcttacct**ag**G..	3.505	..GGG**gt**aagg..	2.994	18	1.082
19	147	..ctttggcttcaaac**ag**A..	8.116	..AAG**gt**atgc..	3.474	19	286
20	130	..gtgctctgctcctc**ag**G..	4.412	..AAG**gt**attt..	5.437	20	1.315
21	154	..ttatttgtttccac**ag**T..	4.791	..CAG**gt**aaat..	2.459	21	90
22	129	..ctcattctctctcc**ag**G..	2.016	..CAG**gt**tttgt..	5.249	22	2.242
23	187	..gttgtctttctcac**ag**A..	3.740	..CAG**gt**cagt..	3.339	23	1.384
24	90	..tgttttccattatc**ag**T..	6.957	..AAG**gt**caga..	4.616	24	2.933
25	190	..acccttgtctcccc**ag**A..	3.950	..CAG**gt**aagg..	0.803	25	4.666
26	160	..tcttctgactcatc**ag**G..	6.244	..CAG**gt**gagg..	0.883	26	9.075
27	79	..ctgtgtttctccct**ag**A..	4.674	..TGT**gt**aagt..	4.885	27	1.303
28	114	..gtttctttctccct**ag**A..	3.357	..AAG**gt**aagg..	1.025	28	1.610
29	165	..tctgcttctctggc**ag**A..	4.977	..CAG**gt**accg..	5.517	29	4.153
30	1.468	..tctccttcttttgc**ag**G..	1.220				

[a] For the splice junctions, intronic and exonic sequences are shown in small and capital letters, respectively. The 5' splice-donor and 3' splice-acceptor sequences are given in bold.
[b] Scores were calculated to references (Berg et al, 1988 and Penotti et al. 1991). Approximately 99% of sites have a score of 0-20 (3'acceptor site) and 0-11 (5'donor site).

```
bp                                                                    aa
                    Exon 1
1        CCGGGCAGGTGGCTCATGCTCGGGAGCGTGGTTGAGCGGCTGGCGCGGTTGTCCTGGAGC

                         Exon 2
61       AGGGGCGCAGGAATTCTGATGTGAAACTAACAGTCTGTGAGCCCTGGAACCTCCACTCAG
121      AGAAGATGAAGGATATCGACATAGGAAAAGAGTATATCATCCCCAGTCCTGGGTATAGAA
              M  K  D  I  D  I  G  K  E  Y  I  I  P  S  P  G  Y  R      18
181      GTGTGAGGGAGAGAACCAGCACTTCTGGGACGCACAGAGACCGTGAAGATTCCAAGTTCA
          S  V  R  E  R  T  S  T  S  G  T  H  R  D  R  E  D  S  K  F    38

                         Exon 3
241      GGAGAACTCGACCGTTGGAATGCCAAGATGCCTTGGAAACAGCAGCCCGAGCCGAGGGCC
          R  R  T  R  P  L  E  C  Q  D  A  L  E  T  A  A  R  A  E  G    58
301      TCTCTCTTGATGCCTCCATGCATTCTCAGCTCAGAATCCTGGATGAGGAGCATCCCAAGG
          L  S  L  D  A  S  M  H  S  Q  L  R  I  L  D  E  E  H  P  K    78

                                                    Exon 4
361      GAAAGTACCATCATGGCTTGAGTGCTCTGAAGCCCATCCGGACTACTTCCAAACACCAGC
          G  K  Y  H  H  G  L  S  A  L  K  P  I  R  T  T  S  K  H  Q    98
421      ACCCAGTGGACAATGCTGGGCTTTTTTCCTGTATGACTTTTTCGTGGCTTTCTTCTCTGG
          H  P  V  D  N  A  G  L  F  S  C  M  T  F  S  W  L  S  S  L    118
481      CCCGTGTGGCCCACAAGAAGGGGGAGCTCTCAATGGAAGACGTGTGGTCTCTGTCCAAGC
          A  R  V  A  H  K  K  G  E  L  S  M  E  D  V  W  S  L  S  K    138

                              Exon 5
541      ACGAGTCTTCTGACGTGAACTGCAGAAGACTAGAGAGACTGTGGCAAGAAGAGCTGAATG
          H  E  S  S  D  V  N  C  R  R  L  E  R  L  W  Q  E  E  L  N    158
601      AAGTTGGGCCAGACGCTGCTTCCCTGCGAAGGGTTGTGTGGATCTTCTGCCGCACCAGGC
          E  V  G  P  D  A  A  S  L  R  R  V  V  W  I  F  C  R  T  R    178

                                                        Exon 5a
661      TCATCCTGTCCATCGTGTGCCTGATGATCACGCAGCTGGCTGGCTTCAGTGGACCACTTG
          L  I  L  S  I  V  C  L  M  I  T  Q  L  A  G  F  S  G  P  L    198
721      CTTGGTGCTGTCAGGATCTGGATTTGGGGGGAGTCTCTCTGTAGTGGAACCAGTGACAGA
          A  W  C  C  Q  D  L  D  L  G  G  V  S  L  -                  212

                    Exon 5b
781      AGCTGTTCTTTAGAATTTTTCAGGATGGCTGTATTCTGCGGTCAGAATGAGAGAGTCAAGC
                  N  F  Q  D  G  C  I  L  R  S  E  -                   208

                    Exon 5c
841      TGGGCAGAATCTCTCGCCAAGAGTTCAGCCTTCCTTTGGAGACTGCTCCATCAGTGCCGA
                              P  S  F  G  D  C  S  I  S  A  E          205
901      GGTGTGTGGGAACAGGCTTCACTGCACCGCCATCTTACTGAGTTGCTTCACGTGAGGAAA
                  V  C  G  N  R  L  H  C  T  A  I  L  L  S  C  F  T  -  225
961      AGGGGGCTTTGGCCCTGTGACTCAGTTCCACATTTTGGATTGCATACTGGAAAAGAAGCC
1021     AATCTTCTTGCTAGTAAACCAGCAACCCGGCTGTATACAGTGGTGACCCAAGCAATGGAT
1081     ATAAACCTAAAAATCTGAGGGAGGGGAGAGGTGGAATACAGTAGTTCTTGGAATCTGAAG
1141     TCTCCTATTTGATCAGGTTATTTCCTGGGACTTGGCAAAAATCTGATTGGTGGGGATCTC
1201     CTAGGACCTAGTGGACATCTGGTATTAATTTAATCTCAGGAAAAACAAGAAATTAACCCA
1261     GAGAGAGTCTGGGTTTTGGAATTCAGCGTAGCTACCTCCAGACCGTGGTGTCTGGCCTCC
1321     ATTTTTGTCTGTCATTCAGCTCTGACTTACAGCTGCAGTCACCTTTGCTATAAGGCACCT
1381     GGGTAGAAGGGTGGATGGGCTTCACATCAATTTTTTTCTTCCTTTAGGGTGGGGGATTGG
1441     TTTGGCTTTCTTTTGTTGTGGTTTTTTGTTTTATTTTTGTCAAGATTGATTTTTAGATGC
1501     AAGGACTTGAAAAGACCCAGAAGGATGCCACCAGTTTTTCCTTGAGGCCTAGGATTTTTT
1561     ATTCTGTCCCGAGCAGAGGTAATTCCTCACAACTTAGTGCACCAGTAGCACCAGCCATTT
1621     TGAGCAGAGTACCTCTTTGGGGAGCTTTTCGTTTTGTTTTGTTTTTAATTCTCTTTCCTT
1681     AGCAGCAAGGTCTTTTTTCCTAGAGAATCTACTCCGTTGCAGAATCATTGCAACCTCAGG
1741     AGCCCTCACTGATTGAGTGCTGTCAGCCTGATATACTACTTTGGACTCTGGAAACAGATA
1801     TGGGTTCTATTCTCTATTTCTACTGTGTGTCGTTAAACAACCGTCGGAGACCAGATGACC
1861     TGTTAGATGGCTAGTCCTGTATAACTCGACTCTGTATGTTTCAATGTATGTTACTGCAAT
1921     GCTTCACCTGCTGTACAGTGTTTGTGAGATGCTCTTTGAAGATGGTACTTTTATATTTTT
1981     TCATTTTCAATAAAAGTACATTCCTTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 17. Full-length cDNA sequence and putative ORFs of ABCC5 splicing variants SV1, SV2 and SV3. Nucleotide sequence is numbered starting with the first base of the ABCC5 cDNA. The corresponding amino acid sequence and amino acid positions are indicated bellow the nucleotide sequence. All three splicing variants contain common 197 amino acid stretch (grey highlighted), whereas 15 aa (yellow background), 11 aa (pink highlighted) and 28 aa (green highlighted) are encoded specifically by SV1, SV2 and SV3 transcripts. Start codon, stop codons (located in exon 5a, 5b and 5c in SV1, SV2 and SV3, respectively) and putative polyadenilation signal (AATAAA) are boxed.

5.3.1.4 Expression analysis

RT-PCR amplification was done with primers L25F2 and L25R, which are amplifying SV1, SV2 and SV3 splicing variants, in a panel of 18 human tissues and showed that all three isoforms are highly expressed in retina with lower expression level in brain, fetal brain, spinal cord, bone marrow, RPE, spleen, skeletal muscle, testis and prostate (Fig. 18A). The reported ubiquitous expression of the ABCC5 transporter determined by Northern blot analysis (McAleer et al., 1999) was confirmed in a PCR with primers F1 and R1 (Fig. 18A). However, high levels of expression of ABCC5 transcript were identified in retina and RPE.

The relative expression of each isoform was determined by qRT-PCR in 10 tissue panel. Analysis confirmed abundant expression in retina for all three splicing variants (Fig. 18B). The level of expression of the SV1 transcript in retina is 2.5-fold higher than in cerebellum and approximately 4-fold higher compared to other tissues. The SV2 isoform is expressed in retina and RPE at almost the same level and approximately 50% less in bladder compared to its expression in retina, whereas expression in other tissues is much lower. Compared to retina, 3-fold weaker expression of the SV3 splicing variant was determined in cerebellum and RPE, whereas transcription was almost not detectable in the remaining tissues of the panel. Quantitative real-time PCR analysis of ABCC5 confirmed abundant expression of the transcript in retina, as shown already with RT-PCR, with approximately 2-fold lower expression in cerebellum, occipital cortex (occ. cortex) and RPE and very weak expression obtained in other tissues.

To determine the size of the various splice variants of ABCC5 and to confirm the hypothesis that exon 5c is used as their internal exon, a radiolabeled PCR fragment was amplified from exon 5c and hybridised to a virtual Northern blot containing cDNAs from six human tissues (Fig. 18C). A strong band of approximately 2 Kb, which corresponds to the calculated length of the novel SV1, SV2 and SV3 isoforms of ABCC5, was detected exclusively in retina.

RT-PCR with primer pair L25F1 and L25rR3, which is amplifying all three splicing variants, demonstrated that they are expressed at different levels in retina (SV3>SV2>SV1, Fig. 16B).
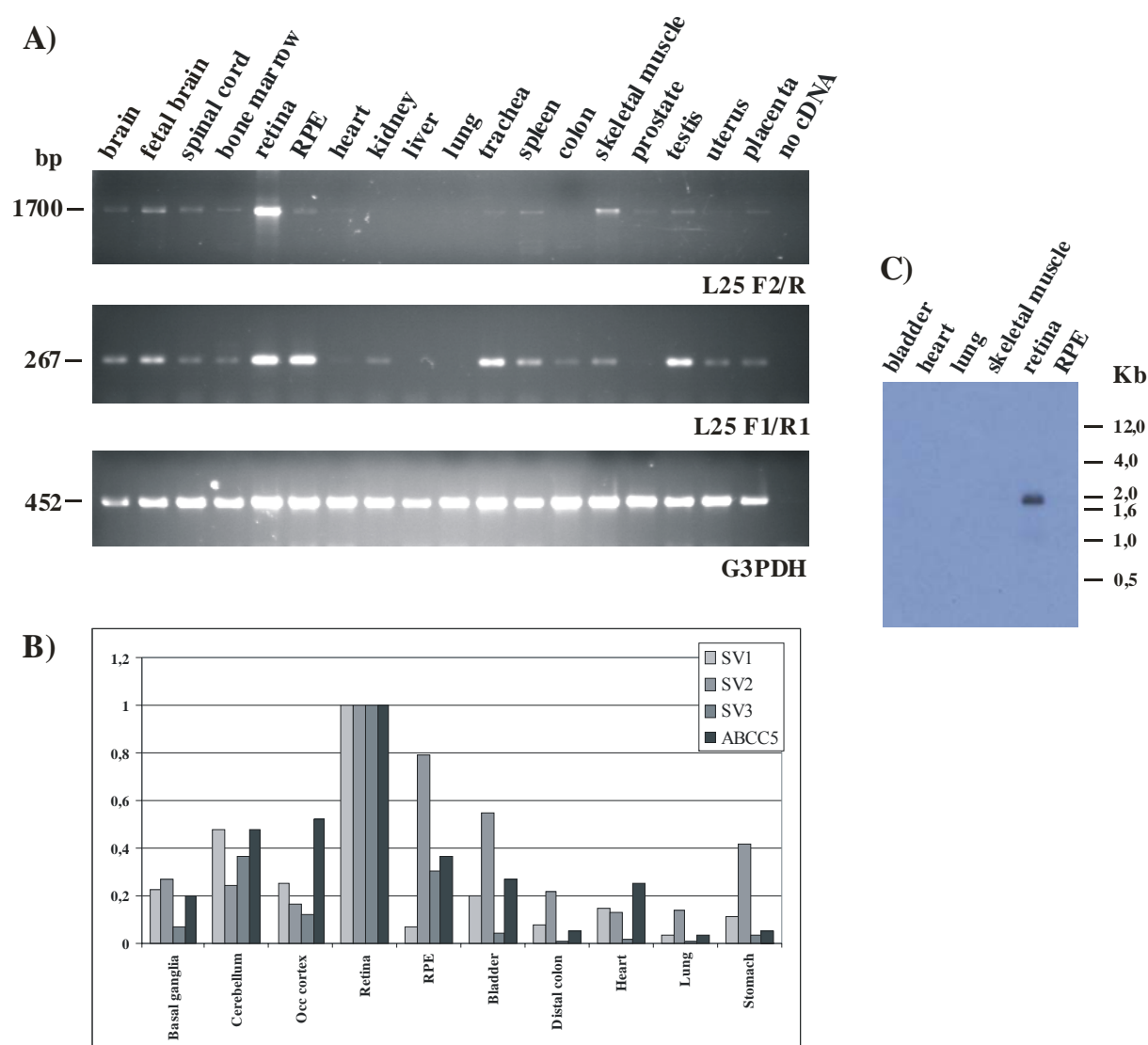
Fig. 18. Expression analysis of the ABCC5 isoforms.

A. Expression analysis of ABCC5 wild type and SV1, SV2 and SV3 alternatively spliced transcript in 18 human tissues done by RT-PCR. Primer pair L25F2 and L25R, amplifying all three splicing variants, showed that they are abundantly expressed in retina and weaker in brain, fetal brain, spinal cord, bone marrow, RPE, skeletal muscle and testis. Expression analysis performed with primers L25F1 and L25R1 determined that ABCC5 gene is ubiquitously expressed in human tissues with high abundance in retina and RPE. The same panel was amplified with G3PDH specific primers as a control of cDNA quality. B. Relative expression of ABCC5 isoforms done by qRT-PCR in panel of 10 human tissues. Expression analyses were performed with primers specific for each isoform. For each isoform expression data were normalized to four housekeeping genes (see Materials and Methods, 4.3.2.2) and to the highest expression of in the panel, which was set as 1. Analysis showed that all tested isoforms are abundantly expressed in retina. C. Transcript size of SV1, SV2 and SV3 was determined by virtual Northern blot analysis and signal at approximately 2.0 Kb was detected in retina.

5.3.1.5 Functional characterization of the splice variants

5.3.1.5.1 Nonsense mediated mRNA decay in isoform SV1 containing a premature termination signal

It has been shown that a stop codon found 50-55 bp upstream of the last exon-exon junction is recognized as premature termination codon (PTC) by the translational machinery. As a consequence, the transcript will undergo nonsense mediated mRNA decay (NMD) (Henze and Kulozik 1999). To investigate whether this mechanism plays a role in the degradation of the SV1 isoform, which fulfils this criterion, a protein synthesis inhibition method was used. It is known that the NMD machinery operates in a translation dependent manner which can be inactivated by blocking protein synthesis with a number of antibiotics (Maquat, 2004). Specifically, human cell lines ARPE-19 and Y79 were treated with either puromycin or anisomycin for two hours. Total RNA samples were extracted from untreated cells and those treated with antibiotic and RT-PCR was conducted to assess the level of expression for the various splicing variants. As expected for a transcript which normally undergoes NMD, treatment of the cell lines with protein synthesis inhibitors demonstrates clear enrichment of the SV1 transcript over those not undergoing NMD (such as SV2 and SV3) (Fig 19).
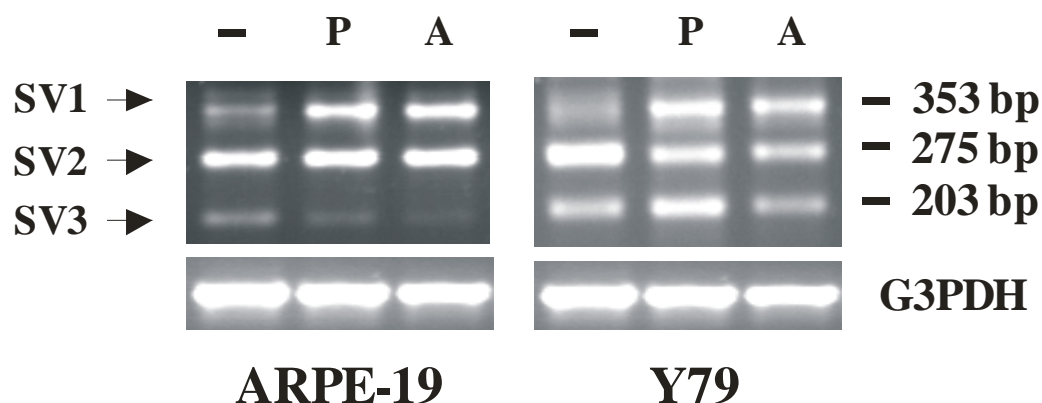
Fig. 19. Protein synthesis inhibition selectively increases PTC containing SV1 RNA. Antibiotic treatment was performed on ARPE-19 and Y79 cell lines. Total RNA was isolated from untreated cells (-) or those treated with 100 µg/ml puromycin (P) or 100 µg/ml anisomycin (A) for 2h. RT-PCR amplification was done with L25F1 and L25rR3 primers which are amplifying all three splicing variants.

To confirm the results obtained, again the cells were treated with antibiotics, the antibiotic was then removed by washing with 1 X PBS and half of the cells were incubated in drug-free medium for an additional 4h (Lamba et al., 2003). In order to generate consistent and reliable results, all experiments were done in triplicates and the amount of all three splicing variants was determined by qRT-PCR. The amount of SV1 drastically increased as a result of protein synthesis inhibition but returned to previous levels after protein synthesis was restored, whereas no changes in the amount of SV2 and SV3 transcripts were detected (Fig. 20). In particular, puromycin treatment increased the relative amount of SV1 mRNA almost 3-fold in the ARPE-19 cell line and more than 2-fold in the Y79 cell line, whereas the amount of wild type ABCC5 mRNA decreased by a factor of two in both cell lines at comparable time points (Fig. 20). The initial transcript concentrations were restored after washing.

In the cell lines treated with anisomycin, an average 4- and 2-fold increase of SV1 mRNA was determined in ARPE-19 and Y79, respectively, whereas ABCC5 mRNA levels decreased to half of the original amount. Even four hours after antibiotic removal, higher expression levels of SV1 and less quantity of ABCC5 mRNA were still detectable in the ARPE-19 cells. The recovery of protein synthesis after anisomycin treatment in the Y79 cell line seemed to be faster as the level of expression for both isoforms were slowly returning to normal (Fig. 20). Again, drastic changes of amounts of SV2 and SV3 mRNA were detected in the course of inactivation of the NMD machinery. This observation suggests that ABCC5 expression might be regulated by the transcription level of ABCC5_SV1.
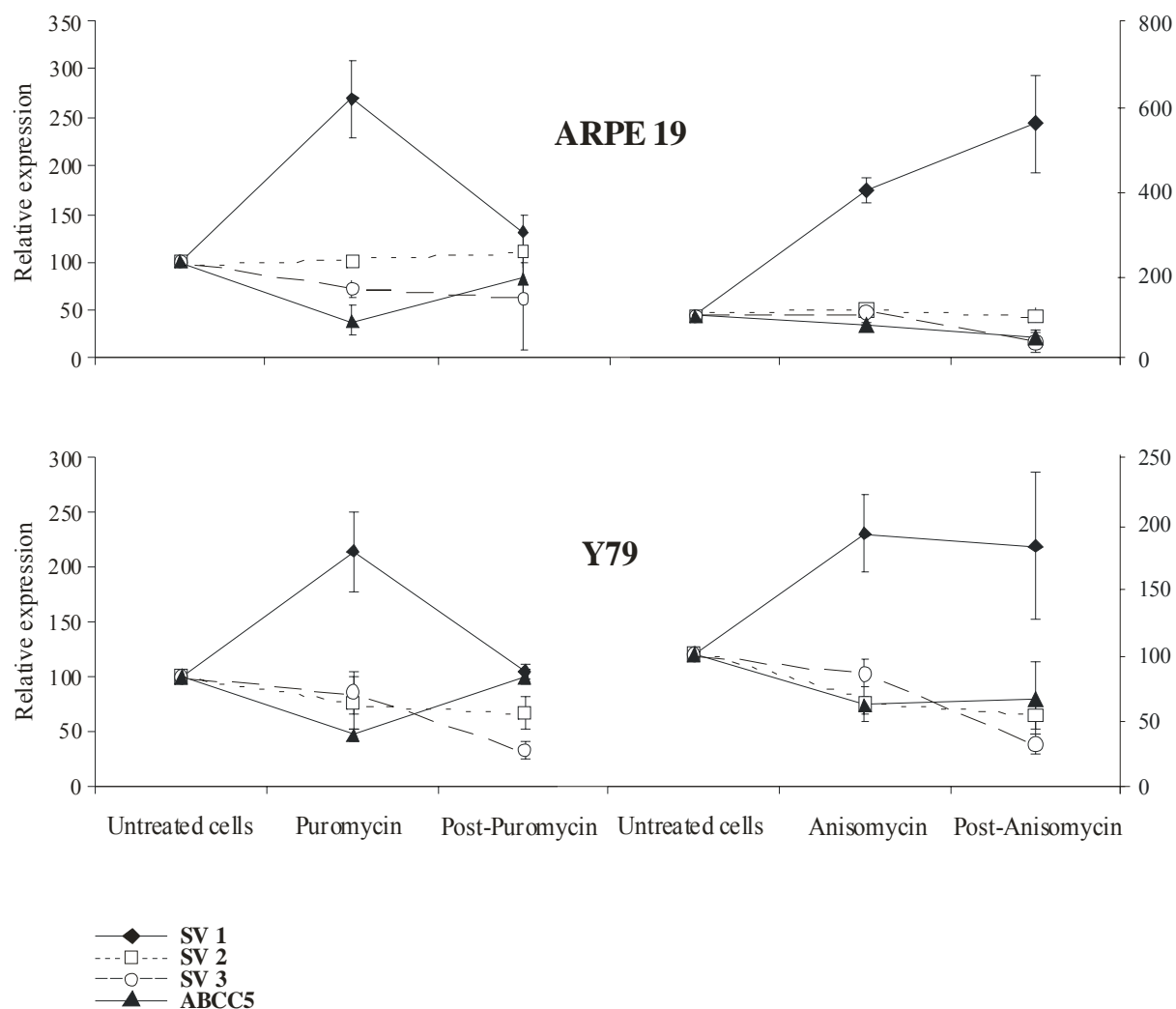
Fig. 20. Effects on inactivation of NMD on wild type and alternatively spliced ABCC5 transcripts.

Treatment of the human cell lines ARPE-19 and Y79 with inhibitors of protein synthesis puromycin or anisomycin resulted in stabilisation of PTC containing SV1 transcript, followed by decrease in amount of ABCC5 mRNA, two hours after the treatment with antibiotic. Expression of the transcripts were monitored by qRT-PCR analysis using samples of the total RNA which was isolated from the untreated cells, cells treated with antibiotic (puromycin or anisomycin) and from the cells 4h after the treatment (Post-Puromycin and Post-Anisomycin). Expression of all transcripts were normalised to their expression level in untreated cells which was set as 100%, and to four housekeeping genes (see Materials and Methods, 4.3.2.2). Error bars represent standard errors.

5.3.1.5.2 Silencing of SV1, SV2 and SV3 isoforms with siRNA duplexes

To investigate whether the expression of the ABCC5 isoforms is co-regulated by the transcription level of ABCC5_SV1 mRNA, three alternative isoforms were systematically silenced by RNA interference. First, the sequence of the siRNA.1 probe was designed to anneal to the 3' untranslated region common to all three isoforms as described in Materials and Methods (see siRNA design and sequences, 4.7.2). Silencing of target transcripts was examined in cultured ARPE-19 cells transfected with siRNA.1 duplexes and their effects were compared with mock-transfected cells (Fig. 21).

To optimize conditions for effective RNA silencing, the time course for siRNA-mediated reduction of SV1, SV2 and SV3 expression was determined. Cells were harvested at 8, 24, 48 and 72 hours after siRNA transfection in order to prepare RNA extracts. Gene silencing potential was monitored by qRT-PCR and analysis revealed that all three transcripts were effectively silenced 48h after transfection with siRNA.1, leading to approximately 80% decrease of expression in comparison to control cells (Fig. 21). The amount of SV3 remained low even 72h posttransfection time, whereas inhibition of SV1 and SV3 increased again at this time point. Obtained data indicate that siRNA.1 have a high degree of specificity which resulted in specific decrease of SV1, SV2 and SV3 expression without detected effects on the expression level of control housekeeping genes (data not shown).
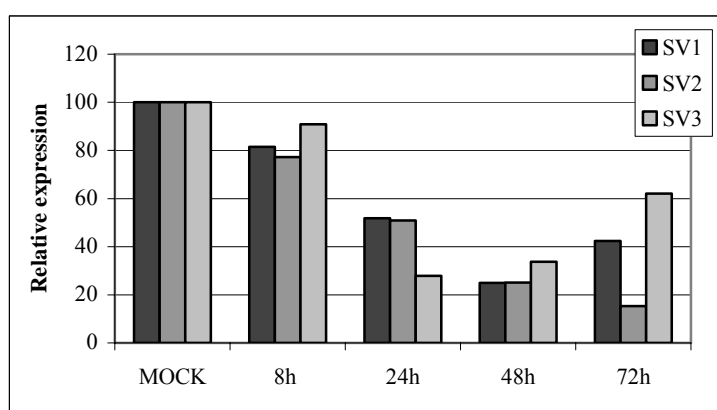
Fig. 21. Optimization of the conditions for effective RNA interference. Time course of reduction of expression of SV1, SV2 and SV3 transcripts was examined in ARPE-19 cells, which were transfected with siRNA.1 duplexes (2.0 µg/well) directed against all three alternatively spliced transcripts. Mock control was generated by addition of the buffer instead of the volume of siRNA used for the transfection of the culture. Extracts of RNA were prepared from each set of these cells at 8, 24, 48 and 72 hours posttransfection and analysed by real-time PCR using primers specific for each transcript. The amounts of SV1, SV2 and SV3 transcripts were calculated according to a model published by Vandesompele et al. (2002), by normalization to housekeeping genes B2M, TBP, SDHA and HPRTI (see Materials and Methods 4.3.2.2) and to mock-transfected cells (expression was set as 100).

After successful optimization of the conditions for specific knockdowns of SV1, SV2 and SV3 using siRNA.1 duplexes, experiments were repeated in triplicates in order to generate consistent results. The 80% decrease of SV1, SV2 and SV3 expression 48h after siRNA transfection was reproducible and it was followed by linear increase of ABCC5 which was observed during this same time period (Fig. 23A). In order to specify which of the three splicing variants may influence expression of the full-length transporter ABCC5, silencing experiments were performed with siRNA.2 duplexes directed selectively against SV1 and SV2 isoforms and with siRNA.3 designed to inhibit exclusively SV1. The most effective silencing of SV1 obtained with siRNA.2 duplexes was monitored 24h after transfection, whereas the most effective inhibition of SV2 mRNA was monitored 48h after transfection. This knockdown of both, SV1 and SV2 led to increase of ABCC5 expression to a level of approximately 50% compared to control cells (Fig. 23B). Incubation of the ARPE-19 cells with siRNA.3 resulted in 50% decrease of SV1 expression one day after transfection, followed by a significant increase of ABCC5 (Fig. 23C). As expected, siRNA.2 did not alter the amount of SV3 mRNA. In contrast, the reason for the significant decrease of SV2 and SV3 transcripts in cells transfected with siRNA.3 is not clear.

A)

B)

C)
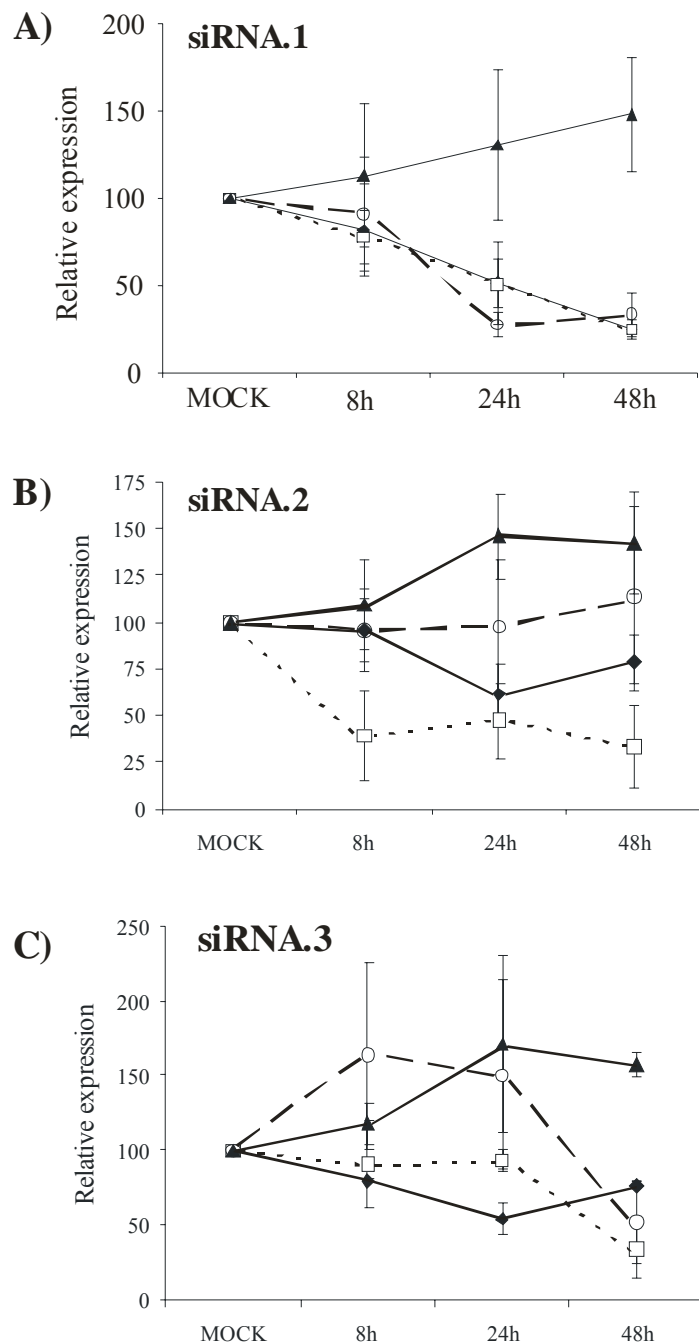
Fig. 22. Silencing of SV1, SV2 and SV3 transcripts. A. The ARPE-19 cell line was transfected with siRNA.1 duplexes (2.0 µg/well) directed against all three alternatively spliced transcripts. B. The ARPE-19 cell line was transfected with siRNA.2 duplexes (4.0 µg/well) designed to silence SV1 and SV2, selectively. C. The ARPE-19 cell line was transfected with siRNA.3 (8.0 µg/well) directed selectively against SV1. Extracts of RNA were prepared from the cells at 8, 24 and 48 hours after transfection and analysed by real-time PCR using primers specific for SV1, SV2 and SV3 transcripts. The amount of each transcript was normalization to housekeeping genes B2M, TBP, SDHA and HPRTI and to mock-transfected cells (calculations were done according to a model published by Vandesompele et al. (2002)). Mock control was generated by transfection of the culture with non-silencing siRNA duplexes (Qiagen) and expression of the transcript obtained in these cells was set as 100. Data are average of three independent experiments and error bars represent standard errors.

5.3.1.5.3 Characterization of truncated proteins

Above, we have shown that SV1 mRNA is degraded by a NMD surveillance mechanism. Transcripts of SV2 and SV3 have stop codons that are 3' of the last splice site and according to the rules of NMD (Henze and Kulozik, 1999), these transcripts should not be targets of NMD. Their mRNAs should be stable and serve as a template for the translation of C-terminally truncated ABCC5 proteins. The length of the putative proteins should be 208 aa and 225 aa, respectively, each sharing the first 197 aa with ABCC5 (Fig. 17). In-silico analysis of the remaining 11 aa encoded by SV1 and 28 aa encoded by SV3 (Fig. 17) did not reveal any sequence homologies to known protein domains, nor did it yield clues as to their putative functions.

It has been reported that similar to other ABC transporters, human ABCC5 contains Walker A, B and C motifs, two nucleotide binding domains (NBD1 and NBD2) and twelve transmembrane spanning helices in two hydrophobic domains (TMD1 and TMD2) (Kool et al., 1999). TMD1 is encoded by exons 5-8; each of the NBDs in encoded by five exons, 11-15 for NBD1 and 25-29 for NBD2, encode each of the NBDs; the TMD2 is encoded by exons 17-23 (Fig. 23). Both TMDs, each have six transmembrane helixes. Based on the membrane protein topology prediction method TMHMM, which is available at http://www.cbs.dtu.dk/services/TMHMM-2.0/, it was determined that the first transmembrane helix of TMD1 corresponds to aa 179-201. This amino acid stretch is almost completely contained in both putative truncated proteins encoded by ABCC5_SV2 and ABCC5_SV3 isoforms. Hydropathy analysis of putative membrane spanning domains by TMHMM did not predict transmembrane domains, suggesting that the two isoforms may be soluble proteins.
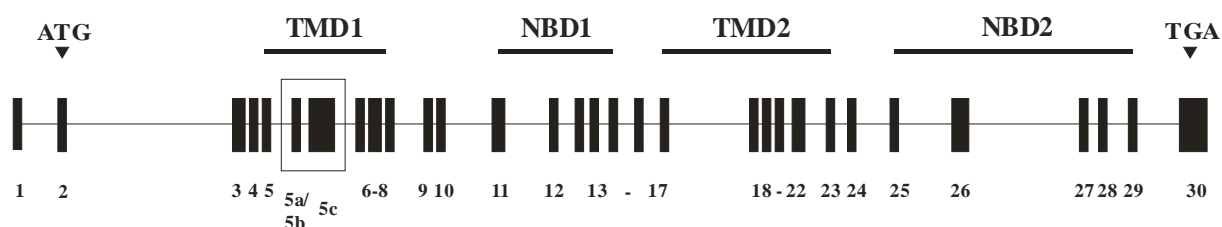
Fig. 23. Genomic organization of the human ABCC5 gene. Exons coding transmembrane domains (TMDs) and nucleotide binding domains (NBDs) of ABCC5 have been underlined. The ABCC5 has 30 exons and inclusion of exons 5a, 5b and/or 5c (exons are boxed) introduce alternative stop codons and activate the NMD machinery (inclusion of exon 5a), or generation of putative C-terminally truncated ABCC5 proteins (transcripts containing 5b and/or 5c exons). Exons and introns are drown in proportional sizes and start and stop codons of ABCC5 are represented with arrows in 2nd and last 30th exon, respectively.

5.3.1.5.4 Polyclonal antibody production

Immunofluorescence analysis of ABCC5 using antibodies raised against its N-terminus (82-168 amino acids) revealed predominant plasma membrane localization with some intracellular staining (Wijnholds et al., 2000). The latter staining may indicate the truncated proteins encoded by SV2 and SV3 transcripts. In order to demonstrate the existence of those two C-terminally truncated ABCC5 proteins, generation of polyclonal antibodies against their unique C-termini was initiated.

To obtain a polyclonal antiserum, mice were immunized with synthetic GST-fusion protein constructs containing the antigenic regions specific for each putative protein. The epitopes selected for generation of polyclonal antibodies were amino acids 198 to 208 for the SV2 protein (5'-NFQDGCILRSE-3') and amino acids 198 to 225 (5'-PSFGDCSISAEVCGNRLHCTAILLSCFT-3') for the SV3 protein (Fig. 17). In order to obtain synthetic constructs, cDNA fragments corresponding to selected amino acid stretches were amplified by RT-PCR using primer pairs L25SV2-EcoRI-F and L25SV2-XhoI-R specific for SV2 and L25SV3-EcoRI-F and L25SV3-XhoI-R1 specific for SV3 (Appendix, Table 21). Obtained products were cloned in frame into the multiple cloning site of the pGEX-4T-3 vector (see Materials and Methods, DNA methods, 3.4.3 and 3.4.4) and expressed in BL21 cells (see Materials and Methods, 4.1 polyclonal antibody production). Different amounts of affinity purified GST-fusion proteins were used for immunization of six Balb/c mice (see Materials and Methods, .2). Mice were boosted twice with corresponding GST-fusion proteins.

The specificity of polyclonal mouse serum was tested by Western blot analysis using the MBP-fusion proteins containing the corresponding SV2 and SV3 peptides (for generation of the constructs see Materials and Methods, 4.4 polyclonal antibody production). Results showed that antisera failed to detect any of the MBP-fusion proteins (data not shown). In addition, the 1D4 tag was inserted in frame in the C terminus of the full-length SV2 and SV3 proteins (see Materials and Methods, 4.4), and the resulting recombinant gene was transfected into the 293-EBNA cell line. Western blot analysis with cellular extracts containing both overexpressed fusion proteins revealed no positive response. As a control, labelling was observed with monoclonal anti Rho-1D4 antibodies recognized corresponding fusion protein of approximately 25 kDa (data not shown). In addition, immunoractivity was not detected in the preimmune serum (data not shown).

5.3.1.6 Conservation of exons 5a, 5b and 5c of isoforms SV1, SV2 and SV3

Amino acid sequence of the human ABCC5 (MRP5) and mouse Abcc5 (mrp5) is highly conserved (94.1%) (Suzuki et al., 2000). In order to analyse whether the sequences of novel exons 5a, 5b and 5c are evolutionary conserved among mammalian species, nucleotide homology searches with human cDNA sequences of those three exons were done and revealed high level of identity to bovine, porcine and mouse ESTs for exon 5b (Fig. 24). Homology searches failed to show significant evolutionary conservation for exons 5a and 5c.

```
                                                     *
Human   1 AATTTTCAGGATGGCTGTATTCTGCGGTCAGAATGAGAGAGTCAAGCTGGGCAGAATCTCTCGCCAAGAGTTCAG
Bovine  1 AATTTTCAGGATGGCTGTATTCTGCGGTCAGAATGAGAGAGTCAAGCTGGGCAGAATCTCTCGCCAAGAGTTCAG
Porcine 1 AATTTTCAGGATGGCTGTATTCTGCGGTCAGAATGAGAGAGTCAAGCTGGGCAGAATCTCTCGCCAAGAGTTCAG
Mouse   1 AATTTTCAGGATGGCTGTATTCTGCGGTCAGAATGAGAGAGTCA-GCTGGGCAGAATCTCTCGCCAAGAGCTCAG
```

Fig. 24. Multiple sequence alignment of the nucleotide sequences of the 5b exon.
Human exon sequence is aligned with bovine, porcine and mouse EST sequences using ClustalW program. The conserved sequence is shaded with BoxShade. Species and EST accession numbers are: bovine, Bos taurus (GeneBank Acc.No. BI848281); porcine, Sus scrofa (GeneBank Acc.No. BI345271) and mouse, Mus musculus (GeneBank Acc.No. AW124768).

To investigate possible splicing events in mouse, RT-PCR was performed using primers L25F1, located in exon 5 of the murine Abcc5 and L25R4_Mus, primer sequence designed based on mouse AW124768 cDNA which sequence is almost identical to the human 5b exon. Amplification resulted in two PCR products ubiquitously expressed in various mouse tissues (Fig. 25B). After sequencing their nucleotide sequences were aligned to the mouse genomic DNA, confirming an orthologous exon 5b in the murine AW124768 cDNA. Analysis of the nucleotide sequences showed that exon 5a is not conserved in mouse, whereas difference in length of two PCR products derives from usage of alternative splice acceptor site in an exon corresponding to human 5b. PCR amplification using L25F1 primer that binds to the fifth exon of the Abcc5 and L25R1 primer that binds to the sixth exon of the Abcc5 resulted in only a single PCR product, suggesting that there are also no additional exons between exon five and six in the full-length transcript of the murine Abcc5 (Fig. 25A).

Analysis of the nucleotide sequences of the PCR products amplified with L25F1 and L25R4_Mus suggest presence of two different splicing variants of the Abcc5 gene, both homologues to ABCC5_SV2 transcript, having a stop codon in an exon that is orthologuous to human 5b (Fig. 25C).
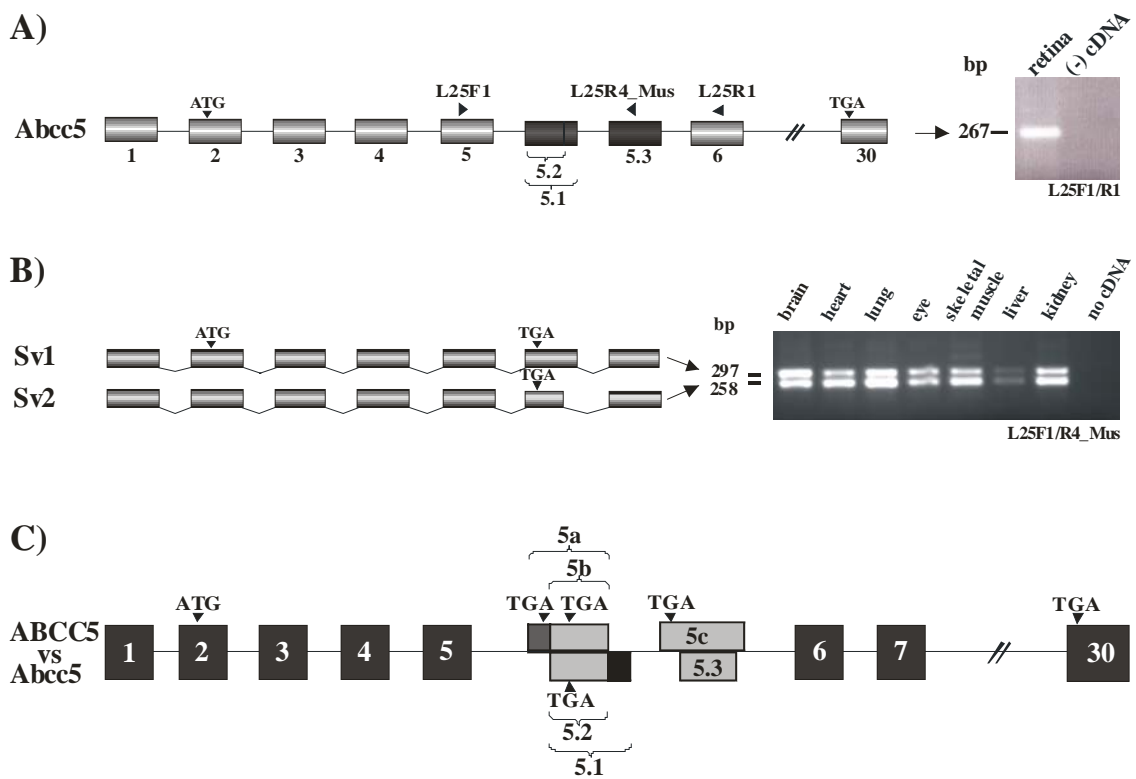
Fig. 25. Identification of novel splicing isoforms of murine Abcc5 and their structural comparison to human homologue sequences. A. RT-PCR with primers L25F1 and L25R1 that anneal to the fifth and sixth exon of Abcc5, respectively, showed that there are no additional exons between those two exons in the transcript of the mouse transporter. B. RT-PCR performed with forward L25F1 primer, which anneals to the fifth exon of Abcc5 and reverse L25R4_Mus primer, which anneals to the exon 5.3 of Abcc5. Two different PCR products, which were 297 bp and 258 bp long, were amplified from mouse brain, heart, lung, eye, skeletal muscle, liver and kidney cDNAs showing ubiquitous expression of the transcripts among those tissues. C. Schematic comparison of human ABCC5 and murine Abcc5 showing differential splicing affecting the fifth intron of this gene in both species.

5.3.2 Cloning and characterization of the L33 gene

5.3.2.1 Identification of the full-length cDNA sequence

Initial sequences of the L33 (laboratory internal ID number) gene were 14 clones isolated from the retina SSH cDNA library. These sequences belong to partial transcripts with retina specific expression.. A number of additional clones were publically available from the same locus on chromosome 15q23. The consensus sequence representing 4F10F, G09_Ret15, E11_Ret15, F12_Ret14, D10_Ret6 and 3A4F clones (Fig. 26B) were subjected to a homology search against GenBank database using the BLASTN program and additional human ESTs were obtained. Analysis of the longest fragments (accession number BQ189587 and BF529111) confirmed the compiled consensus cDNA sequence (Fig. 26B). Moreover, four of the matching EST clones (GenBank accession numbers AI478365, BE220398, BF110735 and BF446762) were obtained from the I.M.A.G.E. consortium, but their sequencing did not reveal any new information.

To obtain the full-length sequence of the L33 gene, the human retina full-length cDNA library (local library ID: Jelena), cloned in λTriplEx2 vector, was screened with a 826 bp long PCR fragment, obtained with PCR primers L33F and L33R1 primers and labelled with [α-^{32}P]dATP. Among 10^6 phage plaques, three positive clones (L33_1/3, L33_2/4 and L33_3/4) were isolated, each with a 1550 bp long insert. Complete sequencing of the L33_1/3 and L33_2/4 clone with the λTriplEx 5' and the λTriplEx 3' primer showed that they were highly similar to previously obtained sequence (Fig. 26B). The 3'-terminus of both clones ended with potential polyadenilation signal (AATAAA) and a stretch of a 28 bp poly (A) tail.

Assembly of analysed sequences resulted in identification of the full-length cDNA of the L33 gene, which is 1663 bp long.

Two human mRNAs, which were made public at a later time point, contain additional sequences from the 5' end (Fig. 26B), indicating that there could be alternative 5' end and possibly alternative splicing of the transcript. This was not studied further.
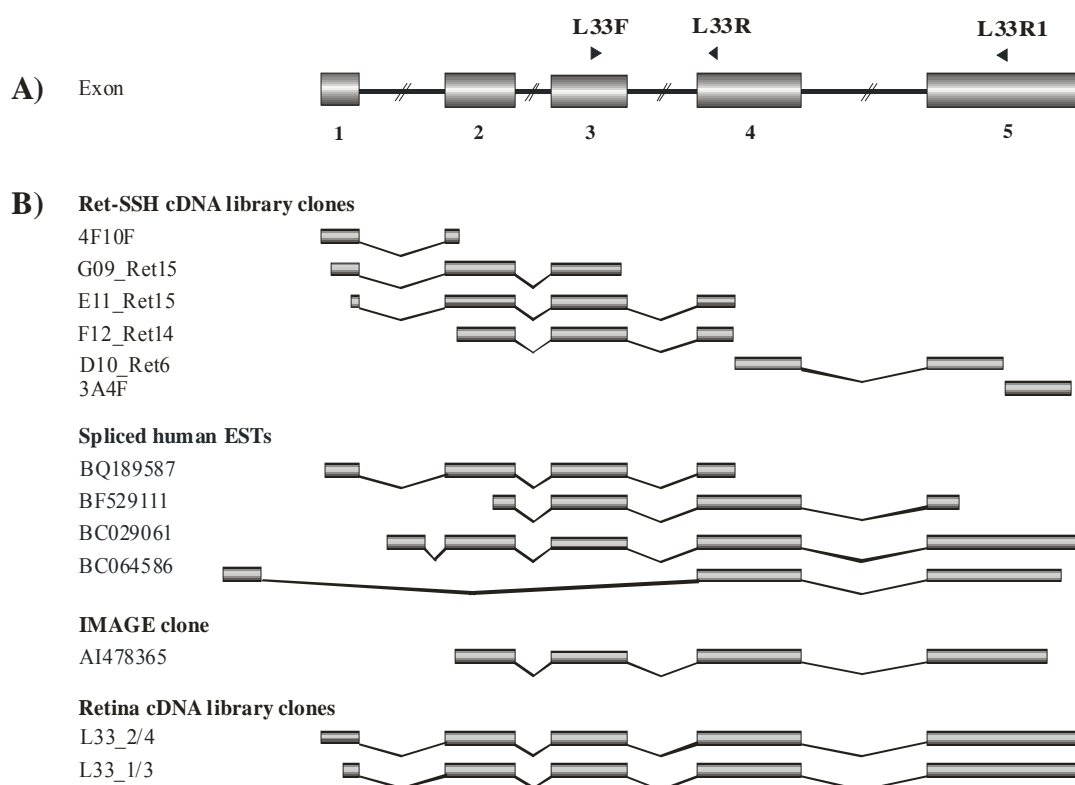
Fig. 26. Genomic structure of L33 gene.
A. Schematic representation of the L33 gene structure. The boxes represent the exons with their relative size and original organization. Numbers below boxes indicate exon numbers. The names and positions of the primers used for the analysis of L33 gene are shown above the schematic illustration. B. A consensus cDNA sequence was assembled using sequence of 14 clones from the human retinal cDNA library which was constructed by suppression subtraction hybridization (Ret-SSH cDNA library) together with the homologues human ESTs, four sequenced I.M.A.G.E. clones and two cDNA clones retrieved from cDNA library screening (clones have internal ID number). Only representative clones were selected for the schematic presentation.

5.3.2.2 Genomic structure of L33

In order to determine the genomic structure of L33 gene, the full-length cDNA sequence was aligned to the human genome draft sequence. The L33 gene spans 13 Kb of genomic sequence. It consists of five exons (Fig. 26A) which are 204 bp, 187 bp, 215 bp, 334 bp and 695 bp long, respectively (Tab 9.). Exons conform to the exon/intron boundary consensus GT/AG and very good scores for the consensus splice junction sequences (Tab 9).

Table 9.
Exon-intron boundaries of the human L33 gene

Exon		3' Splice acceptor site[a]		5' Splice donor site[a]		Intron	
No.	Size (bp)	Sequence	Score[b]	Sequence	Score[b]	No.	Size (bp)
1	204			..GA**gt**aagt..	2,948	1	4,782
2	187	..tttctttgtctgca**ag**A..	7,633	..AAA**gt**gagt..	2,384	2	287
3	215	..ccctctatgattgc**ag**G..	5,352	..GAG**gt**actg..	6,463	3	1,389
4	334	..catcgttgtcctgt**ag**G..	6,184	..TCG**gt**gagt..	2,829	4	5,212
5	695	..gtaaattcattttc**ag**T..	9,349				

[a] For the splice junctions, intronic and exonic sequences are shown in small and capital letters, respectively. The 5' splice-donor and 3' splice-acceptor sequences are given in bold.
[b] Scores were calculated to references (Berg et al, 1988 and Penotti et al. 1991). Approximately 99% of sites have a score of 0-20 (3'acceptor site) and 0-11 (5'donor site).

5.3.2.3 Expression analysis of L33

To examine the tissue expression of L33, RT-PCR amplification was performed with primers L33F annealing to exon three and L33R annealing to exon five in a panel of 20 human tissues. Analysis showed that the gene is highly expressed in retina with lower levels of expression in brain, fetal brain, spinal cord, kidney, trachea, skeletal muscle and testis and very low expression in liver, colon and uterus (Fig. 27A).

Virtual Northern blot analysis of the ds cDNAs from different human tissues was performed using radiolabeled PCR fragment, amplified with L33F and L33R1 primers. A signal at approximately 1.6 Kb was most abundant in retina and very weak in RPE (Fig. 27A), confirming already determined predominant expression of the gene in retina and obtained length of the transcript. Weak bands, one at approximately 1.0 Kb and another around 3.0 Kb could also be detected in retina (Fig. 27B). They could derive from alternatively spliced full-length transcript to which radiolabeled probe was hybridized.
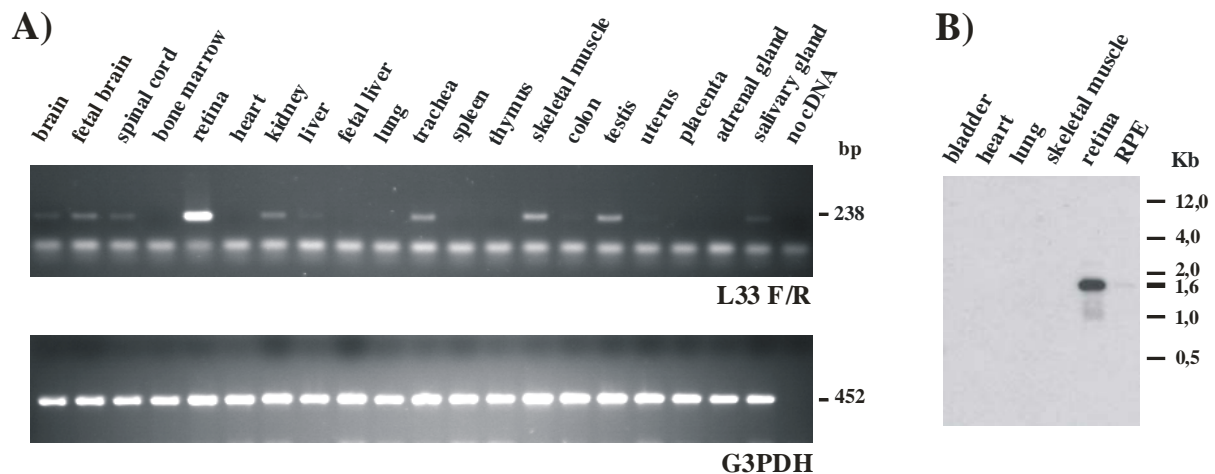
Fig. 27. Expression analysis of L33 gene.
A. Expression analysis of L33 gene in 20 tissue panel, with exon spanning primers L33F and L33R that were designed based on consensus sequence of the clones from the Ret-SSH cDNA library, showed that gene is abundantly expressed in retina and weakly in brain, fetal brain, spinal cord, kidney, trachea, skeletal muscle and testis. The same panel was amplified with G3PDH specific primers as a control of cDNA quality. **B.** The transcript size of the L33 gene was determined by virtual Northern blot analysis with radioactive labelled PCR product amplified with L33F and L33R primers. Signal at approximately 1.6 Kb was determined only in retina.

5.3.2.4 In-silico analysis of the cDNA sequence

To reveal the protein encoded by the L33 transcript, ORFs within the consensus cDNA were investigated. Unexpectedly, termination codons were located all over the sense strand of the cDNA, and no reasonably long ORFs that could encode proteins were observed in any of the three reading frames (Fig. 28).
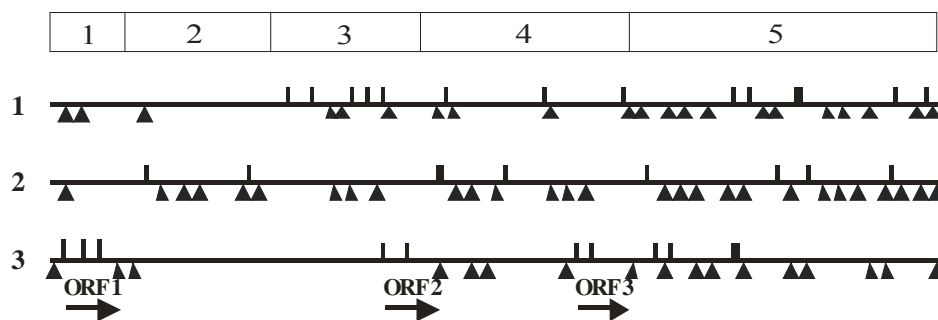
Fig. 28. ORF analysis of L33 cDNA.
The relative positions of all initiation and termination codons are indicated by vertical lines and filled triangles, respectively, for each of the sense open reading frames of the L33 cDNA. Three longest predicted ORFs are represented with arrows.

F bp   **aa**

```
          |Exon 1
     1    CCTGCATCTTAATCCACTCCAGAGCTGAGGCTTCCTGATCCCTCTAACCAGGTGC[ATG]TTCACTGGGCTTCCAGCTTCCA
+3                                                             M   F  T  G  L  P  A  S    8

    81    AGCCTTCTACCTGTGGAATGCTTGGTCCAATGTCTGGGGCACCCACTCTTACTCCAAACTCCTCCAGATCTGCAGAGTGG
+3         K  P  S  T  C  G  M  L  G  P  M  S  G  A  P  T  L  T  P  N  S  S  R  S  A  E  W   35

                                                    |Exon 2
   161    CCCAGCATATTCTCTTCT[TGA]AAGGAGGTCTTTGCCTAACAGAAATGGAATAAAAAGGCCTACCTCATTGGGCTCGTGTG
+3         P  S  I  F  S  S  -                                                               41

   241    TGAGGAGAACTGAAGAGTCTGAGAGCGCGGCACGAGCCAGAGGCTACGGAAAACACTGCCCTCCTACACTCCACCTTG

                                                                            |Exon 3
   321    GAGAGACCCAGAAAAGAACAAGCTTCATTTGTAAAAAAAGGAAAACAACTCAGGCAATGGGGGTGGCTTAAAGGCACTCTA
   401    AAACAACTCAGGCAATGGGGGTGGCTTAAAGGCACTCTACAGTGTGCAGATGCCTTCCACTTCTTCCATCTGCCCTGTCT

   481    CTCCAAGAACCCCTATGGCCCCGGTCTCAGAACAGAGCTGAGTGCAGAA[ATG]AAAATCTATGGCTCTGTGTTCCAAAACG
                                                          M   K  I  Y  G  S  V  F  Q  N     10

   561    ATGAAGAATTTCAAGATGGTGGCAGTGGTAAAATCTTTCTCCAGGAAAATCTGTCCTTGGCCCAATGTGTAAACACTTG
+3         D  E  E  F  Q  D  G  G  S  G  K  I  F  L  Q  E  K  S  V  L  G  P  M  C  K  H  L   37

                  |Exon 4
   641    CTGAGGAACTTGGAA[TAA]CTTGCAGTGTCTTGCAGTATTGTGAAACCAGCAACTTGTTCACAATTCTTCTGAATTTCTTG
+3         L  R  N  L  E  -                                                                  42

   721    GGAAATTTGAAGTGGAGTACCTGTACCAACATGAAATGACACGAATTTAAGTGACGCTCAACAACGAAAAGCAAAAAGAA

   801    CCAAAGAGGAAGCAACTGAAACAACATCTGGATGTATTTAAAAATATACA[ATG]CCTCCAAAATCAGGTGTCATTAATGAA
+3                                                          M   P  P  K  S  G  V  I  N  E    10

   881    AATTCTGAAGAAATGCCACCGGACATAGCCAACGCACCTACGCTGTTGTTATTCATTTCCTGCTTTTCACAGAAAACAAT
+3         N  S  E  E  M  P  P  D  I  A  N  A  P  T  L  L  L  F  I  S  C  F  S  Q  K  T  I   37

                  |Exon 5
   961    TTTGTTGCATGGAAGATCG[TGA]ACTTCAAAGGGCAGAGGGGAAACTGTCCCTTGGCCTCTACCCCTCCAAGGCCCCACTT
+3         L  L  H  G  R  S  -                                                               43

  1041    TTTTGTCAACACTCCTTGGACGCAGCAGAAGTATGAACATAATATGGTCCTGAATGAGGCTGAGTCTTTGGGCGCAGAAG
  1121    ACCCGGGTTAATAAAAATAGGAAGGTAAGAAAAGAAAAGAAAAATCAAGACACATCATAGGACTAAATTCCTATTATTTA
  1201    TCCACTCAGGATTGACCACCCCTTTGGGCCAGATAGTTGTACCCCCATGTACCAGGTGGGCACATGAAGACACAAGAAGT
  1281    GCTGTGATGGTTCATTTTGCACGTCACCTTGCGTGGAGTATGCCAACTCATTGTTTGGTCAAACACTAGTCTGGACATGG
  1361    TGGTAAAGGTATTTTTTAGATGAGATTAACACTTAAATCAGTAAAGCAGGTTACCCACCATACTACGGGTGGGCCCTGTC
  1441    CAATCAGTTGAAGGCATTAAGAACAAAGATTGAGGTTTCCTAAAGAAGATGGAATTCTCCTTGAGACTACAACATAGAAA
  1521    CCCTATCTGAGTTTCCAGCCTGTTGCCCTGTGAATTCAAACTCAGGACTCCGGTCTATGACATTAACCCTCACTTAACT
  1601    TTTCAGCCTGCCAGCCTGCCCTATGGATTTCGGACTTGCCAGCCACACAATTCCTTAAAATAAATCTCTCCGTCAAAAAA
  1681    AAAAAAAAAAAAAAAAAAAAAA
```

Fig. 29. Nucleotide sequence of L33 gene.

The complete cDNA sequence is numbered starting with the first base of the 4F10F clone. Amino acid sequences encoded by putative ORFs are indicated bellow the nucleotide sequence and length of the potential proteins are given on the right side. Potential initiation (ATG) and termination (TGA/TAA) codons of the putative ORFs are boxed and putative polyadenilation sequence (AATAAA) of the L33 transcript is underlined.

Table 10. The longest ORFs for the L33 transcript

ORF no.[a]	Frame	Start	Stop	Length (bp)	Initiator sequence[b]
ORF1	3	56	178	123	a g g t g C **A T G** t
ORF2	3	491	616	126	G C a G a a **A T G** a
ORF3	3	811	940	129	t a t A C a **A T G** c
Consensus Kozak's rule:					A G C C C C **A T G** G G

[a] The ORF no. column shows the assigned numbers of three ORFs ordered in length.
[b] The initiator sequence column shows the -6 to +4 sequence of the putative initiator site. The start (ATG) codon is shown in bold. The nucleotides that match with the Kozak's consensus are shown by upper case.

The longest putative ORF (ORF3; +811 to +940, Table 10, Fig. 29) encodes 43 amino acid residues. There are a total of three ORFs that could encode more than 41 amino acid residues (123 nt, Table 10). Among them, ORF2 would be the most likely one, particularly in regard to the strong consensus of the translation initiation sequence for many mammalian mRNAs proposed by Kozak (1987; Fig. 29; Table 10), with ATG representing the initiation methionine.

To examine whether the longest predicted ORF indeed may encode a protein, it was analysed further under the assumption that a functional ORF should be conserved among species. The overall nucleotide sequence of the L33 transcript had no significant homology to any nucleotide sequence available at the public databases, including genomic or ESTs. In addition, none of the ORFs had putative conserved domains or significant sequence similarity to any other known protein. This finding may indicate that ORF1, ORF2 and ORF3 do not encode amino acid stretches.

Performed analysis of the nucleotide sequence and putative amino acid stretches encoded by the three longest ORFs suggest that the L33 transcript may not encode a protein, but instead may function as a regulator mRNA.

5.3.3 Cloning and characterization of the L35 gene

5.3.3.1 Identification of the full-length cDNA sequence

Identification of the cDNA sequence of the L35 gene was based on four clones (6E1F, 3B6F, C10_Ret10, F04_Ret15 and H01_Ret15) from the retinal SSH cDNA library. These clones belong to the category of partial transcripts with retina specific expression. Genome sequence comparison revealed that these clones are deriving from a locus on chromosome 7p21, all mapping to the first intron of the MEOX2 gene (Fig. 30A). Orientation of splicing of 6E1F was opposite to that of the MEOX2 gene. The remaining clones which partially overlap represent a continuous cDNA sequence. BLASTN homology search with the consensus sequence revealed two ESTs (accession numbers AI417984 and BG184437) which showed homology to 6E1F, thus extending its sequence by 385 bp at the 3' end (Fig. 30A). Analysis of the BG184437 clone suggested that there could be another isoform of the transcript. In order to investigate possible splicing events further, sequence information from the 6E1F cDNA was used to design exon spanning forward and reverse primers L35F and L35R, respectively. RT-PCR amplification yielded a 145 bp long PCR product of the expected size, which was very weak in abundance, and an unexpected highly expressed PCR product of 288 bp in size. Sequencing of the two PCR products revealed that the shorter one was identical to 6E1F clone, whereas the longer one suggested the presence of an additional exon.

In order to retrieve the full-length cDNA sequence encoding the L35 gene, a mix of adult human retina libraries DKFZ3, DKFZ4 and C1F1, was hybridised with a 288 bp long PCR fragment amplified with L35F and L35R primers and labelled with $[\alpha\text{-}^{32}\text{P}]\text{dATP}$. Screening of the libraries in duplicates identified four positive clones (L35_2.3, L35_3.1, L35_3.2 and L35_4.1) which were completely or partially sequenced with λTriplEx 5' and λTriplEx 3' primers. Further analysis of the L35_2.3 clone which was partially sequenced from both ends extended the sequence of 6E1F by 195 bp at its 5' end (Fig. 30A), whereas its 3' end overlapped with remaining novel retina sequences, suggesting that all five clones from the retinal SSH cDNA library derive from the same transcript. The L35_2.3 sequence terminated with putative polyadenilation signal, which was followed after 13 bp by 19 bp long poly (A) tail (also contained in retinal BX538274 mRNA which was sequenced at a later time point).

In addition, the human retina full-length cDNA library (local library ID: Jelena) was screened with the same radiolabeled 288 bp PCR fragment. After two rounds of hybridization two positive clones, L35_2/2 and L35_3/2, with approximately 750 bp long inserts were

obtained from approximately 1×10^6 recombinant phages. Complete sequencing of both inserts with λTriplEx 5' and λTriplEx 3' primers showed that they are highly similar to previously obtained cDNA sequences (Fig. 30A). Interestingly, isolated clones terminate upstream from identified potential polyadenilation signal, both having 28 bp long poly (A) tail.
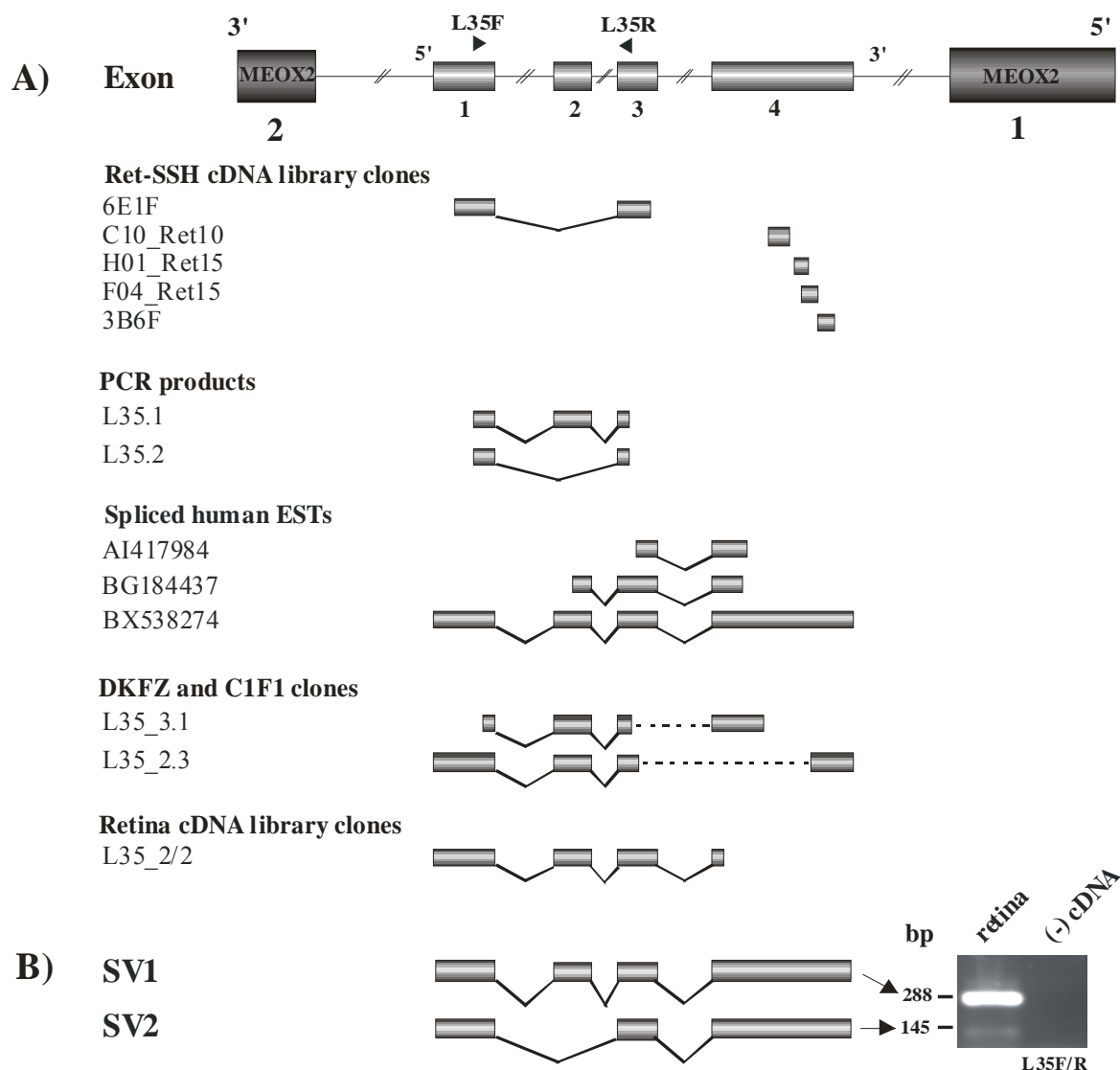
Fig. 30. Genomic structure of L35 gene.
A. Schematic illustration of L35 gene structure, located in the first intron of MEOX2 gene, with opposite orientation. The boxes represent the exons with their relative size and original organization. Numbers below boxes indicate exon numbers. The names and positions of the primers used for the analysis of L35 gene are shown above diagram. Clones from retinal SSH, full-length, DKFZ (3 and 4) and C1F1 cDNA libraries and PCR products have internal ID number and only representative one were selected for the schematic presentation. Broken line represents partially sequenced clones. **B.** Sequences of the L35_SV1 (SV1) and L35_SV2 (SV2) isoforms were assembled by alignment of all clone sequences and two PCR products, which were amplified with gene specific primers L35F and L35R, to the corresponding human genome sequence.

Assembly of analysed sequences resulted in the identification of two cDNA isoforms, named L35_SV1 and L35_SV2. The cDNA sequence of the L35_SV1 is 3759 bp long, whereas L35_SV2 is shorter, due to lack of an exon which is 141 bp in size (Fig. 30B). While a polyadenylation signal is present in both isoforms 32 bp upstream of the most 3'-end of the isolated cDNAs (Fig. 33), the presence of a poly (A) tail in L35_2.3 and L35_3.1 clones suggests usage of an alternative polyadenylation signal generating a 749 bp long transcript.

5.3.3.2 Genomic structure of L35

To determine the genomic structure of L35, the full-length cDNA sequence was aligned to the human genome draft sequence. Accordingly, the L35 gene consists of four exons that span approximately 13 Kb. It is located in the first intron of the MEOX2 gene, which is encoded from the opposite strand. The exons of L35 are 350, 141, 185 and 2.779 bp in size, respectively, and are separated by 7.723, 1.302 and 1.266 bp intervening sequences (Table 11). All exon/intron junctions have appropriate splice donor/acceptor sites and reveal high scores for the consensus splice junction sequences (Table 11).

Table 11.
Exon-intron boundaries of the human L35 gene

Exon		3' Splice acceptor site[a]		5' Splice donor site[a]		Intron	
No.	Size (bp)	Sequence	Score[b]	Sequence	Score[b]	No.	Size (bp)
1	350			..ATA**gt**aaat..	6,246	1	7,723
2	141	..attatttattcaaa**ag**G..	9,839	..GGG**gt**ttgt..	7,439	2	1,302
3	185	..ttcggttgactctt**ag**G..	7,369	..CTG**gt**aaga..	2,539	3	1,266
4	2,779	..tgtttttttccttgt**ag**C..	4,268				

[a] For the splice junctions, intronic and exonic sequences are shown in small and capital letters, respectively. The 5' splice-donor and 3' splice-acceptor sequences are given in bold.
[b] Scores were calculated to references (Berg et al, 1988 and Penotti et al. 1991). Approximately 99% of sites have a score of 0-20 (3'acceptor site) and 0-11 (5'donor site).

5.3.3.3 Expression analysis of L35

The expression pattern of L35 in 20 different human tissues was determined by RT-PCR analysis with forward primer L35F, located in the first exon, and L35R, located in the third exon of the gene (Fig. 30A). Highest expression of the two isoforms was found in retina, whereas in skeletal muscle and testis a longer isoform was present at very low levels (Fig. 31A). A weak PCR product of approximately 400 bp in size was also detected in skeletal muscle and testis and might be amplified from alternatively splicing or unspliced mRNA species (Fig. 31). This was not analysed further.

Virtual Northern blot hybridization experiments using ds cDNAs from six different human tissues was performed using as a probe the 288 bp long PCR fragment labelled with [α-^{32}P]dATP. The experiment showed that L35 is expressed as a transcript of about 3.5 Kb exclusively in retina (Fig. 31B), confirming tissue specific expression results obtained by RT-PCR analysis.
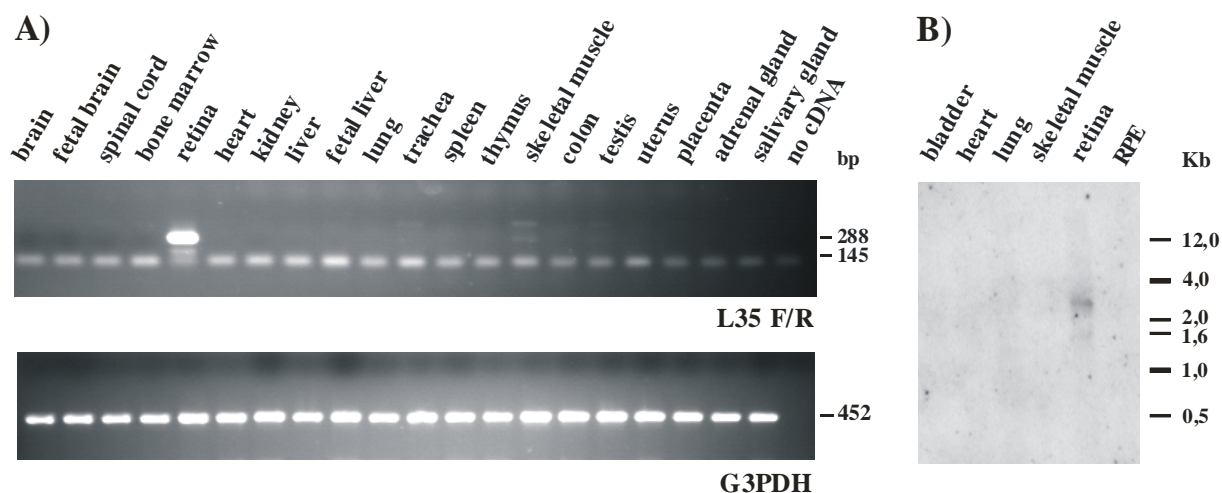
Fig. 31. Expression analysis of the L35 gene

A. Expression analysis of L35 gene in 20 tissue panel, done by RT-PCR, showed that gene is expressed abundantly in retina and very weakly in skeletal muscle. The same panel was amplified with G3PDH specific primers as a control of cDNA quality. **B.** The length of the L35 transcript was determined by virtual Northern blot analysis using radioactively labelled PCR product amplified with L35F and L35R primers. Obtained signal correspond to transcript which is approximately 3.0 Kb long and confirmed retinal expression.

5.3.3.4 In-silico analysis of the cDNA sequence

A striking feature of the full-length L35 cDNA sequence is the high density of stop codons in all three reading frames (Fig. 32), which suggests that if L35 is a coding RNA, the protein product is likely to be very small. Analysis of the coding potential of the L35 cDNA revealed four open reading frames (Table 12). The most probable one would be the ORF2, with regard to a reasonable consensus to the translation initiation sequence proposed by Kozak (Fig. 32; Table 12), whereas the longest ORF4 would encode 54 aa. The alternative splicing of exon 2 would not change the predicted open reading frames.
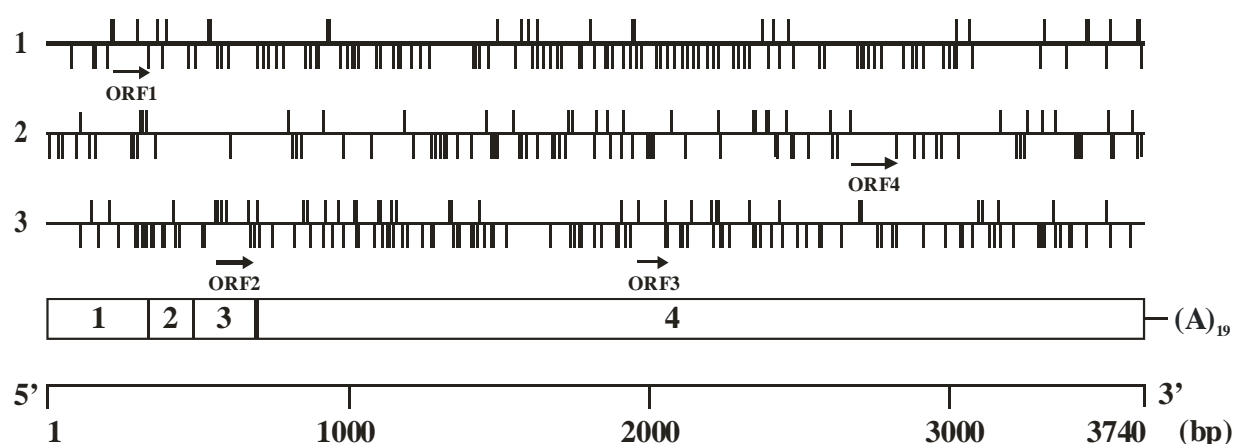
Fig. 32. Open reading frame analysis of L35 cDNA.
The relative positions of ATG codons and termination codons (TGA, TAA and TAG) are indicated by upper and lower vertical lines, respectively, for each of the three open reading frames. Four longest ORFs are represented with arrows. The four open boxes represent exons 1-4 of the L35 gene.

```
bp                                                                                            F    aa
        |Exon 1
1       AGGACACAGTTAAACAACAGATCCACTTTATTCCATCCCCTTGGAATAACATATTTAAATTTCTTATTGGTTTAATTCGT
81      TTTTCTCACATCCCGAGTTTTCAAGgtaaagccaagtaatgtgggaaactaccctatttctactgtgtactaacatgcattt

161     AACTACAGTTGACCTTTGAACAACAGGGGGTGGAAACTGTGTAGGTCTACTTATGC ATG GATTTTCTTCCGCCTCTCCTA
                                                                 M   D  F  L  P  P  L  L   +1   8

241     CCTGAGACAGAAGACCTTTGCCTTCTCTTGCTCCTTCTGCTCAGCCTGCTCAACCTGACGGTGATGAAAACTGAAGCCCT
         P   E   T   E   D   L   C   L   L   L   L   L   L   S   L   L   N   L   T   V   M   K   T   E   A   L   +1   35

                                            |Exon 2
321     TGATGATGATCCACTTCCGCTTAATGAA TAG CACTGTAGAAGTCTCGCATATATGATCATTTAAAATTTCGCTGAAGTCA
         D   D   D   P   L   P   L   N   E   -                                           +1   44

401     TGTCTGAATTCTGCACACTGCCTTGTGGATGTGGGAGAGGTGACTAAACTGAGTCAACGCAGTCAACCACCCAGGACTTG
```

```
                    Exon 3
481     CAATTCCAGGGGGGTTGCATCCACCTAACCCTCCGGGGTTCTGGCTTCGGATAGTGTTTGATCAATGGAAGCAACAAAAGA


561     AGATTGGAGGAC ATG AAGACAGGTTACTGGATGATATTCGACATGAGCTGCAATCCTCCATCTGAGGTCCCAACTTTTCC
                     M   K   T   G   Y   W   M   I   F   D   M   S   C   N   P   P   S   E   V   P   T   F   P  +3 23

                       |Exon 4
641     AGAAGGCTTCTGTCCTGTTGCGCTCATTGGATTCTGCATTATGGGATTCTCT TGA AGTGTGTGAAGAATAATGAGCTAAA
          E   G   F   C   P   V   A   L   I   G   F   C   I   M   G   F   S   -                        +3 40

721     TAAAGAGGAAAAACAAAAGAAAACAAAAACTGCAGGGACTTTTATAACTATCATTGATTGCCAGCTTAAGCCTGCTGGAC
801     TCTTCACAAAACTATATATGCTTCAAGCTTAATATTTAATTTTGTAATTGGAACATTTATAATCATATTATGCATTTAAT
881     AAGGATGCTCTAATTCAGATTTACTATTATCAGGAAATTAAATTTGGAGTTACAGTAATGCATGCATATATATGTGTGGG
961     TGTGAGTGTGCATATACCTATATATGTATATACATATATAAAGTAAACGCCTTCTTTTAACGTGTGTTTTAAGTATAGCT
1041    TCCATGTCCTAGTTAATTCAGCACCCTCCCTTTTATCCTTTCCTTGCCATTGTCCTCTCTACTAGACTAC AATAAAAACA
1121    AAAAACTATGAAGTAAGCAGCTTATAATCATAATTTTAATCAATGTAGCCTATGCAAGCTTTAAGAAACATACGTCTAA
1201    CATCTTAGAAAGAAAGCTTATGAACACAAACAGCAGATACTTAAATACATTTGAAACATTCACATTGAATTGATTCTAC
1281    AAAATTTTGAAGAGCAACAGCTTTTGAATTTAGAGCTATAATCAGTACCCTAAAGTTTTTACCACTGAGTTACTACCTAG
1361    TAAAAATATGTTATATTTGCAATAAAAGTACACATTGTAAGGACATAAGAATTGTCATCAAGGTACACAAACATTTTAGA
1441    ATTATTAAACATTTAAATAAATGGAAATTTTAAAAATAAAATAACAATATTATCATACAAGTTAATGATCCAGTTTATTA
1521    ATTTCTAATTCTATTGATAAAACAAGGGATTATTTTTGAAAGGGTTTTTCTATCTTTCATACAAACCCTTTCCAAAAATG
1601    TATAGATACAGTACGGATTCTACCTGATCATTCTTGTGCATATGGTGATGTAAGAAACCAAAAAAATGTAAATTATTTTC
1681    ATAACACTGTGAAAACCAATATTTAAAAATACAGTGCACAATTAGTAGTGAGTAACATATATAAATTTAGAAATATTATA
1761    TTTCTGTACACAAAGTAACATTATAAATGAAAGCTTCAGGAATAATGAGCTACTTATTATTCCGGAAATTATATATCAAA
1841    AATTTCAAATATAGAGGAAAATTAAAAACCCACTCATGACCCATGAAAATTAATAACCCATTCGATCATACACTTACCAT
1921    GTAACTTTCTGTTAATTTGGTAATAAATAATGTCAAGAGATTTTAAATGAAATACTTATTCTGAAACATAAAAACAACA


2001    TCTGAG ATG AGCACAAATGGATCAATAGATACATTATTGATTATTTTGTATCATAAAATACTATCTAGTACAACATTCTT
               M   S   T   N   G   S   I   D   T   L   L   I   I   L   Y   H   K   I   L   S   S   T   T   F   F  +3 25

2081    TAAAAAATTCAATATTTTTATAATG TAA GTAAACGGGAAGAACTGTTAGAGAAATGGGCATCAGATTTAGATTAATTCT
         K   K   F   N   I   F   I   M   -                                                               +3 33

2161    GCCTCTAACGCAAATTGTAGGCTGAGCCCTATATGTTCTTGACCATTCTAAGTTTTTACTACATTAAAACACATAAAATAT
2241    TCTGCGAGGTTTTCTCCAATGGGATTTTGAATCTCACTGATGACTGTGATTTAGGAGTAATGTAGAGAGTTTTGCAACTA
2321    ACTTTACTTTACACCGTTACTCCTTTTCTACTTGGGGGTTCAGTTCAACACGCTAGTAAAAGAAATAGTCATGAGAGCAT
2401    CCTAAGGAAAGTAGGCAAAATGTTAGTGCCTCATTATATTTACAATCTTTACAAAACATGTATGTCAGTAAACCATATTA
2481    ACATCTAAAGCACATGATATGTAGAAACAAATTGTGACTACAAAATGTGTCTTGTTTCTTAAATATGCTAAATAGGAAAT
2561    AAATCACATCTACTTCAATTTTTAATCTGTGGACTTTGGCCTTAGTTTGTGTTTTATTTTATTTTTTTGCTTTTGATTTT
2641    AACCACTGAAATCATTTACCAATACTATAAGCCAGGAAATACAATGTAACTGGCCAGTTGAGAATTAAGCCTTCTCTTTT


2721    TATTATTACAATCAAGGGGGGAA ATG CTTATATATCTTCCTGCACAAAATGTTTCATTTATCAAAAAAAAAAAGTGAGAAA
                             M   L   I   Y   L   P   A   Q   N   V   S   F   I   K   K   K   S   E   K  +2 19

2801    AATAAAGAAAATTTGGGTGAGATAAATAATCCACAACTATTAATTCAACAGATATTATTTAAGAAATTCCTACCTGTAAT
         N   K   E   N   L   G   E   I   N   N   P   Q   L   L   I   Q   Q   I   L   F   K   K   F   L   P   V   I  +2 46

2881    TTATTTTCTAATATTTCTGCATCCT TAG TATTTATCGCTCAAAGATATTTGCAGCCCTCATAGCAGTGCTTGGAAAAGAG
         Y   F   L   I   F   L   H   P   -                                                               +2 54

2961    ATTTGCTCTGAAAATTTAGAAATAAATTCCTGAGGTGAATACAGTTCAAAACTAAAATATATTGCTGTATTATATTGGTT
3041    TTTCTCATTTGAAATTAAAATCTAAAAAACCTTCATATTAGGTTGTGTGTGTCAAGCTGCCATTATAAAAACACTAAATG
3121    AATAAAATTTCGCAAGCAAGGCAGCATTATCAGGCTAAGTACATGTAGTGGGCTTATATATGGTGATGTTTTACATAATT
3201    CAGCTATTAATTCATCTATAGATTTACATTTCTTAAGTGCAAAAAAAAATGTAATTATTTCATTACATATGGAATTCACC
3281    TCCCAAATTTCTCAGAATCTAAGAGAAAGGATTGACATTTTCTGATATTTCATTTTTGCTTAAAATTAAGTTCTTAGGAT
3361    GTTTAGATTTTATTATAAATATTACAACTTAAATTCCAGAGTTATAATGAACGGTCTATGTTTTGAAACACAGAGGGAGA
3441    ATGGGGTAATGACTGAAGTACTGTTTTTTCAGTTTATTAGTAAAGATCAGCTTATAAGATGTCATTGTAAAGGTTTCTAT
3521    TGATAATTACCATAAACAAAGATTAAAACTTTAAATACTAAGTTCTGGATTTCTAAGCCCGGTGAAAGTTTGTGCAATCA
3601    CATACCTTGCATTTATGATTTTTCTTTTAAAATGAACTATAAATGAAGAAATTTACACAGAATTGGAAACTATTCCTACT
3681    TTTATTTGCATTCTTCCTTCCATTCATTGCTAATGTTCAAC AATAAAG CATTTTAATTCTAAAAAAAAAAAAAAAAAAAA
```

Fig. 33. The complete nucleotide sequence and amino acid translations of the L35 gene.
The complete cDNA sequence is numbered starting with the first base of the 6E1F clone from the retinal SSH cDNA library. Amino acid sequences encoded by the four putative open reading frames are indicated bellow the nucleotide sequence. Potential initiation (ATG) and termination (TAG/TGA) codons of the putative ORFs are boxed. Putative polyadenilation sequence (AATAAA) of the L35 transcript is underlined.

Table 12. The longest ORFs for the L35 transcript

ORF no.[a]	Frame	Start	Stop	Length (bp)	Initiator sequence[b]
ORF1	1	218	349	132	t t a t g **C A T G** G
ORF2	3	574	693	120	G a g G a **C A T G** a
ORF3	3	2.006	2.107	99	t C t G a g **A T G** a
ORF4	2	2.744	2.906	162	G g g G a a **A T G** c
Consensus Kozak's rule:					A G C C C C **A T G** G G

[a] The ORF no. column shows the assigned numbers of three ORFs ordered in length.
[b] The initiator sequence column shows the -6 to +4 sequence of the putative initiator site. The start (ATG) codon is shown in bold. The nucleotides that match with the Kozak's consensus are shown by upper case.

The complete cDNA sequence of the L35 transcript has no significant homology to any genomic or ESTs sequence available in the public databases, suggesting that the putative ORFs are not conserved among species. In addition, none of ORFs show homology to putative conserved domains or reveal significant sequence similarity to any other known proteins. These findings may suggest that the predicted ORFs do not encode proteins, but may function as a regulatory mRNA.

5.3.4   Cloning and characterization of the L37 gene

5.3.4.1 Identification of the full-length cDNA sequence

Identification of L37 was initiated with the 3H5F clone isolated from the human retinal SSH cDNA library. This clone represents a partial transcript for which expression was determined as retina specific and which is mapping to chromosome 3q22. It is located in the 36 Kb first intron of ATP2C1, but spliced with opposite orientation to the ATP2C1 gene (Fig. 34). The clone sequence was subjected to BLASTN homology searches but homologues

cDNA sequences could not be obtained. In the next step, the human retina full-length cDNA library (local library ID: Jelena) was screened with a 275 bp fragment obtained by PCR amplification with L37F and L37R primers and labelled with [α-^{32}P]dATP. Among 10^6 phage plaques, three positive clones (lab ID numbers: L37_3/1, L37_5/3 and L37_6/2) were isolated, with inserts of approximately 400 bp. Complete sequencing of the clones with λTriplEx 5' and λTriplEx 3' primers showed that they were highly similar to the previously obtained sequences of 3H5F (Fig. 34), which terminated with a potential polyadenilation signal. The 3'-termini of both 3/1 and L37_5/3 ended with a 29 bp long poly (A) tail. Sequence information from the 3H5F cDNA fragment was used to design specific primers L37F and L37R. RT-PCR analysis resulted in a single band of expected size verifying the identified sequence (Fig. 34). Thus, assembly of obtained nucleotide sequences resulted in identification of the full-length L37 cDNA of 419 bp.
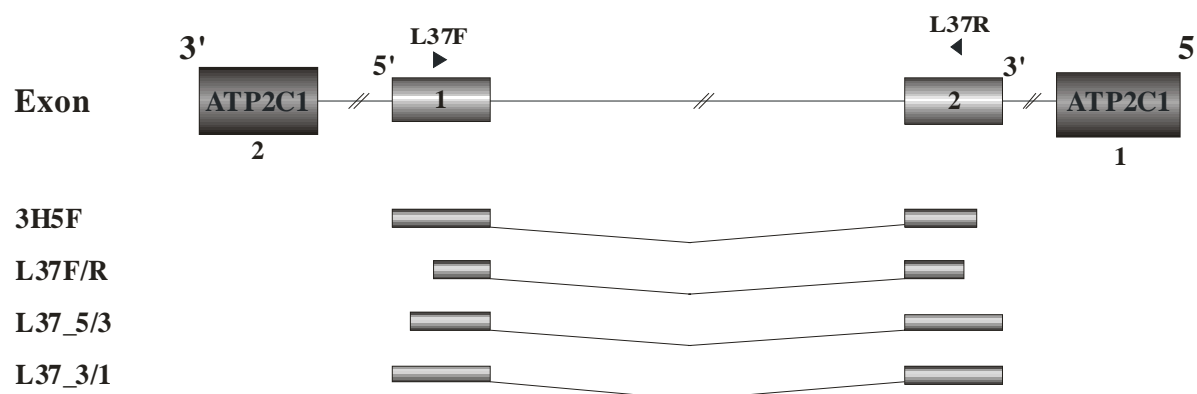
Fig. 34. Genomic structure of L37 gene.
Schematic representation of two exon gene, located in the first intron of ATP2C1, but spliced with opposite orientation. A consensus cDNA sequence of L37 was assembled using sequence of 3H5F clone from the human retinal SSH cDNA library together with two cDNA clones isolated from the full-length retinal cDNA library (L37_5/3 and L37_3/1). PCR product, which was amplified with L37F and L37R primers, verified already identified sequence. Exons of both genes are represented in proportional sizes and original organization.

5.3.4.2 Genomic structure of L37

To determine the genomic structure of L37, the identified full-length cDNA sequence was aligned to the human genome draft sequence. L37 consists of two exons (Fig. 34) of which the first is 171 bp and the second is 219 bp (Table 13). The two exons are separated by approximately 30 Kb of intervening sequence (Table 13). They have conserved splice donor and acceptor sequences that conform to the GT-AG rule by reasonable scores for the consensus splice junction sequences (Table 13).

Table 13.
Exon-intron boundaries of the human L37 gene

Exon		3' Splice acceptor site[a]		5' Splice donor site[a]		Intron	
No.	Size (bp)	Sequence	Score[b]	Sequence	Score[b]	No.	Size (bp)
1	171			..CAG**gt**tagt..	2.921	1	30.223
2	219	..tatacttttgtcgt**ag**C..	7.648				

[a] For the splice junctions, intronic and exonic sequences are shown in small and capital letters, respectively. The 5' splice-donor and 3' splice-acceptor sequences are given in bold.
[b] Scores were calculated to references (Berg et al, 1988 and Penotti et al. 1991). Approximately 99% of sites have a score of 0-20 (3'acceptor site) and 0-11 (5'donor site).

5.3.4.3 Expression analysis

RT-PCR analysis with gene specific primers L37F and L37R located in the first and second exon, respectively, examines the expression of the L37 mRNA in various human adult tissues. As shown in Fig. 35, L37 is expressed exclusively in the retina. To determine the transcript size of the L37 mRNA, the analysis of virtual Northern blot containing the ds cDNAs from bladder, heart, lung, skeletal muscle, retina and RPE were performed. The radiolabelled PCR probe was generated with the same primer pair used for the RT-PCR analysis. As shown in Fig. 36, a specific transcript of approximately 400 bp was detected exclusively in retina.
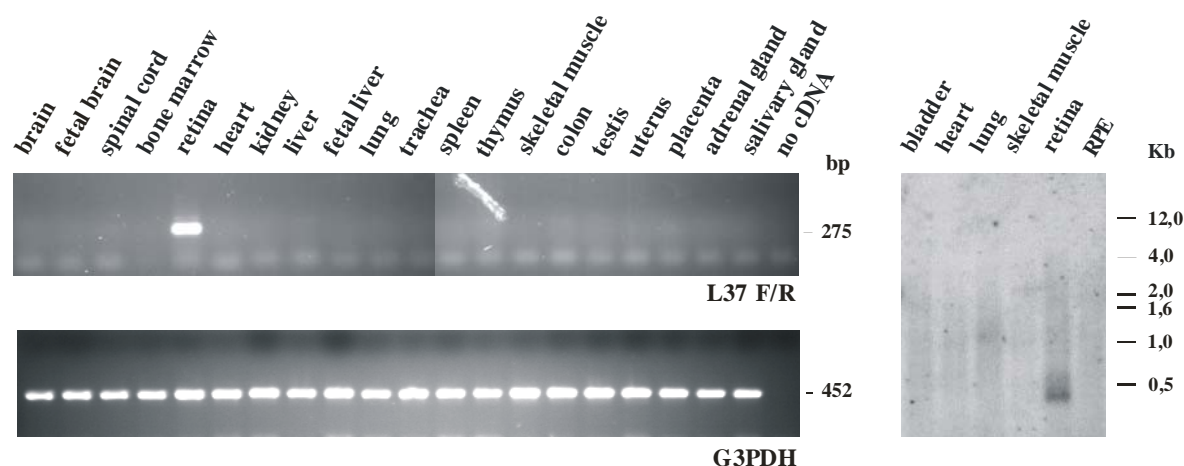
Fig. 35. Expression analysis of L37 gene
A. Expression analysis of L37 gene was performed by RT-PCR in the panel of 20 human tissues using exon spanning L37F and L37R primers. Analysis showed that gene is exclusively expressed in the retina. B. Hybridization of the radioactively labelled PCR product, amplified with the same primer pair, to the virtual Northern blot. A signal of approximately 400 bp was detected exclusively in retina.

5.3.4.4 Analysis of the L37 cDNA sequence

Analysis of the coding potential of the novel retinal L37 gene predicts two putative proteins consisting of 33 and 35 amino acid residues, respectively (Fig. 36). Both ORFs are predicted in the first reading frame. While the first ORF is spanning exon 1 and 2, the second ORF is located in the second exon of L37. The amino acid sequences of the two predicted ORFs was subjected to a homology search against public protein databases using the BLAST P algorithm, but no homologuous sequences could be detected. Conserved putative domains were not found for both proteins.

Fig. 36. Sequence of L37 gene (see next page)
The 419 bp long full-length cDNA sequence of the L37 gene has two putative ORFs. ORF1 and ORF2 are both predicted in the first frame and 33 aa and 35 aa long, respectively. Start and stop codons of the putative ORFs are boxed. Corresponding amino acid sequence and their positions are indicated bellow the nucleotide sequence.

```
 bp                                                                    F       aa

                                                                            ORF1   ORF2
        | Exon 1
   1    TCTCCATTTTAACTACTCCTCTTAGTCAGTCAGCTCATAGCTGATCTACCCAGAAAAGAC
  61    AACTGGCTACTCTTTCAACGCACC ATG ACAACAAGTCAGACTTTAAGAGCTATTCACAAA
                              M   T   T   S   Q   T   L   R   A   I   H   K    +1      12

                                                        | Exon 2
 121    CTAGGAGAAAAACAAGAAGTAAAAGAGACCAGACAAGCAAAGGATGCACAGCAACTTAAA
         L   G   E   K   Q   E   V   K   E   T   R   Q   A   K   D   A   Q   Q   L   K   +1      32

 181    GCC TGA AGACAGAAGATACACAGTTGACTCTTTAACAAC ATG AGTTTCAACTGTAAAGGT
         A   -                                     M   S   F   N   C   K   G    +1      33      7

 241    CTACTTATACTCAGATTTGCTTCTGTCACCCGAGACAGCAAGACCAACCCCTCCTTTTCT
         L   L   I   L   R   F   A   S   V   T   R   D   S   K   T   N   P   S   F   S   +1              27

 301    TCAGCCTACTCAAAGCGAAGACAA TGA GGATGAAGACCTTTATAGTGATCCACTTCTGCT
         S   A   Y   S   K   R   R   Q   -                                       +1              35

 361    TAATGAATTATAAATATATTTTCTCCTCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

5.3.5 Cloning and characterization of the L38 gene

5.3.5.1 Identification of the full-length cDNA sequence

Sequence analysis of the clones isolated from the human retinal SSH cDNA library identified clone 4E9R which localizes to chromosome 12q21 (Fig. 37A). Sequence comparison of the clone using the BLAST algorithm did not reveal homologuous cDNA sequences.

To retrieve full-length cDNA sequences of L38, a radiolabeled 221 bp PCR fragment was amplified with exon primers L38F and L38R, which were designed based on sequence information from the 4E9R cDNA fragment and used to screen the human retina full-length cDNA library (local library ID: Jelena) . After two rounds of screening, three positive clones (L38_1/2, L38_2/2 and L38_3/2) were identified among approximately 1 X 10^6 recombinant phages. Cloned inserts ranged from 750 to 950 bp and were partially or completely sequence. They revealed high sequence identities to the previously analysed retinal fragment (Fig. 37A). Their consensus sequence extended the 4E9R cDNA for 459 bp and 354 bp at the 5' and 3' end, respectively. The 3'-termini of the longest L38_1/2 clone contained a potential polyadenylation signal AATAAA and terminated with a 29 bp long poly (A) tail. Analysed nucleotide sequences of isolated clones suggested that there could be another isoforms of the transcript.

To examine the splicing event, primer L38R1 was designed from the sequence of the first known exon and L38F was designed from the sequence of the last identified exon. RT-

PCR amplification with retinal cDNA as template yielded the expected 711 and 576 bp products, and in addition, there were also two products of 440 and 270 bp in size (Fig. 37B). Sequencing of the four amplicons determined that the two longer ones, 711 and 576 bp in size, were amplified from variants identical to the L38_1/2 and L38_2/2 clones, respectively, Moreover, sequence analysis of the remaining two PCR products that are 440 bp and 270 bp in size determined that they represent the transcripts which have an internal deletion of 271 bp and 431 bp, respectively, due to the use of an alternative splice acceptor sites.

Assembly of the analysed sequences obtained from the retinal cDNA libraries and by RT-PCR resulted in the identification of a 1164 bp full-length cDNA sequence for the L38 gene.
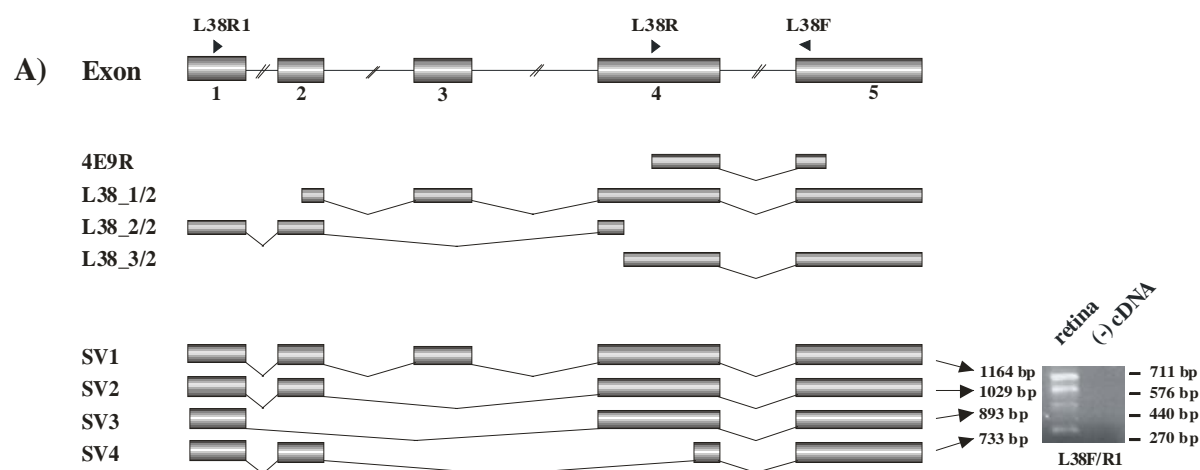
Fig. 37. Genomic structure of the L38 gene.
A. L38 gene contains five exons, it is located on the chromosome 12q21 spanning approximately 21 Kb of genomic sequence. The boxes represent the exons with their relative size and original organization. Numbers below boxes indicate exon numbers. The names and positions of the primers used for the analysis of L38 gene are shown above the schematic illustration. Based on cDNA sequence of the 4E9R clone from the retina SSH cDNA library primers L38F and L38R were designed and used for the amplification of the retinal cDNA. Obtained radioactive amplicon was used for the screening of the retinal full-length cDNA library which resulted in isolation of the L38_1/2, L38_2/2 and L38_3/2 clones. B. A consensus cDNA sequence was assembled using sequence of obtained clones and four different PCR products which were amplified with L38F and L38R1 primer pair. Calculated length of the transcripts is shown on diagram.

5.3.5.2 Genomic structure of the L38 gene

To determine the genomic structure of L38, the full-length cDNA sequence was aligned to the human genome draft sequence. This resulted in the identification of five exons with sizes of 93, 136, 135, 367 and 433 bp, respectively (Fig. 37A, Table 14). Their respective splice donor and acceptor sequences correspond to the consensus splice junction sequences (Table 14) and conform to the GT-AG rule (Burset et al, 2000). The exons are separated by four intervening sequences that are 1.352, 9.786, 3.844 and 3.244 bp in size (Table 14).

The above mentioned analysis resulted in the identification of four isoforms for the L38 gene, named L38_SV1 (SV1), L38_SV2 (SV2), L38_SV3 (SV3) and L38_SV4 (SV4). Their respective mRNAs are 1164 bp, 1029 bp, 893 and 733 bp in size, respectively (Fig 36B).

Table 14.
Exon-intron boundaries of the human L38 gene

Exon		3' Splice acceptor site[a]		5' Splice donor site[a]		Intron	
No.	Size (bp)	Sequence	Score[b]	Sequence	Score[b]	No.	Size (bp)
1	93			..CAG**gt**aaga..	1,057	1	1,352
2	136	..acttcatgaatttc**ag**T..	9,701	..CAG**gt**ataa..	5,843	2	9,786
3	135	..tcttttccttctgc**ag**G..	0,957	..GCT**gt**aagt..	4,542	3	3,844
4	367	..tttttatcatttcc**ag**A..	4,638	..CAG**gt**aaga..	1,057	4	3,244
5	433	..actcactgttccac**ag**G..	5,518				

[a] For the splice junctions, intronic and exonic sequences are shown in small and capital letters, respectively. The 5' splice-donor and 3' splice-acceptor sequences are given in bold.
[b] Scores were calculated to references (Berg et al, 1988 and Penotti et al. 1991). Approximately 99% of sites have a score of 0-20 (3'acceptor site) and 0-11 (5'donor site).

5.3.5.3 Expression analysis

To examine the tissue expression of L38, RT-PCR analysis was performed with gene specific primers L38F located in exon 4 and L38 located in exon 5 using cDNAs from 20 different human tissues as template. The results demonstrate that L38 is expressed

preferentially in the retina with minor expression in bone marrow, heart, placenta and adrenal gland (Fig. 38A).

Virtual Northern blot analysis of the ds cDNAs from different human tissues was performed using a radiolabeled PCR fragment, which was amplified with the L38F and L38R primers. A signal at approximately 1.1 Kb was detected exclusively in retina (Fig. 38B), thus confirming the previous results obtained by RT-PCR.
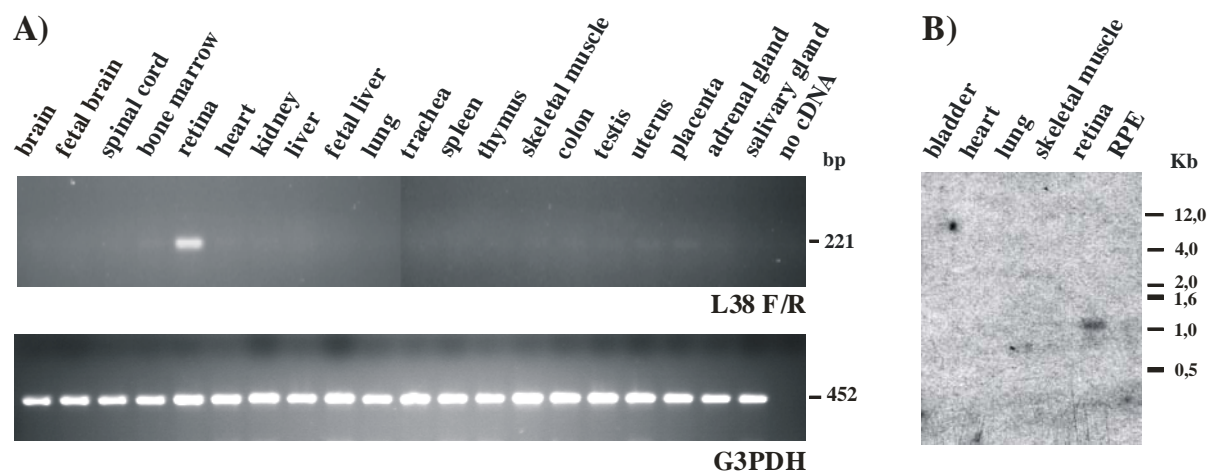
Fig. 38. Expression analysis of L38
A. Expression analysis of L38 in 20 tissue panel was done by RT-PCR using gene specific primers L38F and L38R. Analysis showed that L38 is exclusively expressed in human retina. The same panel was amplified with G3PDH specific primers as a control of cDNA quality. B. The length of the L38 transcript was determined by virtual Northern blot analysis using radioactively labelled PCR product amplified with the same primer pair which was used for the RT-PCR analysis. Obtained signal correspond to transcript which is approximately 1.1 Kb long and confirmed retinal expression.

5.3.5.4 In-silico analysis of the L38 cDNA sequence

Analysis of the consensus sequence of the L38 transcript revealed that stop codons were present throughout the sense strand of the cDNA (Fig. 39). The peptide encoded by the longest ORF is predicted in the first reading frame and is only 34 aa (Fig. 39 and 40).
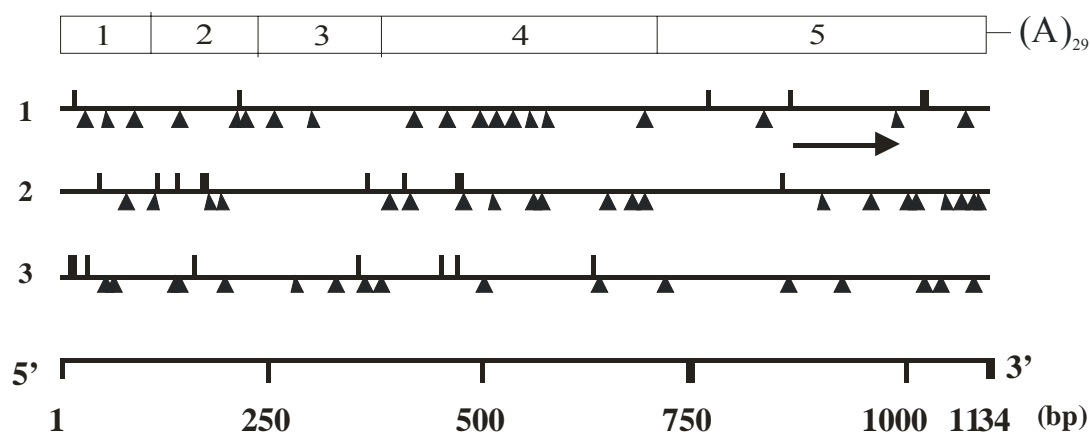
Fig. 39. Open reading frames of the L38 gene.
Positions of potential translation start and stop codons in all three reading frames are indicated by vertical lines and filled triangles, respectively. The longest predicted ORF is represented with arrow.

The nucleotide sequence flanking the ATG codon (ACCCCC<u>ATG</u>A) of the putative ORF corresponds reasonably well to the Kozak consensus sequence (GCCPurCC<u>ATG</u>G). However, this ORF is located within the LTR/MaLR repeat sequence of the last exon of the gene. The putative protein has no conserved domain nor does it reveal significant sequence similarities to other known proteins. It was also not possible to obtain significant homology of the complete L38 transcript to any genomic or cDNA sequences among different species.

Performed analysis of the cDNA sequence and protein encoded by the longest ORF, suggest that L38 transcript does not encode a protein, but may act as a regulatory mRNA.

```
   bp                                                   F        aa

      | Exon 1
    1 CTTATACATATATGGAAAAAAATGTGGACATGAAGAAGGAAAAAAAAATGAGCAGTGATCT

                                         | Exon 2
   61 GATCCATCAGGACATTCAAATTTTCTGAGTCAGTGACTTCGAGTGGCAGAAAATTTAAGA
  121 GATGCAGTATCTTCGGAGAGTAATGACATATAATCCACACATTCATGGAAGAATGGAGAC

                                               | Exon 3
  181 CTGAAGTCTTTGAAAGGGAGTAAGCAGAGAGGAAAATAGATGTGAACAGGAAGTCATTGT
  241 TCCAGGGACTGTTTATAGAGGCAGATTTTCAAAGGCCTCCAAACTCCAATTAAGATCATT
  301 TTGGCTTAAGAACACCCTCTCATTTATTACAGGATTTTCTTTAAGTACACAAAAGTTGCA

        | Exon 4
  361 TGCTATGATCTACCTTATTTTGGAATTAAGAAACACTTGAAGATTATTCTCAGAAATGCC
  421 ATAAGGGATATTTAAAAATCTTGGAGCACACAAGAAGTACAAGAGGAATGAAAATGTCAT
  481 GTCTTAATTTTCCAAAAAAGAGAATATTTTAGTTTTTAGTATATAAAGGCAGATTAACAA
  541 ACCTTCACACAAATCTAAATTTTATTACTAAATTAATTGATTACACTTTGTGTTTACTAG
  601 GTCACAGGGAATTTCAGTGGGTTTACTCTGCAAAAGATTTCATGTGAAATACATTTTTCT
  661 ATTCTGAAAGGTTTTTCAAGATTTTCAAGGCAACTAAGAAGTTTTCTAAAAAGACAGTGT

          | Exon 5
  721 TTGTAATACAGGCTCTATAGAAAGCATCGTT*GTGGAGGCCTCAGGAAACTTACAATCATG*
  781 *GCAGAAGGCGAAGGGAAGCAGGCACATCTTACGTGGCTGGAGAAGGAGGAAGAGAGAGCA*
  841 *GAGGAAGGTGCTACACATTTTTAAACAACCGGATCTCATGAGAACTCACTCACTATCACA*
```

101

```
 901    AGAACAGCAGGGAAATCCACCCCC[ATG]ATTCTATCACCTCCCACTAGCCCTCTCCTCCTT
                                 M   I   L   S   P   P   T   S   P   L   L   L   +1   12


 961    CAACACTGGGGATTACAACTCCACATAAGGTTTGGGCAGGGACACACATCCAAACCATAT
         Q   H   W   G   L   Q   L   H   I   R   F   G   Q   G   H   T   S   K   P   Y   +1   32


1021    CATCAC[TAA]TTAATGTTTCTGAAAGCTAGAGTTACATTACAACCACCTGAGGAGCTTTTA
         H   H   -                                                            +1   34


1081    GAAAATTACAGATCCCTGAGCTTCCATCTAGATTGATTAAATAAAATATCTGGGAAAAA
1141    AAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 40. Sequence of L38 gene.
Complete cDNA sequence of the L38 gene is numbered starting with the first base of the 2/2 clone from the retinal full-length cDNA library. Amino acid sequences encoded by one putative open reading frame are indicated bellow the nucleotide sequence and length of the potential protein is given on the right side. ATG and TAA codons of the putative ORFs are boxed and putative polyadenilation sequence of the L38 transcript is underlined.

5.3.6   Cloning and characterization of the L40 gene

5.3.6.1 Identification of the full-length cDNA sequence

Initially two clones (5B4F and B04_Ret5) from the retinal SSH cDNA library were isolated belonging to a partial transcript with retina specific expression. Partially overlapping clones were derived from the same locus on chromosome 7p12 by bioinformatics searches in the nucleotide databases. The consensus sequence representing both clones (Fig. 41B) was subjected to a homology search against GenBank database using the BLASTN program. Results obtained showed that the first exon of the B04_Ret5 clone overlapped with the seventh exon of the Zona pellucida-binding protein (ZPBP) gene, suggesting that the retina sequences identified could represent a novel splicing variant of ZPBP.

Screening of the human retina full-length cDNA library (local library ID: Jelena) was performed with the most 3' end of the previously isolated cDNA as a probe, which was amplified with exon spanning primers L40F and L40R. Three positive clones, L40_2/4, L40_3/3 and L40_6/3, were obtained from approximately 1 X 10^6 recombinant phages, with inserts ranging from 750 bp to 1150 bp. Partial or complete sequencing of the clone inserts revealed that they were highly similar to the previously identified retinal sequences. The largest clone L40_2/4 was characterized further and analysis of its nucleotide sequence demonstrated that it extends the sequence of the B04_Ret5 clone for 508 bp to the 5'-end, whereas the 3'-termini of both clones terminate with a potential polyadenylation signal

(AATAAA) and a 48 bp long poly (A) tail. Interestingly, the second exon of the L40_2/4 clone overlapped with the fifth exon of the ZPBP gene, whereas its third exon overlapped with the seventh exon of ZPBP. Nucleotide sequence analysis of clone L40_6/3 showed that it was amplified from a shorter transcript with an internal deletion of 310 bp, suggesting that there could be another isoform of the transcript (Fig. 41B). To investigate the splicing behaviour in more detail, sequence information from the identified cDNA fragments was used to design isoform-specific primers. Retinal cDNA was amplified with primer pair L40R2, which was designed at the 5' end of the L40_2/4 and L40_6/3 clones and L40F3, which anneals to the seventh exon of the ZPBP. PCR amplification revealed a product with unpredicted size, which contained an additional 219 bp, thus suggesting this product to represent novel splicing variant of the L40 transcript. Another primer pair L40R1, which anneals to the seventh exon of the ZPBP and L40F, which was designed from the 3'end of the cDNA sequences of the 5B4F and B04_Ret5 clones was used for the amplification of the retinal cDNA, which revealed a PCR product of the expected length of 175 bp in size (Fig. 41B).

The 1.037 bp consensus cDNA sequence of L40 gene was determined by assembling the sequences of representative clones and PCR products.

Fig. 41. Genomic organization of the L40 and ZPBP genes (see next page).
A. Schematic representation of the genomic organization of two overlapping genes, ZPBP (zona pellucida-binding protein) and L40. Two promoters, P1 and P2, used for the initiation of the transcription of mRNAs of both genes are indicated by arrows and shared exons are boxed. The approximate positions of the primers used for PCR amplification are indicated above diagram. **B.** Transcripts of ZPBP and L40 genes. White boxes represent the protein coding region, black boxes represent untranslated regions. The ZPBP gene is made up of eight exons, whereas L40 gene has seven exons. Three exons which are common to both transcripts are boxed. **C.** The complete cDNA sequence was assembled using sequences of the clones isolated from the human retinal SSH and full-length cDNA library together with PCR products amplified with L40F3/R2 and L40F1/R1 primers. Only representative clones were selected for the schematic presentation.

A) Genomic organization

B) ZPBP

L40

C) Ret-SSH cDNA library clones
5B4F
B04_Ret5

Retina cDNA library clones
L40_6/3
L40_2/4

PCR products
L40F3R2
L40F1R1
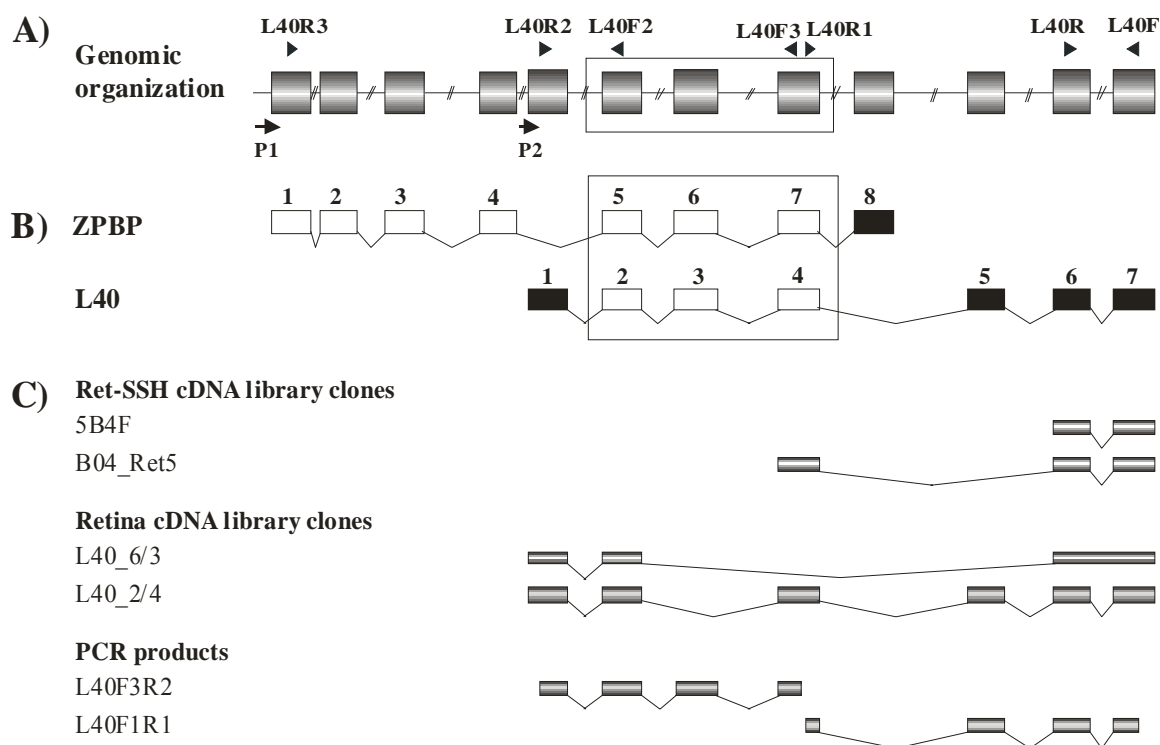
5.3.6.2 Genomic structure of the L40 gene

To determine the genomic structure of L40, the full-length cDNA sequence was aligned to the human genome draft sequence. This resulted in the identification of the gene that spans approximately 156 Kb of genomic sequence and is organized in seven exons (Fig. 41A) which are 200 bp, 219 bp, 77 bp, 178 bp, 132 bp, 98 bp and 83 bp in size, respectively (Table 15). Alignment of its full-length cDNA sequence to the human genome also showed that three exons of L40 (exons 2-4) overlap with three exons (exons 5-7) of ZPBP gene (Fig. 41A). All exon/intron boundaries of L40 reveal conserved splice donor and acceptor sequences that conform to the GT-AG rule with excellent scores for the consensus splice junction sequences (Table 15).

Table 15.
Exon-intron boundaries of the human L40 gene

	Exon		3' Splice acceptor site[a]		5' Splice donor site[a]		Intron	
No.	Size (bp)	Sequence	Score[b]	Sequence	Score[b]	No.	Size (bp)	
1	200			..AAG**gt**aata..	4,166	1	24,874	
2	219	..tgttaattttttac**ag**C..	6,889	..CAG**gt**aagt..	0	2	12,775	
3	77	..taatatattttctc**ag**T..	8,469	..AAG**gt**atgt..	2,549	3	34,72	
4	178	..tgtatttatgtttt**ag**G..	7,596	..GTG**gt**atgt..	4,675	4	64,294	
5	132	..ttttttttttaaac**ag**A..	4,859	..CAG**gt**aatt..	2,888	5	17,7	
6	98	..tcttgtcttattgc**ag**T..	5,161	..ACA**gt**aagt..	3,738	6	666	
7	83	..cttttgtttcttata**ag**C..	4,579					

[a] For the splice junctions, intronic and exonic sequences are shown in small and capital letters, respectively. The 5' splice-donor and 3' splice-acceptor sequences are given in bold.
[b] Scores were calculated to references (Berg et al, 1988 and Penotti et al. 1991). Approximately 99% of sites have a score of 0-20 (3'acceptor site) and 0-11 (5'donor site).

5.3.6.3 Expression analysis

The expression pattern of the L40 and the ZPBP genes in various human tissues was determined by RT-PCR analysis. In order to examine the expression profile specific for each of the two overlapping genes, L40F2 was designed as forward primer in the first shared exon, whereas reverse primers L40R2 and L40R3 were designed within isoform-specific sequences within the first exons of L40 and ZPBP, respectively. As shown in Fig. 42, high level of expression of L40 is restricted to the retina and minor expression was detected in RPE, whereas expression of ZPBP was specific for the testis.

A radiolabelled PCR probe, amplified with the L40F and L40R primers, was hybridised to a virtual Northern blot containing ds cDNAs from several human tissues including bladder, heart, lung, skeletal muscle, retina and RPE. Hybridization failed to show clear signals for the expression of the L40 gene. Thus, the transcript length could not be determined (data not shown).
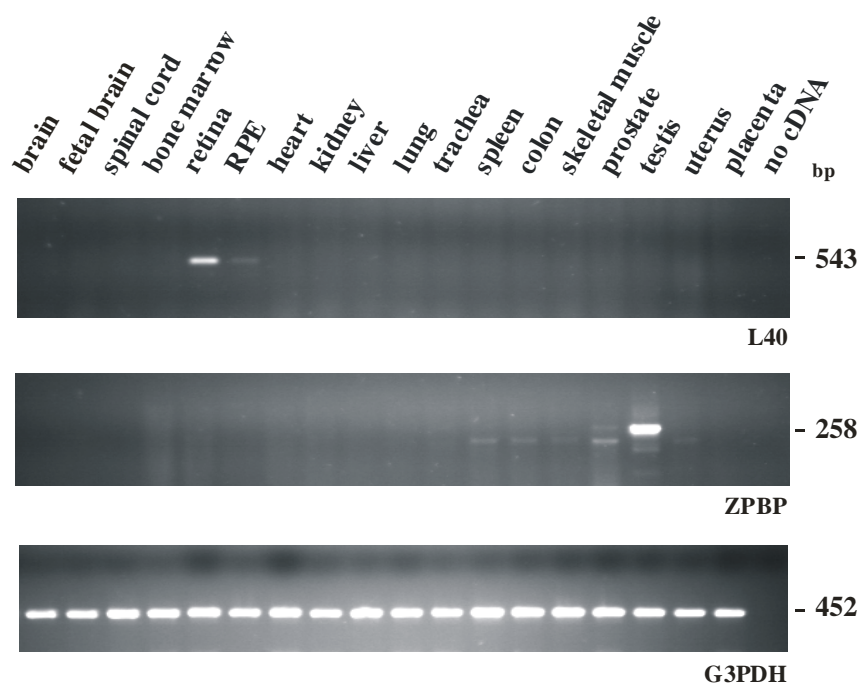
Fig. 42. Expression analysis of L40 and ZPBP genes.
A. Expression pattern of two overlapping genes was determined by RT-PCR in the panel of 18 human tissues, using primer pair specific for each gene. For those analysis forward primer was designed in the shared exon and used for both expression analysis, whereas reverse primer was designed in the first exons of L40 and ZPBP, which are specific for both genes (see Fig. 41A). Expression of L40 was determined using primers L40F2 and L40R2, whereas expression of ZPBP was determined using the same forward primer and reverse primer L40R3. Analysis showed that L40 mRNA is exclusively expressed in the human retina, whereas ZPBP expression was restricted to testis. The same panel was amplified with G3PDH specific primers as a control of cDNA quality.

5.3.6.4 Analysis of the cDNA sequence

Bioinformatic analysis of the L40 nucleotide sequence indicated that its largest open reading frame is 333 bp by exons 2-5 possibly encoding 111 amino acid residues (Fig. 43). The L40 gene overlaps ZPBP in coding sequences (Fig. 41A) and both genes share a common stretch of 100 amino acids (Fig. 43). ZPBP is one of several proteins that are thought to participate in secondary binding between acrosome-reacted sperm and the egg-specific extracellular matrix, the zona pellucida (McLeskey et al., 1998). It was determined that the deduced 351-amino acid ZPBP protein has no N-terminal signal peptide Katoh and Katoh (2003), but contains an N-terminal transmembrane domain and a long extracellular domain of almost 300 amino acids with which the L40 gene overlaps.

```
bp                                                                          aa
        | Exon 1
1       GGAAATGCGCTGACAAAAGGAGAACTTGCAGTGATTTTCTTTGAAAATCAAAGAAGAAGC
61      CACATGCCAAATAATGAAGGCAGCCACTACAGCTAGAAATGACAAAGAATTGGATTCCCC
121     CTAAAGCTTTCAGAATGAATACAGCTTGATTTTTGCTTAGTAAGACTCATTTTGGATTTC

                    | Exon 2
181     TGATCTCCAGAGCCATAAGCTTATCGTGAGCCTCATTATTATTATCAGTTCACAGCTCGA
241     TATCATGCAGCTCCCTGCAATAGCATTTATAATATTTCTTTTGAGAAGAAACTTCTTCAG
301     ATTTTAAGCAAACTGCTTCTTGACCTTTCATGTGAAATTTCCTTACTTAAGTCTGAATGC

                                                        |Exon 3
361     CATCGCGTTAAAATGCAAAGAGCTGGTTTGCAAAATGAATTGTTCTTTGCATTTTCAGTT
                   M   Q   R   A   G   L   Q   N   E   L   F   F   A   F   S   V        16

421     TCATCTCTAGACACTGAAAAAGGACCCAAGCGATGTACAGACCATAACTGTGAACCTTAC
         S   S   L   D   T   E   K   G   P   K   R   C   T   D   H   N   C   E   P   Y    36

             | Exon 4
481     AAAAGACTTTTTAAGGCTAAAAATCTCATAGAGAGATTTTTTAATCAACAAGTAGAAATT
         K   R   L   F   K   A   K   N   L   I   E   R   F   F   N   Q   Q   V   E   I    56

541     CTTGGCAGACGTGCAGAACAATTACCTCAAATATACTATATTGAAGGTACTCTCCAAATG
         L   G   R   R   A   E   Q   L   P   Q   I   Y   Y   I   E   G   T   L   Q   M    76

601     GTTTGGATTAATCGCTGCTTTCCAGGATATGGAATGAATGTCCAGCAACATCCAAAATGT
         V   W   I   N   R   C   F   P   G   Y   G   M   N   V   Q   Q   H   P   K   C    96

                 |Exon 5
661     CCTGAGTGCTGTGAGTCTTGCTCTGTTACCCAGGCTGGAGTGCAGTGACGCGATCTCGGC
          P   E   C   C   E   S   C   S   V   T   Q   A   G   V   Q   -                 111

721     TCAAGCAACCTCCACCCCCTCGGTTTAAGGGATTCTCCTGCCTCAGCCTCCCGAGTAGAT

                          | Exon 6
781     TACAGGCATGTGCCACCATGCCCAGTAGTTGGGACTTCCACTATGATGTTGAAAGGAAAT
841     GAAGAAAGGAGACAGCCTTATTTTGTTCCTAATCCTGGTGGGAAAGTTTCAAGTTTCTCA

            |Exon 7
901     ACACAATTTGCATCCTCTTTTTTGTTCTTAGCTTGGTGAGATACTTAGAACTTTTATTAA
961     TCTTTTCAATAAACAAGCTTTTGTTTCGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
1021    AAAAAAAAAAAAAAAA
```

Fig. 43. The complete cDNA sequence of L40 gene.
The 1037 bp long full-length cDNA sequence of the L40 gene is coding 111 aa. Amino acid sequence overlapping to ZPBP gene is underlined. Start and stop codons are boxed. Corresponding amino acid sequence is indicated bellow the nucleotide sequence.

This study revealed that two major mRNA species of the ZPBP and the L40 genes share three exons but differ in their 5'-untranslated regions. They are transcribed in a tissue-specific manner through the use of different promoters (Figure 40A). The first mRNA form is transcribed through the P1 promoter and is expressed in testis, whereas the second gene is transcribed through the P2 promoter and is expressed in the retina.

5.3.7 Identification of the single nucleotide polymorphisms (SNPs) in L33, L35 and L38 genes

Cloned genes with retinal specific expression are potential candidates for the retina disease genes. This was used as criterion for the selection of putative RNA genes: L33, L35 and L38 for identification of single nucleotide polymorphisms which are going to be used as markers for disease mapping and as target disorder, complex disease AMD was chosen. SNPs of these genes were identified by direct DNA sequencing or by denaturing high performance liquid chromatography (dHPLC). The genomic loci of these genes together with 5 Kb upstream (5' UTR) of the first exon and 5 Kb downstream (3' UTR) of the last exon were analysed. Primers were designed to amplify the known exons together with exon/intron junctions, intervening sequences (IVS) and UTRs (primers for each gene are listed in the Table 23, Appendix). For the primer design regions containing repeats were avoid. Location of the identified SNPs within the exon sequences was counted starting from the first nucleotide of the first exon which is not in correlation with nomenclature rules, since it is supposed that those genes are not coding proteins and thus do not have start codon.

5.3.7.1 Identification of the single nucleotide polymorphisms (SNPs) in L33 gene

As already described above, the L33 gene spans 13 Kb of genomic sequence on chromosome 15q23 and consists of five exons. In total, 16 fragments covering 23 Kb of the genomic locus were amplified from eight control individuals. Amplicons were directly sequenced or analyzed on the DHPLC. Analysis resulted in the identification of 25 SNPs. Of these, seven had allele frequency above 20% for the minor allele (marked with bold in Table 16), including one localized in the 5'UTR, one in the first intron of the gene, two in the forth intron and remaining four in the 3'UTR. Some of them were analysed further and the allele frequencies were determined in 48 control individuals. Although, frequencies of their minor allele were above 20% in eight controls, further analysis of the polymorphisms I1-120G>C and I4-1540G>A determined their lower frequency in the larger control samples, whereas for other nucleotide changes frequencies are higher then previously identified (Table 16). Nucleotide changes which are marked with the same colour are in strong linkage disequilibrium (Table 16).

Table 16. SNP card of the L33 gene

Part of the gene	The name of the fragment	Nucleotide change[1]	Allele frequency	Frequency of minor allele
5' UTR	**5'U-F1/R1**	**-532C>A**	**23/56**	**0.41**
IVS1	I1-F1/R1	IVS1+385G>A	1/16	0.06
IVS1	I1-F4/R4	IVS1+1274 A>G	1/12	0.08
IVS1	I1-F4/R4	IVS1+1492 G>T	1/12	0.08
IVS1	I1-F4/R4	IVS1+1562 C>T	1/12	0.08
IVS1	I1-F4/R4	IVS1+1647 A>G	1/12	0.08
IVS1	I1-F3/R3	IVS1-1672G>A	1/16	0.06
IVS1	I1-F6/R6	IVS1-1293C>T	1/12	0.08
IVS1	I1-F6/R6	IVS1-1149 G>C	1/12	0.08
IVS1	Ex2F/R	IVS1-165T>A	1/14	0.07
IVS1	**Ex2F/R**	**I1-120G>C**	**5/84**	**0.06**
IVS2	Ex2F/R	IVS2+18delG	1/14	0.07
IVS2	Ex3F/R	IVS2-24A>G	1/16	0.06
Exon 4	Ex4F/R	737C>A	1/16	0.06
Exon 4	Ex4F/R	812A>G	1/16	0.06
Exon 4	Ex4F/R	939C>T	1/16	0.06
IVS4	**I4- F2/R2**	**I4-1540G>A**	**11/88**	**0.13**
IVS4	**I4- F2/R2**	**IVS4-1487C>G**	**4/16**	**0.25**
3' UTR	3'U-F1/R1	+750G>T	1/16	0.06
3' UTR	3'U-F4/R4	+1130C>T	1/14	0.07
3' UTR	3'U-F4/R4	+1450T>C	1/14	0.07
3' UTR	**3'U- F2/R2**	**+2283A>T**	**19/92**	**0.21**
3' UTR	3'U- F2/R2	+2359G>A	1/14	0.07
3' UTR	**3'U- F2/R2**	**+2631C>T**	**3/14**	**0.21**
3' UTR	**3'U-F3/R3**	**+3419delC**	**3/14**	**0.21**

[1]Location of the SNP, counted from the first base of the reference genomic sequence.
[2]The same colours of the SNPs mean that they are in linkage disequilibrium.

5.3.7.2 SNP analysis of the L35 gene

L35 is a four exon gene, spanning 13 Kb of genomic sequence on chromosome 7p21. In total, 17 fragments were designed covering 23 Kb. Conditions for the PCR amplification could not be optimized for 5'U-F3/R3, 5'U-F4/R4 and Ex4F5/Ex4R1 fragments. The

remaining 14 fragments were amplified from eight control individuals. Generated amplicons were directly sequenced or run on the DHPLC. Analysis resulted in the identification of 13 SNPs. Of those, ten had allele frequencies of the minor allele of above 20% (marked with bold in Table 17), including one localized in the 5'UTR (-127G>A), one in the first intron of the gene (IVS1+7723T>C), three in the second exon (+1484A>G, +1775A>T, +4264C>T), one in the second exon (+364A>G), and the remaining four in the 3'UTR (+150G>A, +989C>G, +1519G>A, +2436C>T). Some of these were analysed further and included the sequence analysis of the fragments in 48 control individuals. Although, frequencies of their minor allele were above 20% in the eight controls, further analysis of the +1519G>A polymorphism resulted in a lower frequency in the larger control sample, whereas for other nucleotide changes frequencies are higher then previously identified (Table 17). Nucleotide changes which are marked with the same colour are in strong linkage disequilibrium (Table 17).

Table 17. SNP card of the L35 gene

Part of the gene	The name of the fragment	Nucleotide change[1]	Allele frequency	Frequency of minor allele
5' UTR	**Ex1F1/R1**	**-127G>A**	**11/32**	**0.34**
IVS1	I1-F1/R1	IVS1+1818T>C	1/16	0.06
IVS1	I1-F2/R2	IVS1+3496C>T	2/16	0.13
IVS1	I1-F2/R2	IVS1+3664G>A	1/16	0.06
IVS1	**I1-F3/R3**	**IVS1+7723T>C**	**13/20**	**0.65**
Exon 2	**Ex2F/R**	**+364A>G**	**40/80**	**0.50**
Exon 4	**Ex4-F6/R5**	**+1484A>G**	**3/14**	**0.21**
Exon 4	**Ex4-F6/R5**	**+1775A>T**	**5/14**	**0.36**
Exon 4	**Ex4F4/L35R1**	**+4264C>T**	**4/12**	**0.33**
3' UTR	**3'U-F1/R1**	**+150G>A**	**4/16**	**0.25**
3' UTR	**3'U-F2/R2**	**+989C>G**	**4/14**	**0.29**
3' UTR	**3'U-F2/R2**	**+1519G>A**	**12/96**	**0.13**
3' UTR	**3'U-F3/R3**	**+2436C>T**	**4/14**	**0.29**

[1]Location of the SNP, counted from the first base of the reference genomic sequence.
[2]Gray background means that SNPs are in linkage disequilibrium.

5.3.7.3 SNP analysis of the L38 gene

The L38 gene comprises five exos which span 25 Kb of genomic DNA on chromosome 12q21. In total, 14 fragments were designed to covering 35 Kb of the genomic locus (long repeat regions within the locus were the main reason for choosing the small number of fragments). Conditions for PCR amplifications could not be optimized for the Ex3F/R, I3-F1/R1 and I3-F2/R1 fragments. The remaining 11 fragments were amplified from eight control individuals. Generated amplicons were directly sequenced or run on DHPLC. Analysis resulted in the identification of 11 nucleotide changes including one deletion. Of these, only two (+1062C>T and +del:2186-2191) had allele frequencies for the minor allel above 20% (marked with bold in Table 18) and both of these are localized in the 3'UTR of L38. They were both analysed further and were screened in 48 control individuals. Although, frequencies of their minor allele were above 20% in eight controls, further analysis determined that the frequencies indeed were below 20% in the larger control samples (Table 18). It was also determined that these markers are not in linkage disequilibrium (Table 18).

Table 18. SNP card of the L38 gene

Part of the gene	The name of the fragment	Nucleotide change[1]	Allele frequency	Frequency of minor allele
5' UTR	5'U-F3/R3	-3733G>T	44/84	0.52
5' UTR	5'U-F2/R2	-2216C>T	1/16	0.06
5' UTR	5'U-F1/R1	-1205G>A	2/16	0.13
5' UTR	5'U-F1/R1	-1144insT	2/16	0.13
5' UTR	5'U-F1/R1	-858C>T	1/16	0.06
Exon 1	Ex1F/R	+50A>G	1/16	0.06
Exon 1	Ex1F/R	+67T>C	1/16	0.06
IVS1	I1-F1/R1	+3315A>C	1/16	0.06
Exon 4	Ex4F/R	+627(G>T)	6/12	0.5
3' UTR	**3'U-F1/R1**	**+1062C>T**	**7/96**	**0.07**
3' UTR	3'U-F2/R2	+2161C>T	1/16	0.06
3' UTR	**3'U-F2/R2**	**+del:2186-2191**	**17/88**	**0.19**

[1]Location of the SNP, counted from the first base of the reference genomic sequence.
[2]Gray background means that SNPs are in linkage disequilibrium.

6   DISCUSSION

The goal of this project was identification of the novel genes which are expressed preferentially in the human retina. The retina is highly-complex system specialized for the reception and processing of visual information and identification of the genes with expression restricted to this tissue will further our understanding of its physiology. In addition, expression analysis of the known genes which are associated with retinal degenerations showed that most of them are exclusively or abundantly expressed in the retina. Therefore, we decided to constructed cDNA library by SSH technique in order to identify novel retinal transcripts. Our efforts concentrated on the generation of a catalogue of the retina-specific genes, which will be discussed at the beginning of this chapter. Several of the retina-specific genes were selected for further analysis and challenges encountered in cloning and functional characterization of the these genes opened many questions and brought us to the following topics including i) alternative splicing, ii) mechanism of NMD and its possible involvement in the regulation of gene expression and RNAi as a mechanism to study gene expression (e.g. ABCC5 and its three newly identified isoforms), iii) regulatory ncRNA genes (e.g. L33, L35 and L38) and iv) genes overlapping on the same strand (e.g. L40), which are going to be discussed in the following paragraphs.

6.1   Identification of novel retinal transcripts

Towards the identification of the novel genes that are expressed exclusively in the human retina, cDNA library was constructed using suppression subtractive hybridization (SSH) technique (Diatchenko et al., 1996) which combines normalization and subtraction step in a single procedure in order to obtain enrichment of differentially expressed retinal genes of both high and low abundance. The normalization step equalized the abundance of cDNAs within retinal population, whereas subtraction step excluded the common sequences between retina and two other nonocular cDNA populations which were used.

In total, 1113 clones were isolated randomly from the retina SSH cDNA library and subjected for the further analysis. After correcting for redundancy, clustering of the clones showed that 162 of them derived from 92 partial transcripts representing unknown human genes, whereas 234 known human genes were represented by 766 clones (Table 1). The 92 transcripts were selected for the expression analysis which showed that sixteen genes were

expressed specifically in the retina, thirteen showed abundant expression in the retina, whereas for twelve genes expression was determined as neuronal (Table 2). Thus, evaluation of the cDNA library demonstrated enrichment of retinal transcripts and the presence of a large number of as yet unidentified tissue-specific genes. Of these, L25, L33, L35, L37, L38 and L40 transcripts with retina specific expression were selected for further analysis including cloning of the full-length cDNA and their functional characterization.

6.2   Cloning and characterization of L25 gene

The cloning and characterization of the L25 gene was begun by identification of two clones from retinal SSH cDNA library that derived from the chromosome 3q27, for which expression was determined as retina specific (Fig. 12). The fact that clones are mapping to the fifth intron of ABCC5 gene initiated analysis of the splicing events of this gene. Performed analysis resulted in the identification of the three novel splicing variants of ABCC5, named ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3, which share the first five exons with ABCC5 but have different 3' ends, and their length was determined as 2039 bp, 1962 bp, and 1887 bp, respectively.

ABCC5 protein belongs to the ATP-binding cassette (ABC) transporter superfamily, which represent the largest family of transmembrane proteins. In general, ABC proteins bind ATP and use the energy to drive the transport of various molecules across all cell membranes (Higgins 1992; Childs and Ling 1994; Dean and Allikmets, 1995). Most of the ABC transporters translocate a wide variety of substrates including sugars, amino acids, metal ions, peptides, proteins and a large number of hydrophobic compounds and metabolites across extra- and intracellular membranes. They usually move compounds into endoplasmatic reticulum, mitochondria and peroxisome or from the cytoplasm to the outside of the cell (Dean et al., 2001). Mammalian ABC genes are divided into seven subfamilies (Dean et al., 2001). One of them is ABCC subfamily, which contains 12 full transporters that perform different functions including ion transport, toxin secretion, and signal transduction (Borst et al., 2000). For example, ABCC1, ABCC2 and ABCC3 transport drug conjugates and other organic anions. Other transporters, like ABCC8 and ABCC9 proteins bind sulfonylurea and regulate potassium channels involved in modulating insulin levels (Dean et al., 2001). Nine members of ABCC subfamily are also known as multidrug resistance proteins (MRPs) (Borst et al., 2000). They play an important role in multidrug resistance, in which the cell is resistant

to several drugs in addition to the initial compound, which is a particular limitation to cancer chemotherapy (Dean et al., 2001). One of these proteins is ABCC5, also known as multidrug resistance protein 5 (MRP5), which was identified as cellular export pump for cyclic nucleotides with 3', 5'-cyclic GMP (cGMP) as a high-affinity substrate and cAMP as low affinity substrate (Jedlitschky et al., 2000). The cyclic nucleotides cAMP and cGMP are major second messengers in the cell. cGMP is formed from GTP by the action of guanylate cyclase, whereas cAMP is formed from ATP by adenylate cyclase. Both cyclases can be activated by a variety of intracellular and extracellular stimuli. Guanylate cyclase is stimulated by nitric oxide, which leads to a build up of cGMP in the cell, whereas adenylate cyclase is activated by the binding of ligands to membrane receptors (Beavo and Brunton, 2002). Afterwards, cGMP and cAMP are hydrolyzed by the phosphodiesterases to 5'-GMP and 5'-AMP, respectively, thus returning cyclic nucleotide concentrations to basal levels and making the cell susceptible to a new stimulus. In addition to metabolic degradation by phosphodiesterases, active cyclic nucleotide export as an elimination route has been observed in many tissues (Woods and Houslay, 1991; Billiar et al., 1992; Patel et al., 1995; Sager et al., 1998). In addition to a possible role in the regulation of intracellular cyclic nucleotide level, efflux of cyclic nucleotides may also have a signaling function, as suggested by reports showing biological effects of extracellular cGMP (Poulopoulou and Nowak, 1998; Touyz et al, 1997).

As already mentioned, ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3 transcripts are generated by alternative splicing of the ABCC5 mRNA. Alternative splicing of ABCC5 gene has already been reported. Suzuki et al. (1997) demonstrated existence of the short type of multidrug resistance protein homologue, SMRP. This protein composed of 946 amino acids, represents an N-terminally truncated ABCC5. Its coding sequence is starting from the seventh exon of the gene. There are also a number of reports which have demonstrated alternate splicing of ABCC family members. The earliest report is for ABCC1 gene, which contains 31 exons and several alternate spliced forms, which, interestingly, retain their reading frames (Grant et al., 1997). ABCC11 and ABCC12 also contain alternate forms, some of which retain their reading frame, but have lost certain functional domains (Yabuuchi et al., 2001; Bera et al., 2002).

Length of the exons 5a, 5b and 5c is 152 bp, 75 bp and 1.142 bp, respectively, and their splice donor and acceptor sequences correspond to the consensus splice junction sequences (Table 8). It has been shown that the usual size of the human exon is 300 nucleotides with the average internal exon measuring only 145 nucleotides in length (Venter

et al., 2001; Lander et al., 2001). Exon recognition in the context of the vastly larger introns is guided in part by partially conserved sequence elements at the exon-intron junctions, denoted 5' and 3' splice sites, which is a challenge, given their small size. Usage of the alternative 5' or 3' splice sites can result in exons of different sizes. In addition, inclusion or skipping of one or more exons is a common form of alternative splicing.

In general, alternative splicing is known as the major source of proteome diversity in humans. This is the process by which a single primary transcript yields different mature RNAs leading to the production of protein isoforms with diverse and even antagonistic functions (Modrek et al., 2003; Johnson et al., 2003). Alternative splicing leading to the production of transcripts with different properties from one locus has been proposed as an important mechanism underlying the complex phenotypes governed by mammalian genomes (Lander et al., 2001; Tupler et al., 2001). Annotation of the human genome has revealed that the majority of intron-containing transcripts are alternatively spliced (Modrek et al., 2003; Johnson et al., 2003). For example, estimations, based on analyses of expressed sequence tags (ESTs), suggested that the transcripts from 35% of human genes are alternatively spliced (Hanke et al. 1999; de Souza et al. 2000). However, this number is likely to be an underestimate, considering the fact that the human EST collection does not represent all protein coding sequences and that many alternative splicing events are very rare and occur only in a specific tissue at a specific time in development and/or under certain physiological conditions. Another study of 10,000 human genes, which are expressed in more than 50 tissues and cell lines, concluded that transcripts from at least 74% of all multi-exon genes were alternatively spliced (Johnson et al., 2003). These findings confirm predictions that majority of transcripts expressed from human genes are alternatively spliced. An extreme example of complex alternative splicing events is the gene encoding the Down syndrome cell adhesion molecule (Dscam) in flies which may generate more than 38,000 distinct isoforms, which is 2 to 3 times the number of predicted genes in the entire organism. There are 12 different variants of exon 4, 48 variants of exon 6, 33 variants of exon 9 and two variants of exon 17. However, each mature Dscam mRNA contains only one variant of these exons. Although nothing is known about regulation of the Dscam splicing, there must be some degree of control because Dscam transcripts always contain one variant exon at each position, yet never more than one variant (Schmucker et al., 2000).

It has been predicted, using the TMHMM method, that the cDNA sequences of three novel isoforms of ABCC5 encode putative protein with one transmembrane helix (Fig. 23). This amino acid stretch corresponds to the first transmembrane helix of the first hydrophobic

transmembrane domain (TMD1) of ABCC5 transporter. Typically the plasma membrane encoded ABC transporters are composed of two nucleotide binding domains (NBDs), also known as ATP-binding domains, and two transmembrane domains (TMDs), which contain 6-11 membrane-spanning α-helices and provide the specificity for the substrate (Dean et al., 2001). The NBD are highly conserved among species, whereas the membrane spanning domains TMDs are much more divergent (Higgins, 1992). However, hydropathy analysis of putative proteins encoded by these three isofoms did not predict transmembrane domains, suggesting that these transcripts may encode soluble proteins.

Sequences analysis of the human ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3 cDNAs showed that only exon 5a is highly conserved among different species including bovine, porcine and mouse, whereas for exons 5b and 5c this could not be shown (Fig. 24). The nucleotide sequences of functionally neutral alternative exons are not under selective pressure and therefore should not be conserved in distantly related species, whereas functionally relevant alternative exons may be conserved, which leads to the suggestion that exon 5a might be functionally relevant alternative exon of ABCC5 gene.

Computational analysis of the ABCC5_SV1 nucleotide sequence showed that inclusion of exon 5a produces a frame shift which results in introduction of stop codon that is located more than 100 bp upstream of the last exon-exon junction (Fig. 16B). Stop codons that are not located within the last exon and are followed by one or rarely two exon-exon junctions that reside no more than 50–55 nt downstream (Fig. 44) (Nagy and Maquat, 1998) are called premature termination codons (PTCs) (Hentze and Kulozik, 1999), whereas normal termination codons are usually located within the last exon and are not followed by an exon-exon junction.

Eukaryotes possess an mRNA surveillance system, termed nonsense-mediated mRNA decay (NMD), which scans newly synthesized mRNAs for faulty open reading frames (ORFs) and eliminates "imperfect messages" that contain PTCs, which are coding for nonfunctional or even harmful polypeptides (Hentze and Kulozik, 1999).
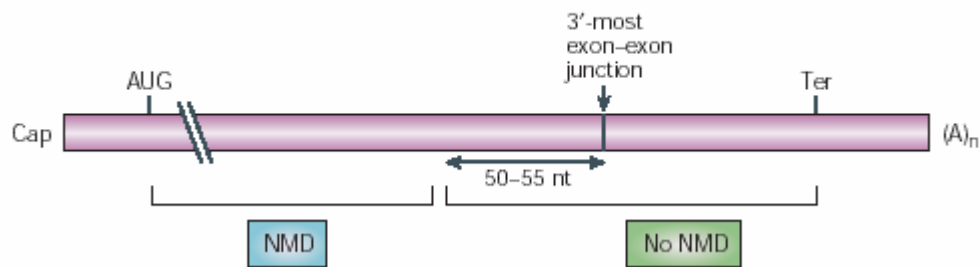
Figure 44. NMD and the 'position-of-an-exon–exon-junction' rule (Reproduced from Ref. Maquat, 2004)**.** Only the 3'-most exon–exon junction within a generic mammalian mRNA is shown. Mammalian mRNAs generally have an average of 7-8 splicing-generated exon-exon junctions (Lander et al, 2001; Venter et al., 2001). A premature termination codon (PTC) that is located in the region indicated in blue, which is followed by an exon-exon junction more than 50-55 nucleotides (nt) downstream, elicits nonsense-mediated mRNA decay (NMD), whereas a PTC that is located in the region indicated in green fails to elicit NMD. The normal termination codon (Ter) usually resides within the 3'-most exon (Nagy and Maquat, 1998).

In general, during pre-mRNA processing the splicosome deposits exon-junction complexes (EJCs) at sites of intron removal (Fig. 45) (Le Hir et al., 2001). With the pioneering first round of translation, the ribosome displaces the EJCs in its path and subsequently disassociates from the mRNA at the site of termination (Ishigaki *et al.*, 2001). If the ribosome reaches a stop codon which is located more than 50 nucleotides upstream of the final exon-exon junction, one or more EJCs will remain bound to the mRNA. Interactions between EJC proteins and release factors recruit a decapping enzyme that triggers rapid mRNA decay (Lykke-Andersen et al., 2000). Thus, a general rule for specifying whether a transcript will be targeted by the NMD pathway has been stated as follows (Nagy and Maquat, 1998): if an intron is located more than 50 nucleotides downstream of the stop codon, then the termination codon is recognized as premature and the transcript will be down-regulated by NMD.
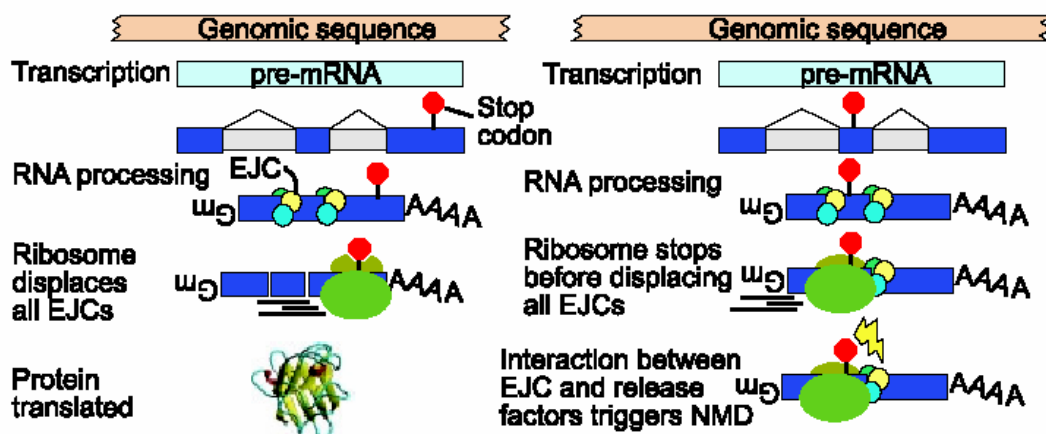
Figure 45. Nonsense-mediated decay is important surveillance mechanism. Normal translation involves displacement of all exon junction complexes (EJCs). Any remaining EJCs will trigger NMD (Reproduced from Ref. Green et al, 2003).

In mammalian cells, NMD was discovered in studies of β-thalassemias which are caused by introduction of the premature termination codons (Maquat et al., 1981; Kinniburgh et al., 1982). It was also identified that the PTC mutations of the yeast *ura3* gene destabilized the encoded mRNA (Losson and Lacroute, 1979). These findings led to the suggestion that NMD serves as an mRNA surveillance mechanism to eliminate mRNAs encoding truncated polypeptides that could act in a dominant-negative fashion or other deleterious mechanisms (Hentze and Kulozik, 1999. In addition, NMD also degrades natural substrates such as certain selenoprotein mRNAs (Moriarty et al., 1998; Sun et al., 2001). Finding that mouse embryos which are inactive in NMD resorb shortly after implantation provides the evidence for the importance of NMD (Medghalchi et al., 2001). Furthermore, blastocysts that are isolated 3.5 days post-coitum and that are inactive in NMD undergo apoptosis in culture after a brief growth period (Medghalchi et al., 2001).

Present study demonstrated that ABCC5_SV1 splice variant is target of NMD by using antibiotics puromycin and anisomycin which are interfering with the efficiency of translation and thus blocking NMD (Fig. 20) (Noensie and Dietz, 2001; Lei et al., 2001; Rajavel and Neufeld, 2001; Carter et al., 1995). Our experiments with human cell lines ARPE-19 and Y79 revealed that the SV1 transcript containing PTC increased in abundance relative to other isoforms after treatment with protein synthesis inhibitors. Moreover, it was shown that the amount of this PTC containing transcript in NMD-inhibited cells was

absolutely dependent upon protein synthesis, thus revealing classic characteristics of NMD (Carter et al., 1995). These findings suggest that exon 5a plays a role in regulation of ABCC5 expression facilitating general, conserved regulatory mechanism.

To confirm the correlation between expression levels of ABCC5 and its various short N-terminal isoforms RNAi method was used. Results of RNAi experiments demonstrated that reduction in SV1 levels was correlated with increase of ABCC5 mRNA amount in treated A-RPE 19 cell line, just opposite to the effect obtained by blocking NMD, thus clearly confirming correlation between those two transcripts and proposed hypothesis of the regulation of ABCC5 expression. For this purpose ARPE-19 cell line was treated with synthetic siRNA duplexes and the effect on the induction of RNAi was investigated. Our experiments showed that siRNA1 duplexes effectively silenced SV1, SV2 and SV3 isoform expression to a level of approximately 20% of normal. Other siRNAs tested were less efficient in silencing the various isoforms (Fig. 22). RNAi obtained for SV2 and SV3 transcripts by using siRNA2 duplexes was 30-50%, whereas inhibition obtained for SV3 mRNA using siRNA3 was approximately 50%. Silencing was measured between 24 to 48 hours posttransfection time since the most effective inhibition was observed at these time points (Fig. 21). Currently it is not known weather the lower effectiveness of specific RNAi duplexes arises from a local secondary structure phenomenon of the mRNA, protection of the mRNA by a cellular binding protein, translational silencing through miRNA effect (Doench et al., 2003), or an as yet unidentified feature in the sequence of the duplex. Alternatively, a minor error in the cDNA sequence or a polymorphism has to be considered, as already a single base change will render the duplex ineffective (Harborth et al., 2001).

Expression analysis of ABCC5 as well as SV1, SV2 and SV3 isoforms performed by RT-PCR, virtual Northern blot and quantitative RT- PCR showed high level of expression of these transcripts in retina compared to their expression levels in other human tissues, thus demonstrating tissue dependent expression of the ABCC5 as well as ABCC5_SV1, ABCC5_SV2 and ABCC5_SV3 splice variants (Fig. 18).

All those data strongly suggest that tissue specific splicing of exon 5a containing PTC and the subsequent degradation of the respective message via NMD rather seam to be a general mechanism of tissue-specific regulation of ABCC5 expression than just elimination of "imperfect" message. This could occur by the 5a exon facilitating regulated unproductive splicing and translation (RUST) (Fig.46) (Lewis et al., 2003). Just as the binding of transcription factors may turn on (or off) gene expression, the binding of splicing factors can turn on (or off) gene expression by splicing productive (or unproductive) isoforms. This mode

of regulation has been demonstrated for the splicing factor SC35, which auto-regulates its own expression through coupling alternative splicing and NMD (Sureau et al., 2001). Additional cases of RUST have been observed for ribosomal proteins in C.elegans (Mitrovich and Anderson, 2000).
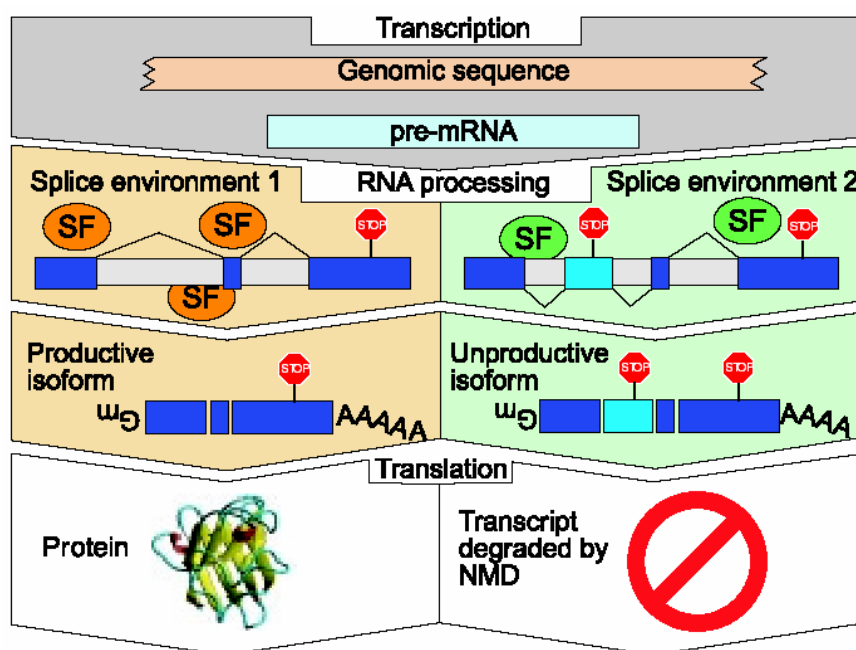

Figure 46. Regulated unproductive splicing and translation (RUST). Protein expression is controlled by regulated variations in the splice environment, the presence and abundance of splicing factors (SF). This determines whether productive or prematurely terminating (unproductive) mRNAs are spliced (Figure taken from Ref. Lewis et al., 2003).

In summary, present study provides an evidence for the importance of a complex alternative splicing pattern of ABCC5 gene involved in the regulation of its retina-specific expression and suggests NMD to be critical in the translational regulation of a ubiquitously transcribed ABCC5 mRNA. Tight regulation of intracellular cyclic nucleotide concentrations exported by ABCC5 is important, because excessive cGMP production in cells is toxic and may cause neurotoxicity (Li et al., 1997; Montoliu et al., 1999) or retinal degeneration (Heasley and Brunton, 1985; Bowes et al., 1990). It can be assumed that cellular environment regulates ABCC5 splicing and provides a checkpoint, after translation, for fine tuning and selecting the type of ABCC5 required. Although it is unknown at this point, such a checkpoint

may serve to allow a cell to determine the amount of ABCC5 transporter that is required and then modulate its expression accordingly.

Complex pattern of alternative splicing in a tissue-specific manner was reported for regulation of protein function of the MID1 (Winter et al., 2004). Identified splicing variants, which are highly conserved among species, cause loss-of-function effects via negative effects of C-terminally truncated proteins. Others carry PTC and are degraded by NMD. Lamba and co-workers (2003) showed NMD dependent down-regulation of alternatively spliced ABCC4 transcripts. Highly conserved exons of ABCC4 bearing PTC are suggested to play a role in regulation of the gene expression by facilitating translation re-initiation. A similar regulatory mode has been suggested for the expression of the AUF1 gene. Instead of introducing exons containing PTCs into the message, alternative splicing of the 3'UTR of the AUF1 gene has been shown. While the inclusion of an unspliced intron in the 3' part of this gene leads to the transcription of a stable message, a final exon-exon junction downstream of the termination signal is being produced by alternative splicing of this intron (Willson et al., 1999). Alternative splicing was also reported for neuronal nitric oxide synthase (nNOS) gene by alternate splicing of exon 1, thus regulating tissue and developmental stage-specific expression (Wang et al., 1999). Alternate splicing of CYP4F3, with inclusion of either exon 3 or exon 4, results in enhanced protein diversity to accommodate different substrates as well as impact tissue specific expression (Christmas et al., 2001).

The connection between alternative splicing and NMD seems to be the result of inappropriate inclusion or skipping of exons (e.g., FGFR2 transcripts that include both exons 8 and 9 (Jones et al., 2001)), which is perhaps an unavoidable side effect of the flexibility required for regulating splicing (Wagner et al., 2003). Thus, NMD provides surveillance for errors in alternative splicing. In other cases, alternative splicing-coupled surveillance can be used to downregulate transcript levels by diverting a larger fraction of these to be dispensed by NMD. Several alternative splicing factors can regulate their own synthesis by this mechanism (Wollerton et al., 2004).

6.2.1   Evidence for the coupling of alternative splicing and NMD

Lewis and co-workers recently suggested a genome-wide coupling of alternative splicing and nonsense-mediated mRNA decay in humans based on in-silico analysis of reliable alternative isoforms of known human genes (Lewis et al., 2003). In total, 16.163 well-

characterized human coding mRNAs from RefSeq and LocusLink were analysed. To detect alternative isoforms, 4.6 million EST sequences from dbEST were aligned to the genomic sequence. Application of the 50-nucleotide NMD rule revealed that 35% of alternative isoforms are candidates for NMD. Additionally, 35% of alternatively - spliced RefSeq genes were found to generate at least one NMD-candidate isoform. They also reported that the distribution of predicted polyadenylation signals in NMD-candidate splices is biased against regions just downstream of premature termination codons, undermining the likelihood that alternative polyadenylation stabilizes many of the NMD-candidate transcripts.

Though alternative splicing and NMD combine to make RUST a general mechanism of gene regulation, the impact of RUST is not yet fully known and the prevalence is still hypothetical. Preliminary phylogenetic sequence analyses hint that NMD might have evolved before alternative splicing. This allows the speculation that the predominant ancestral role of alternative splicing could possibly have been RUST, rather than the generation of protein diversity. Initial studies suggest that among the genes with NMD-candidate mRNA isoforms, certain functional classes are overrepresented, including chaperone proteins (Lewis et al., 2003).

It appears that NMD may function not only as a surveillance complex for the rare transcript but also as part of a pathway that controls expression of a large number of genes. The importance of NMD is underscored by its potential involvement in degrading many disease-linked gene transcripts. More complete understanding of the interplay between alternative splicing and NMD will provide insight possibly leading to effective treatments for some genetic diseases and a better understanding of gene regulation.

It seems unlikely that the complexity of an organism can be explained by the *one gene - one protein* hypothesis. Alternative splicing expands the coding capacity of the human genome, and in some instances related, diverse or antagonistic functions of alternatively spliced isoforms have been determined. This part of the thesis showed the importance of a complex alternative splicing pattern for the tissue-specific function of the ABCC5 transporter. In the majority of cases, however, the precise function of each isoform is not understood. Thus, what some consider noise might actually be crucial in facilitating the development of complex organisms from a limited number of genes. Developing a full catalogue of alternatively spliced transcripts and determining each of their functions will be a major challenge of the upcoming proteomic era.

6.2.2 RNA interference (RNAi) for gene knockdown

In general, RNA interference provides a new and reliable method to investigate gene function. This technology has many advantages over other nucleic-acid-based approaches, and therefore is currently the most widely used gene-silencing technique in functional genomics (Dorsett et al., 2004). RNAi is the process of sequence-specific posttranscriptional gene silencing triggered by double-stranded RNAs (dsRNAs) homologues to the silenced genes. RNAi was originally described in Caenorhabditis elegans (Fire et al., 1998; Timmons and Fire., 1998) and Drosophila melanogaster (Kennerdell and Carthew, 2000; Yang et al., 2000) as a mechanism to protect against invasion by foreign genes and has subsequently been demonstrated to be utilized by diverse eukaryotes such as insects, plants, fungi, and vertebrate (Agami, 2002; Cottrell and Doering, 2003; Hannon, 2002). The dsRNAs introduced or generated in cells are subjected to digestion with an RNase III-like enzyme, Dicer, into 21-25 nucleotide RNA duplexes, and the resultant duplexes, referred to as short interfering RNA (siRNA) duplexes, function as essential sequence-specific mediators of RNAi in the RNA-induced silencing complexes (RISCs) (Bernstein et al., 2001; Hammond et al., 2000). Elbashir et al. (2001) have shown that synthetic 21-nt siRNA duplexes can induce the sequence-specific RNAi activity in cultured cells without triggering the rapid and non-specific RNA degradation and translation inhibition. RNAi induction by synthetic siRNA duplexes appears to have paved the way for studying the molecular mechanism of mammalian RNAi, and also provided us with a powerful reverse genetic tool for suppressing the expression of a gene of interest in mammalian cells (Harborth et al., 2001).

6.3   Cloning and characterization of L33, L35 and L38 genes

The cloning and characterization of L33, L35 and L38 genes was initiated after expression analysis of the clones from the retina SSH cDNA library, which derived from these three genes, showed that their transcripts are expressed abundantly in the human retina (Fig. 12). For the identification of the full-length cDNA sequences of these genes full-length human retina cDNA library (local library ID: Jelena) was screened, as mentioned above, and sequence assembly of the isolated clones together with homologue clones from the public databases resulted in identification of the full-length cDNA sequences of the L33, L35 and L38 genes.

Full-length cDNA sequence of the L33 gene, which was identified as five exon gene on chromosome 15q23, is 1.600 bp long. Computational analysis of its full-length cDNA sequence predicted that longest identified open reading frames ORF1, ORF2 and ORF3 could encode 41, 42 and 43 amino acid residues, respectively (Table 10).

Full-length cDNA sequence of the L35, four exon gene which was localized to chromosome 7p21, was determined as 3.759 bp. Alternatively spliced transcript that lacks 141 bp long exon was also detected (Fig. 30). In addition, sequencing of the clones isolated from the full-length human retina cDNA library suggested generation of the transcript which is only 749 bp long due to the usage of an alternative polyadenylation signal. Analysis of the complete cDNA sequence of the L35 gene identified four open reading frames, ORF1 to ORF4 and of these, ORF4 as the longest one would encode 54 amino acids (Table 12).

Complete cDNA sequence of the L38 gene, which was localized to chromosome 12q21 and identified as five exon gene, is 1.164 bp long. Additional three isoforms for the L38 gene that are 1029 bp, 893 and 733 bp in size were also identified (Fig 36B). Computational analysis of the full-length cDNA sequence of this gene identified one open reading frame which encodes 34 amino acids residue (Fig. 40).

Considering the lengths of the full-length cDNA sequences of the L33, L35 and L38 genes which are approximately 1.6 Kb, 3.7 Kb and 1.1 Kb long, respectively, lengths of the longest open reading frames predicted for these three transcripts, which are 43, 54 and 34 amino acids, respectively, are very short. Moreover, in silico analysis of the putative proteins encoded by these three transcripts did not result in identification of the conserved domains or significant sequence similarity to any other known protein. These findings led us to suggest that L33, L35 and L38 genes may belong to the growing group of non-coding RNA (ncRNA) genes, producing functional RNA molecules rather than encoding proteins.

In general, for the sequences with short ORFs coding potential is less convincing, whereas many genes having long ORFs are obviously coding. Length of the ORF and coding potential alone is often insufficient to decide whether a gene is coding or non-coding, thus it is difficult to differentiate coding genes with short ORFs from ncRNA genes. For a small gene to be called a protein-coding gene, evidence may be that the ORF is significantly conserved in other related species. Comparative genome analysis may also provide a tool to clarify whether a locus produces ncRNA as opposed to a functional protein. For example, sequence analysis of the plant Medicago ENOD40 cDNA did not show significant coding potential, so it has been thought that this is ncRNA gene (Crespi et al., 1994). Subsequently, on the basis of comparative genome analysis, the ENOD40 transcript appeared to encode two small proteins

that consist of only 13 and 27 amino acids (Sousa et al., 2001). Another example is CsrB, 360-nucleotide bacterial regulatory ncRNA (Romeo, 1998), for which was originally thought to encode a 47-amino-acid protein.

Relatively few ncRNAs have been identified or described in detail. In most cases such RNAs either have not been predicted or, if detected, have been set aside in favour of focusing on the study of functional proteins. The ncRNAs have been documented by a number of research groups and provide some insight into this largely unexplored facet of coding potential in the genomes of higher organisms. Examples of well characterized ncRNAs include i) a synapse-associated 7H4 (Velleca et al., 1994), ii) brain-specific cytoplasmic RNA 1 (BC1), suggested to operate in the regulation of local translation-related processes in postsynaptic microdomains (Lin et al., 2001), iii) brain specific repetitive (Bsr) RNA and Ntab, which are expressed abundantly in rat brain, respectively (Komine et al., 1999; French et al., 2001), iv) Polar granule component (Pgc), required for germ cell formation in Drosophila (Nakamura et al., 1996), v) adapt33, oxidant-inducible RNA (Wang et al., 1996), vi) noncoding transcript in T cells (NTT) (Liu et al., 1997), vii) BORG, induced following exposure to bone morphogenetic proteins (Takeda, 1998), viii) H19, an abundant hepatic fetal-specific ncRNA, suggested to function as a tumor suppressor (Hurst and Smith, 1999), ix) roX1 and roX2, Xist and Tsix, which are involved in X chromosome dosage compensation in insects and mammals, respectively (Kelley and Kuroda, 2000; Lee et al., 1999), x) fibroblast growth factor-2 antisense (FGF-AS), differentially spliced to produce coding and noncoding RNAs in different tissues and regulates the expression of fibroblast growth factor-2 (Li and Murphy, 2000), xi) SCA8, within which an expansion of a CTG trinucleotide repeat causes neurodegenerative disorder spinocerebellar ataxia type 8 (Nemes, 2000), xii) DISC2, a single-exon noncoding structural RNA gene involved in the molecular etiology of schizophrenia (Millar et al., 2000), xiii) brain-specific dendritic BC200, expressed specifically in neuronal cells (Kuryshev et al., 2001), xiv) Bic, which is strongly upregulated in come B-cell lymphomas (Tam, 2001), xv) the Y-chromosome-specific TTY2 family, expressed specifically in human testis and kidney (Makrinou et al., 2001) and xvi) ncR-uPAR, which regulates human protease-activated receptor-1 gene during embryogenesis (Madamanchi, 2002). In addition, Scherer and colleagues (2003) showed that one of the loci associated with autism encodes a large ncRNA. There are also many ncRNAs documented at imprinted loci in animals (Charlier et al., 2001).

Over the past few years, RIKEN in Tokyo have taken a systematic approach to determining the mouse transcriptome by constructing the normalized cDNAs libraries from

different tissues. In total, 60.770 full-length cDNAs were sequenced and after computational analysis of the nucleotide sequences clustered into 33.409 transcription units. Of these, over 11.000 did not contain more than 100 coding nucleotide triplets. Since these transcripts do not code for a substantial open reading frame it was suggested that they belong to the group of ncRNAs, thus indicating that non-coding RNA is a major component of the transcriptome (Okazaki et al., 2002). A large number of all transcriptional units are represented by more than one independent clone and many have been shown to be differentially expressed in different tissues, and thus represent real transcripts. Whole-transcriptome analysis also showed that 42 % of the transcripts are spliced. In addition, over 2.400 pairs of overlapping sense-antisense transcripts were identified (Okazaki et al., 2002).

6.3.1   Non-coding RNA genes

The central dogma of molecular biology is that genetic information flows from DNA to RNA to proteins (DNA>RNA>protein). Usual interpretation of this statement is that genetic information flows from DNA to proteins via mRNA, which means that genes are generally synonymous with proteins and that genetic output is entirely or almost entirely transacted by proteins (Mattick, 2004). This statement is actually valid for the prokaryotes, whose genomes are almost entirely composed of closely packed protein-coding sequences with associated 5' and/or 3' cis-regulatory sequences. However, it has recently been found that prokaryotes do in fact contain a number of non-protein coding RNA genes, apart from those encoding rRNAs and tRNAs and these may number 200 or more in Escherichia coli, but account for no more than about 0.5% of the total number of genes and about 0.2% of the transcriptional output (Argaman et al., 2001). It is supposed that the same assumption is true for the multicellular organisms, although the proportion of protein-coding sequences is only a small minority of the genomic programming of complex organisms such as mammals (Fig. 47).

More recently, it became obvious that the hypothesis that most genetic information is expressed as proteins may be incorrect, at least in the higher organisms, which led to the re-definition of a gene as a ''transcription unit'' (Okazaki et al., 2002) or ''a complete chromosomal segment responsible for making a functional product'' (Snyder and Gerstein, 2003). It has been assumed that a very large proportion of the genome is in fact expressed, although protein-coding sequences constitute only about 1.6 % of the human genome (Wong

et al., 2001). It is estimated that approximately 97–98% of the transcriptional output of the human genome is non-protein-coding RNA (Mattick, 2001), based upon the fact that intronic RNA constitutes 95% of primary protein-coding transcripts (pre-mRNAs) (Lander et al., 2001; Venter et al., 2001) and on the observations which suggest that there is large number of ncRNA transcripts that do not contain substantial open reading frames, which may represent at least half of all transcripts. For example, detailed analysis of the regions such as β-globin cluster and various imprinted loci in mammals show that almost all of these regions are transcribed, mostly into ncRNAs (Ashe et al., 1997). Moreover, analysis of the whole chromosomes 21 and 22 using oligonucleotide arrays showed that the level of transcription is an order of magnitude higher than accounted for by the characterized and predicted exons (Kapranov et al., 2002). However, what fraction of the expressed RNAs serves some biological purpose is greatly unknown.

It is predicted that non-coding RNAs might function in the control of molecular genetic activity at various levels, which could occur via a variety of RNA–DNA, RNA–RNA and RNA–protein interactions (Mattick, 2003).

There are huge variations in estimates of the number of genes encoded by the human genome. An initial estimate of the number of protein-coding genes in the human genome was 30–40.000, on the basis of genome sequence analysis (Lander et al., 2001; Venter et al., 2001). In addition, Zhuo and colleagues (2001) suggested that the number of genes or transcription units in mammals ranged from 60.000 to 100.000, on the basis of cDNA cluster analysis. They also proposed that many, if not most of the genes do not code for protein, but rather function directly as structural, catalytic or regulatory RNAs. These estimations are based on cDNA cloning and EST sequencing of polyadenylated mRNAs, identification of conserved coding exons by comparative genome analysis and computational gene predictions. These three methods work well for large and highly expressed protein-coding genes which are evolutionary conserved, but they significantly underestimate the number of other genes (Eddy, 2001).
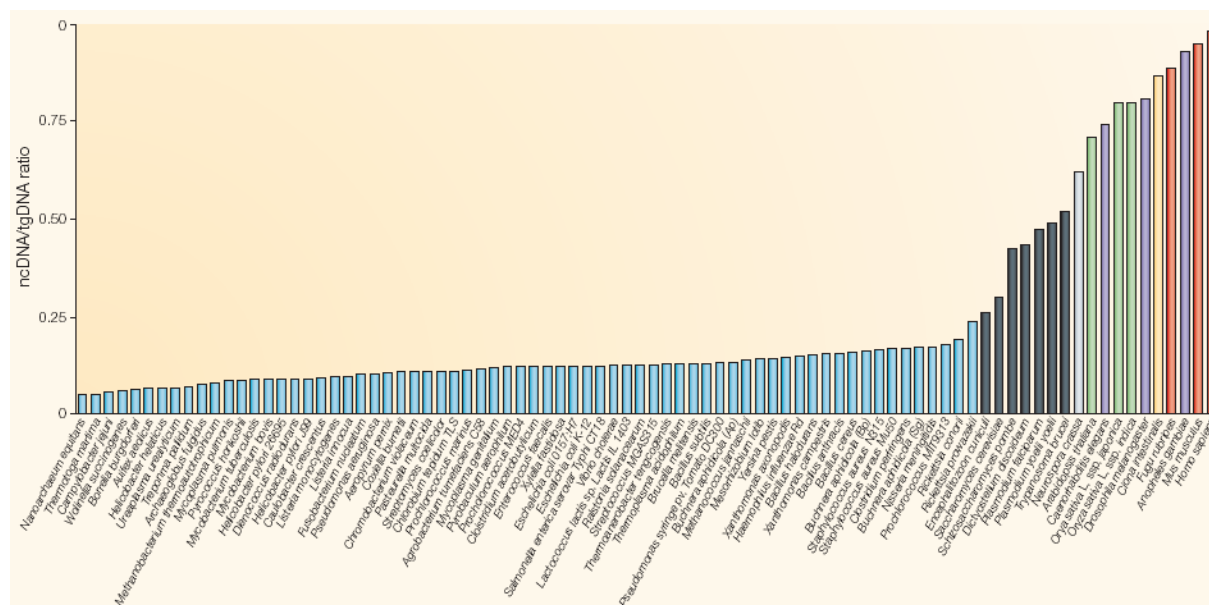
Figure 47. The ratio of non-coding to protein-coding DNA rises as a function of developmental complexity. Prokaryotes have less than 25% non-coding DNA, simple eukaryotes have between 25 and 50% non-coding DNA and more complex fungi, plants and animals have more than 50%, rising to approximately 98.5% non-coding DNA in humans — which also have a genome size that is three orders of magnitude larger than prokaryotes. The different colours represent prokaryotes (bacteria and archaea) (blue), simple eucharyotes (black), *Neurospora crassa* (grey), plants (green), non-chordate invertebrates (nematodes, insects) (purple), *Ciona intestinalis* (urochordate) (yellow) and vertebrates (red). ncDNA, non-coding DNA; tgDNA, total genomic DNA. Figure taken from Ref. Mattick, 2004.

It has been proposed that the actual number of ncRNA genes in mammals is probably in the order of tens of thousands (Mattick, 2003). The majority of these genes remain to be identified, considering the fact that RNA may be fairly unstable and that most biochemical analyses of cell fractions are not designed to detect ncRNAs. It should also be taken into account that the protocols for analysing the functions of ncRNAs are not well established so far. Non-coding RNAs have begun recently to be studied in systematic ways and it has been found that they have been regulated, i.e. expressed in a gender-, tissue- or cell-specific manner (Mattick, 2003). In addition, many ncRNAs are expressed at low levels in the cell (Kapranov et al., 2002; Okazaki et al., 2002).

Considering the fact that many transcripts are initiated from alternate promoters with alternate splicing and polyadenylation signals, the definition of a gene may need to be expanded to include a transcription cluster (defined as a separate locus producing one or more related transcripts from the same DNA strand, together with associated cis-acting regulatory elements) which may encode one or more related protein products (encoded by alternatively

spliced isoforms) and/or one or more different ncRNAs (such as transcripts that specify micro RNAs - miRNAs) (Lau et al., 2001) or small nucleolar RNAs (snoRNAs) (Cavaille et al., 2000; 2002) together with proteins encoded by the exonic sequences (Fig. 48). Even this definition is inadequate if we take in account that different mRNAs may be produced by trans-splicing between different primary transcripts (Pirrotta, 2002, Fire, 1999) and that overlapping transcripts may encode different products or a different repertoire of products with different functional consequences (Snyder and Gerstein, 2003). It is also worth noting that an overlapping antisense transcript may be involved in the cis-regulation of its sense pair and therefore may need to be considered part of the same ''gene''. The term ''gene'' is at best a concept to describe a sequence of genetic information, expressed either as RNA (per se) and / or protein (via mRNA), that has functional consequences (Mattick, 2003).
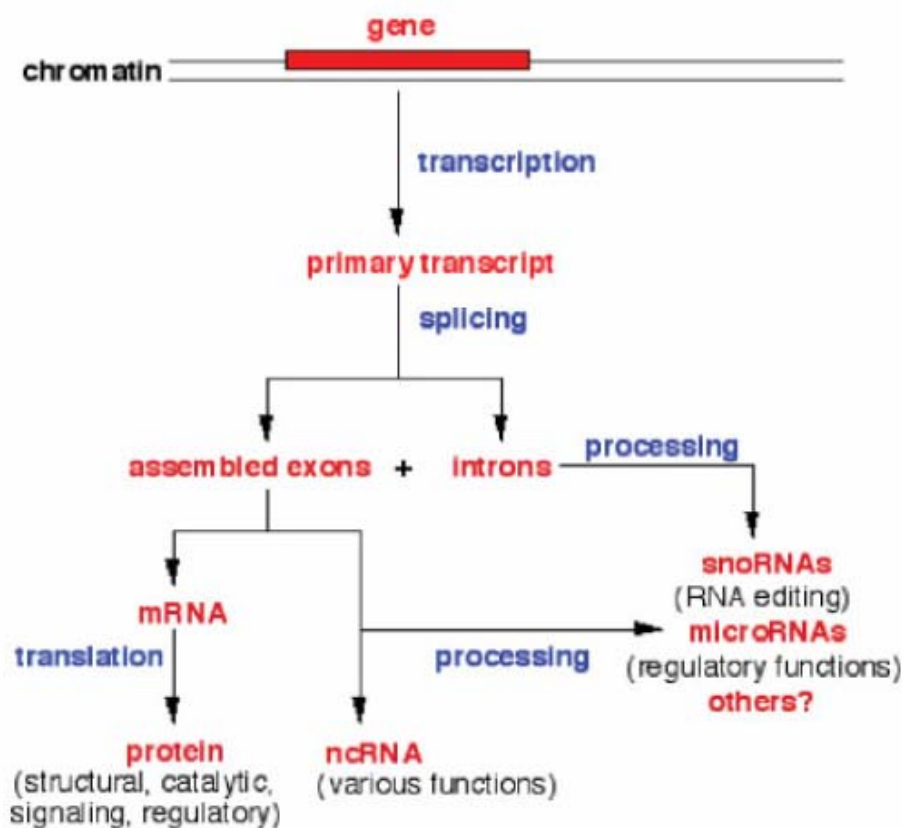

Figure 48. Modified view of the flow of genetic information in the higher eukaryotes. Primary transcripts may be (alternatively) spliced and further processed to produce a range of protein isoforms and/or ncRNAs of various types, which are involved in complex networks of structural, functional and regulatory interactions (figure taken from Ref. Mattick, 2003).

The RNA regulatory system might have been the essential prerequisite to both the evolution of multicellular organisms and the rapid expansion of phenotypic complexity into environments (Mattick, 1994; Mattick, 2001). This also suggests that the principal source of the evolutionary diversity of complex organisms is the regulatory architecture, which is primarily encoded in non-coding RNA genes and introns, and therefore also indicates that most of their genomes are dedicated to developmental programming and are under selective influences (Mattick, 2004).

RNA genes are immune to frameshift or nonsense mutations (Eddy, 2001). These are usually small and multicopy genes, which makes them difficult or even impossible targets for recessive mutational screens. Ridanpaa and colleagues (2001) provided an excellent example for human genetics. Cartilage-hair hypoplasia (CHH), a short-limbed dwarfism, was first described by Victor McKusick, almost 30 years ago (McKusick et al., 1965). Positional cloning failed to identify the gene linked to this disease, despite accurate genetic mapping. Ridanpaa et al. (2001) afterwards increased the resolution of the genetic map by almost an order of magnitude and sequenced the entire genomic region. Ten protein-coding genes, which were selected as potential candidate genes, were studied in further, but without success. Finally, CHH-associated mutations were discovered in the 267-nucleotide RMRP ncRNA gene, which produces the essential RNA component of the ribonucleoprotein endoribonuclease MRP (mitochondrial RNA processing). Interestingly, the only reason why RMRP was considered as a candidate gene was that its RNA has previously been isolated and its sequence was deposited in GenBank. Another example of human genetic disorder which has been mapped to a nuclear-encoded ncRNA candidate gene by positional cloning is autosomal-dominant dyskeratosis congenital. Here again, telomerase RNA was considered as candidate gene because its sequence was already deposited in the GenBank database. Mutations in telomerase RNA have been found in patients with dyskeratosis congenital (Vulliamy et al., 2001). Moreover, this gene was an obvious candidate, since an X-linked form of dyskeratosis had already been associated with nucleolar protein dyskerin which is known to interact with telomerase RNA. Several other ncRNAs have also links to human disease. For example, B-cell neoplasia-associated gene with multiple splicing (BCMS) is candidate for the tumor suppressor gene in the leukemogenesis of B-cell lymphoma (Wolf et al., 2001), expression of MALAT-1 is associated with early-stage non-small cell lung cancer (Ji et al., 2003), overexpression of DD3 gene is identified in prostate cancer (Bussemakers et al., 1999), translocation within DGCR5 gene is associated with DiGeorge syndrome (Sutherland et al, 1996).

The regulatory regions of higher organisms are generally much less conserved than protein-coding regions and it has been shown that the majority of human genomic polymorphisms occur outside of protein-coding regions (Venter et al., 2000). However, mutations or polymorphisms in promoters or other non-coding sequences that have strong effects on phenotype are very rare. This may be for the simple reason that, while a single base change can have fatal effects on the structure and function of a protein component which is critical for the protein function and it may only have subtle effects on cis- and trans-interactions in regulatory networks (Albert et al., 2000). Sequence differences in this controlled architecture probably include a large proportion of the genetic contributions to susceptibility to complex diseases (Mattick, 2003).

6.4   Cloning and characterization of L40 gene

The L40 gene potentially encodes a protein of 111 amino acid residues. Interestingly, the L40 gene locus is overlapping with ZPBP in the same orientation. It was determined that the deduced 351-amino acid ZPBP protein has no N-terminal signal peptide (Katoh and Katoh 2003), but contains an N-terminal transmembrane domain and a long extracellular domain of almost 300 amino acids, to which L40 gene overlaps. However, both genes have different promoter regions, which suggest that regulation of their expression is not related and that two genes have distinct tissue-specific elements required for the restricted expression. Expression analysis of L40 and ZPBP revealed that the L40 gene is exclusively expressed in human retina while the expression of ZPBP is testis-specific. Previous studies in the ZPBP gene showed that it is one of several proteins that are thought to participate in secondary binding between acrosome-reacted sperm and the egg-specific extracellular matrix, the zona pellucida (McLeskey et al., 1998).

The gamete interaction, mediated by the molecules specific for the mammalian sperm and egg, is crucial for the process of fertilization. The first interaction between sperm and egg occurs at the extracellular matrix which is surrounding ovulated eggs, known as zona pellucida. The zona pellucida is composed of three separate glycoproteins, zona pellucida glycoprotein 1 (ZP1), zona pellucida glycoprotein 2 (ZP2), and zona pellucida glycoprotein 3 (ZP3), which have different roles in gamete recognition (Brewis and Wong, 1999). Galactosyltransferase, sp56, zona receptor kinase and spermadhesins are proposed to interact with the zona pellucida proteins and are thought to participate in the exocytotic acrosome

reaction, which represents the primary binding between sperm and zona pellucida (Brewis and Wong, 1999). Digestion of the zona pellucida by these enzymes enables sperm to cross it, and at this time the proteins PH20, proacrosin, ZPBP, and Sp17 are thought to participate in secondary binding between the acrosome-reacted sperm and zona pellucida (McLeskey et al., 1998). Once the sperm passes through the zona pellucida, its plasma membrane is fused with egg plasma membrane in a process known to involve the egg integrin alpha 6 beta 1 and the sperm proteins DE and fertilin.

The unusual genomic structure of the ZPBP and the L40 gene and overlap between their coding regions indicated a possible link between their function in testis and retina, respectively, thus suggesting that L40 gene could be involved in interaction between retina cells.

6.4.1   Overlapping Genes

Veeramachaneni and colleagues identified over 774 gene pairs sharing a common locus in the human genome and 542 in the mouse genome. Although it was expected that evolutionary conservation of overlapping regions would be under stronger selection against mutations which would cause changes in both genes, reported identity between overlapping human and mouse orthologs was not higher than average. Information about identified human and mouse overlapping genes are stored in Overlapping Genes Database (http://posnania.cbio.psu.edu/research/overlapping_genes.html). The database facilitates the search for known overlapping genes by name, accession number, or LocusLink ID. A second form allows users to browse the set of overlaps by overlap type (coding-coding, 3'-UTR-coding, etc).

Overlapping genes occur frequently in viral and cellular prokaryotic genomes as well as in organelles such as mitochondria (Normark et al. 1983), but it was believed that they occur less frequently in eukaryotic genomes. Although their presence in human and genomes of other species was reported previously (Williams and Fried 1986; Lazar et al. 1989; Miyajima et al. 1989; Cooper et al. 1998; Bachman et al. 1999; Shintani et al. 1999; Misener and Walker 2000; Morelli et al. 2000; Slavov et al. 2000), until lately very little was known about their frequency and genome-wide distribution. Recent reports show that overlapping genes do indeed occur relatively frequently in human and other mammalian genomes

(Okazaki et al. 2002). Nevertheless, there is still little known about the origin, evolution, or cross-species conservation of overlapping genes.

6.5   SNP analysis in candidate genes – search for AMD susceptibility genes

Haplotype analysis to investigate associations between disease loci and multiple markers could be a valuable tool for the use of SNPs in complex disease such as AMD. In this study L33, L35 and L38 genes were selected for the analysis of single nucleotide polymorphisms (SNPs) since they are candidates for the retina disease genes concerning their retina specific expression. In the present study, the regions selected for the SNP identification were within the coding region and exon-intron boundaries of the gene, as well as 5 kb upstream and downstream of the 5´ and 3´ UTRs, respectively. Sequencing of 13 Kb long region of L33 gene resulted in identification of 25 novel SNPs, of which 7 are of high frequency and were identified in 5' and 3' UTR and intronic regions (Table 16). The same analysis of 13 Kb long genomic region of L35 gene resulted in identification 13 novel SNPs, of which 10 are of high frequency (Table 17). Identified SNPs are in UTR and intronic region of the gene as well as in exon regions. Sequencing of 25 Kb long region of L38 gene resulted in identification of 12 novel SNPs, of which only two are of high frequency and were identified in 3' UTR of the gene (Table 18). It has been suggested that, for identification of complex disease loci, SNPs in coding regions may be more useful than those in noncoding regions. However, our preliminary analysis showed that there is no clear advantage to SNPs in exons rather than in introns.

The process of performing an association study involves simply determining the frequency of a test factor (e.g. an SNP allele) in many patients and age and race matched controls (Brookes, 1999). Therefore, in order to determine if the SNPs identified in this study are related to the disease, comparison of allele frequencies between patients with AMD phenotype and healthy controls is required. If the SNP contributes an increased risk for disease occurrence, then its frequency should be found at higher level in individuals with disease compared to controls without disease (Brookes, 1999).

Among the best genetic examples of an association between a genetic marker and a complex disorder is that of Alzheimer's disease, in which the e4 allele of the apolipoprotein E gene (APOE) has been strongly correlated with the risk of this disease (Strittmatter and Roses, 1995). Possession of the e4 allele of apolipoprotein E is neither necessary nor sufficient to

cause Alzheimer's disease, but it represents a single, well defined and substantiated common genetic risk factor identified by positional approaches and association studies. Another example would be non-insulin-dependent diabetes mellitus (NIDDM) in which at least 16 common SNPs have been associated with this disorder. The strongest association validated in confirmation studies has been reported between a variant of the peroxisome proliferator-activated receptor-γ gene (PPARG) and diabetes mellitus (Altshuler, 2000). One more example would be identification of the polymorphism in the beta 2 adrenergic receptor which is associated with the blood pressure level (Bray et al., 2000).

7   REFERENCES

Agami R. RNAi and related mechanisms and their potential use for therapy. Curr Opin Chem Biol. 2002 Dec;6(6):829-34.

Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks Nature. 2000 Jul 27;406(6794):378-82.

Allikmets R, Shroyer NF, Singh N, Seddon JM, Lewis RA, Bernstein PS, Peiffer A, Zabriskie NA, Li Y, Hutchinson A, Dean M, Lupski JR, Leppert M. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. Science. 1997 Sep 19;277(5333):1805-7.

Allikmets R. Further evidence for an association of ABCR allelesyelinm with age-related macular degeneration. The International ABCR Screening Consortium. Am J Hum Genet. 2000 Aug;67(2):487-91.

Ambati J, Anand A, Fernandez S, Sakurai E, Lynn BC, Kuziel WA, Rollins BJ, Ambati BK. An animal model of age-related macular degeneration in senescent Ccl-2- or Ccr-2-deficient mice. Nat Med. 2003 Nov;9(11):1390-7.

Angleson JK, Wensel TG. Enhancement of rod outer segment GTPase accelerating protein activity by the inhibitory subunit of cGMP phosphodiesterase. J Biol Chem. 1994 Jun 10;269(23):16290-6.

Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. Curr Biol. 2001 Jun 26;11(12):941-50.

Arshavsky V. Like night and day: rods and cones have different pigment regeneration pathways. Neuron. 2002 Sep 26;36(1):1-3.

Arshavsky V, Bownds MD. Regulation of deactivation of photoreceptor G protein by its target enzyme and cGMP. Nature. 1992 Jun 4;357(6377):416-7.

Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev 1997;11:2494–2509.

Bachman NJ, Wu W, Schmidt TR, Grossman LI, Lomax MI. The 5' region of the COX4 gene contains a novel overlapping gene, NOC4. Mamm Genome. 1999 May;10(5):506-12.

Baylor D. How photons start vision. Proc Natl Acad Sci U S A. 1996 Jan 23;93(2):560-5.

Beavo JA, Brunton LL. Cyclic nucleotide research -- still expanding after half a century. Nat Rev Mol Cell Biol. 2002 Sep;3(9):710-8.

Bera TK, Iavarone C, Kumar V, Lee S, Lee B, Pastan I. MRP9, an unusual truncated member of the ABC transporter superfamily, is highly expressed in breast cancer. Proc Natl Acad Sci U S A. 2002 May 14;99(10):6997-7002.

Berg OG, von Hippel PH. Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. J Mol Biol. 1988 Apr 20;200(4):709-23.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001 Jan 18;409(6818):363-6.

Billiar TR, Curran RD, Harbrecht BG, Stadler J, Williams DL, Ochoa JB, Di Silvio M, Simmons RL, Murray SA: Association between synthesis and release of cGMP and nitric oxide biosynthesis by hepatocytes. Am J Physiol 1992, 262:C1077–C1082

Bird AC. The Bowman lecture. Towards an understanding of age-related macular disease. Eye. 2003 May;17(4):457-66.

Blencowe BJ. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci. 2000 Mar;25(3):106-10.

Borst P, Evers R, Kool M, Wijnholds J. A family of drug transporters: the multidrug resistance-associated proteins. J Natl Cancer Inst. 2000 Aug 16;92(16):1295-302.

Bowes C, Li T, Danciger M, Baxter LC, Applebury ML, Farber DB. Retinal degeneration in the rd mouse is caused by a defect in the beta subunit of rod cGMP-phosphodiesterase. Nature. 1990 Oct 18;347(6294):677-80.

Brennan P. Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? Carcinogenesis. 2002 Mar;23(3):381-7.

Brewis IA, Wong CH. Gamete recognition: sperm proteins that interact with the egg zona pellucida. Rev Reprod. 1999 Sep;4(3):135-42.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell. 1992 Oct 30;71(3):515-26.

Brocke KS, Neu-Yilik G, Gehring NH, Hentze MW, Kulozik AE. The human intronless melanocortin 4-receptor gene is NMD insensitive. Hum Mol Genet. 2002 Feb 1;11(3):331-5.

Brookes AJ. The essence of SNPs. Gene. 1999 Jul 8;234(2):177-86.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell. 1992 Oct 30;71(3):527-42.

Burns ME, Baylor DA. Activation, deactivation, and adaptation in vertebrate photoreceptor cells. Annu Rev Neurosci. 2001;24:779-805.

Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999 Dec 1;59(23):5975-9.

Carrero-Valenzuela RD, Klein ML, Weleber RG, Murphey WH, Litt M. Sorsby fundus dystrophy. A family with the Ser181Cys mutation of the tissue inhibitor of metalloproteinases 3. Arch Ophthalmol. 1996 Jun;114(6):737-8.

Carter MS, Doskow J, Morris P, Li S, Nhim RP, Sandstedt S, Wilkinson MF. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. J Biol Chem. 1995 Dec 1;270(48):28995-9003.

Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A. Identification of brainspecific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc Natl Acad Sci USA 2000;97:14311–14316.

Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/ Angelman syndrome region. Hum Mol Genet 2002;11:1527–1538.

Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G, Georges M. Humanovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. Genome Res 2001;11:850–862.

Chen C, Okayama H. High-efficiency transformation of mammalian cells by plasmid DNA. Mol Cell Biol. 1987 Aug;7(8):2745-52.

Christmas P, Jones JP, Patten CJ, Rock DA, Zheng Y, Cheng SM, Weber BM, Carlesso N, Scadden DT, Rettie AE, Soberman RJ. Alternative splicing determines the function of CYP4F3 by switching substrate specificity. J Biol Chem. 2001 Oct 12;276(41):38166-72.

Ciardella AP, Donsoff IM, Guyer DR, Adamis A, Yannuzzi LA. Antiangiogenesis agents. Ophthalmol Clin North Am. 2002 Dec;15(4):453-8.

Cooper TA, Mattox W. The regulation of splice-site selection, and its role in human disease. Am J Hum Genet. 1997 Aug;61(2):259-66.

Cooper PR, Smilinich NJ, Day CD, Nowak NJ, Reid LH, Pearsall RS, Reece M, Prawitt D, Landers J, Housman DE, Winterpacht A, Zabel BU, Pelletier J, Weissman BE, Shows TB, Higgins MJ. Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. Genomics. 1998 Apr 1;49(1):38-51.

Cottrell TR, Doering TL. Silence of the strands: RNA interference in eukaryotic pathogens. Trends Microbiol. 2003 Jan;11(1):37-43.

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A. enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. EMBO J. 1994 Nov 1;13(21):5099-112.

Das M, Harvey I, Chu LL, Sinha M, Pelletier J. Full-length cDNAs: more than just reaching the ends. Physiol Genomics. 2001 Jul 17;6(2):57-80.

Shuey DJ, McCallus DE, Giordano T. RNAi: gene-silencing in therapeutic intervention. Drug Discov Today. 2002 Oct 15;7(20):1040-6.
De Jong PT, Klaver CC, Wolfs RC, Assink JJ, Hofman A. Familial aggregation of age-related maculopathy. Am J Ophthalmol. 1997 Dec;124(6):862-3.

de Souza, S.J. et al. Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. Proc Natl Acad Sci U S A. 2000 Nov 7;97(23):12690-3.
Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. J Lipid Res. 2001 Jul;42(7):1007-17.

Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 2001 Jul;11(7):1156-66.

Delcourt C, Michel F, Colvez A, Lacroux A, Delage M, Vernet MH; POLA Study Group. Associations of cardiovascular disease and its risk factors with age-related macular degeneration: the POLA study. Ophthalmic Epidemiol. 2001 Sep;8(4):237-49.

Diatchenko L, Lukyanov S, Lau YF, Siebert PD. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods Enzymol. 1999;303:349-80.

Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol. 2002 Mar;3(3):195-205.

Dunn KC, Aotaki-Keen AE, Putkey FR, Hjelmeland LM.: ARPE-19, a human retinal pigment epithelial cell line with differentiated properties. Exp Eye Res. 1996 Feb;62(2):155-69.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature. 2001 May 24;411(6836):494-8.

Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science. 1999 Oct 15;286(5439):487-91.

Evans JR. (2001) Risk factors for age-related macular degeneration. Prog Retin Eye Res. Mar;20(2):227-53.

Felbor U, Stohr H, Amann T, Schonherr U, Weber BH. A novel Ser156Cys mutation in the tissue inhibitor of metalloproteinases-3 (TIMP3) in Sorsby's fundus dystrophy with unusual clinical features. Hum Mol Genet. 1995 Dec;4(12):2415-6.

Felbor U, Stohr H, Amann T, Schonherr U, Apfelstedt-Sylla E, Weber BH. A second independent Tyr168Cys mutation in the tissue inhibitor of metalloproteinases-3 (TIMP3) in Sorsby's fundus dystrophy. J Med Genet. 1996 Mar;33(3):233-6.

Felbor U, Suvanto EA, Forsius HR, Eriksson AW, Weber BH. Autosomal recessive Sorsby fundus dystrophy revisited: molecular evidence for dominant inheritance. Am J Hum Genet. 1997 Jan;60(1):57-62.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998 Feb 19;391(6669):806-11.

Fire A. RNA-triggered gene silencing. Trends Genet 1999;15:358–363.

Frank RN. "Oxidative protector" enzymes in the macular retinal pigment epithelium of aging eyes and eyes with age-related macular degeneration. Trans Am Ophthalmol Soc. 1998;96:635-89.

Frank RN, Amin RH, Puklin JE. Antioxidant enzymes in the macular retinal pigment epithelium of eyes with neovascular age-related macular degeneration. Am J Ophthalmol. 1999 Jun;127(6):694-709.

French PJ, Bliss TV, O'Connor V. Ntab, a novel non-coding RNA abundantly expressed in rat brain. Neuroscience 2001;108:207–215.

Fujii GY, de Juan E Jr, Humayun MS, Chang TS. Limited macular translocation for the management of subfoveal choroidal neovascularization after photodynamic therapy. Am J Ophthalmol. 2003 Jan;135(1):109-12.

Gale M Jr, Katze MG. Molecular mechanisms of interferon resistance mediated by viral-directed inhibition of PKR, the interferon-induced protein kinase. Pharmacol Ther. 1998 Apr;78(1):29-46.

Garcia-Blanco MA, Ghosh S, Lindsey-Boltz LA. The phosphoryl transfer reactions in pre-messenger RNA splicing in RNA. (eds. Soll, D., Nishimura, S., & Moore. P.B.) 109-123 (Pergamon, Amsterdam, 2001).

Gillard EF, Chamberlain JS, Murphy EG, Duff CL, Smith B, Burghes AH, Thompson MW, Sutherland J, Oss I, Bodrug SE, et al. Molecular and phenotypic analysis of patients with deletions within the deletion-rich region of the Duchenne muscular dystrophy (DMD) gene. Am J Hum Genet. 1989 Oct;45(4):507-20.

Gehrig A. Untersuchungen zu den molekularen Ursachen der X-gebundenen juvenilen Retinoschisis - vom Gendefekt zum Mausmodell. PhD Thesis. University of Wuerzburg, Institut for Human Genetics. 2003.

Gorin MB, Breitner JC, De Jong PT, Hageman GS, Klaver CC, Kuehn MH, Seddon JM. The genetics of age-related macular degeneration. Mol Vis. 1999 Nov 03;5:29.

Gottfredsdottir MS, Sverrisson T, Musch DC, Stefansson E. Age related macular degeneration in monozygotic twins and their spouses in Iceland. Acta Ophthalmol Scand. 1999 Aug;77(4):422-5.

Grant CE, Kurz EU, Cole SP, Deeley RG. Analysis of the intron-exon organization of the human multidrug-resistance protein gene (MRP) and alternative splicing of its mRNA. Genomics. 1997 Oct 15;45(2):368-78.

Graveley BR. Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 2001 Feb;17(2):100-7.

Green AL, Meek ES, White DW, Stevens RH, Ackerman LD, Judisch GF, Patil SR. Retinoblastoma Y79 cell line: a study of membrane structures. Albrecht Von Graefes Arch Klin Exp Ophthalmol. 1979 Oct;211(4):279-87.

Gregory-Evans K. What is Sorsby's fundus dystrophy? Br J Ophthalmol. 2000 Jul;84(7):679-80.

Gulcher J, Kong A, Stefansson K. The genealogic approach to human genetics of disease. Cancer J. 2001 Jan-Feb;7(1):61-8.

Guymer R. The genetics of age-related macular degeneration. Clin Exp Optom. 2001 Jul;84(4):182-189.

Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet. 1999 Jun;22(2):164-7.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet. 1999 Jul;22(3):239-47.

Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature. 2000 Mar 16;404(6775):293-6.

Hanke J, Brett D, Zastrow I, Aydin A, Delbruck S, Lehmann G, Luft F, Reich J, Bork P. Alternative splicing of human genes: more the rule than the exception? Trends Genet. 1999 Oct;15(10):389-90.

Hannon GJ. RNA interference. Nature. 2002 Jul 11;418(6894):244-51.

Harborth J, Elbashir SM, Bechert K, Tuschl T, Weber K. Identification of essential genes in cultured mammalian cells using small interfering RNAs. J Cell Sci. 2001 Dec;114(Pt 24):4557-65.

Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, Gerstein M. Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res 2002;12:272–280.

Hayward C, Shu X, Cideciyan AV, Lennon A, Barran P, Zareparsi S, Sawyer L, Hendry G, Dhillon B, Milam AH, Luthert PJ, Swaroop A, Hastie ND, Jacobson SG, Wright AF. Mutation in a short-chain collagen gene, CTRP5, results in extracellular deposit formation in late-onset retinal degeneration: a genetic model for age-related macular degeneration. Hum Mol Genet. 2003 Oct 15;12(20):2657-67.

Heasley LE, Brunton LL. Prostaglandin A1 metabolism and inhibition of cyclic AMP extrusion by avian erythrocytes. J Biol Chem. 1985 Sep 25;260(21):11514-9.

Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell. 1999 Feb 5;96(3):307-10.

Higgins CF. ABC transporters: from microorganisms to man. Annu Rev Cell Biol. 1992;8:67-113.

Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 2003;423:91–96.

Hisatomi O, Tokunaga F. Molecular evolution of proteins involved in vertebrate phototransduction. Comp Biochem Physiol B Biochem Mol Biol. 2002 Dec;133(4):509-22.

Hohjoh H. RNA interference (RNA(i)) induction with various types of synthetic oligonucleotide duplexes in cultured human cells. FEBS Lett. 2002 Jun 19;521(1-3):195-9.

Hurst LD, Smith NG. Molecular evolutionary evidence that H19 mRNA is functional. Trends Genet 1999;15:134–135.

Husain D, Ambati B, Adamis AP, Miller JW. Mechanisms of age-related macular degeneration. Ophthalmol Clin North Am. 2002 Mar;15(1):87-91.
Hyde SC, Emsley P, Hartshorn MJ, Mimmack MM, Gileadi U, Pearce SR, Gallagher MP, Gill DR, Hubbard RE, Higgins CF. Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature. 1990 Jul 26;346(6282):362-5.

Hyman L, Neborsky R. Risk factors for age-related macular degeneration: an update. Curr Opin Ophthalmol. 2002 Jun;13(3):171-5.

Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell. 2001 Sep 7;106(5):607-17.

Jacobson SG, Cideciyan AV, Bennett J, Kingsley RM, Sheffield VC, Stone EM. Novel mutation in the TIMP3 gene causes Sorsby fundus dystrophy. Arch Ophthalmol. 2002 Mar;120(3):376-9.

Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Muller-Tidow C. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003 Sep 11;22(39):8031-41.

Jedlitschky G, Burchell B, Keppler D: The multidrug resistance protein 5 functions as an ATP-dependent export pump for cyclic nucleotides. J Biol Chem 2000, 275:30069–30074

Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, Okano HJ, Yang YY, Darnell RB. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron. 2000 Feb;25(2):359-71.

Yates JR, Moore AT. Genetic susceptibility to age related macular degeneration. J Med Genet. 2000 Feb;37(2):83-7.

Johnson GC, Todd JA. Strategies in complex disease mapping. Curr Opin Genet Dev. 2000 Jun;10(3):330-4.

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. 2003 Dec 19;302(5653):2141-4.

Jomary C, Neal MJ, Jones SE. Increased expression of retinal TIMP3 mRNA in simplex retinitis pigmentosa is localized to photoreceptor-retaining regions. J Neurochem. 1995 May;64(5):2370-3.

Jones RB, Wang F, Luo Y, Yu C, Jin C, Suzuki T, Kan M, McKeehan WL. The nonsense-mediated decay pathway and mutually exclusive expression of alternatively spliced FGFR2IIIb and -IIIc mRNAs. J Biol Chem. 2001 Feb 9;276(6):4158-67.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet. 2000 Mar;66(3):979-88.

Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. Mol Cell. 2003 Jul;12(1):5-14.

Kan Z, Rouchka EC, Gish WR, States DJ. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. Genome Res. 2001 May;11(5):889-900.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR. Large-scale transcriptional activity in chromosomes 21 and 22. Science 2002;296:916–919.

Katoh M, Katoh M. Identification and characterization of human ZPBP-like gene in silico. Int J Mol Med. 2003 Sep;12(3):399-404.

Katsanis N, Worley KC, Lupski JR. An evaluation of the draft human genome sequence. Nat Genet. 2001 Sep;29(1):88-91.

Kelley RL, Kuroda MI. Noncoding RNA genes in dosage compensation and imprinting. Cell 2000;103:9–12.

Kennerdell JR, Carthew RW. Heritable gene silencing in Drosophila using double-stranded RNA.
Nat Biotechnol. 2000 Aug;18(8):896-8.

Kinniburgh AJ, Maquat LE, Schedl T, Rachmilewitz E, Ross J. mRNA-deficient beta o-thalassemia results from a single nucleotide deletion. Nucleic Acids Res. 1982 Sep 25;10(18):5421-7.

Klaver CC, Wolfs RC, Assink JJ, van Duijn CM, Hofman A, de Jong PT. Genetic risk of age-related maculopathy. Population-based familial aggregation study. Arch Ophthalmol. 1998 Dec;116(12):1646-51.

Komine Y, Tanaka NK, Yano R, Takai S, Yuasa S, Shiroishi T, Tsuchiya K, Yamamori T. A novel type of non-coding RNA expressed in the rat brain. Brain Res Mol Brain Res 1999;66:1–13.

Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 1987 Oct 26; 15(20):8125-48.

Kramer F, Mohr N, Kellner U, Rudolph G, Weber BH. Ten novel mutations in VMD2 associated with Best macular dystrophy (BMD). Hum Mutat. 2003 Nov;22(5):418.

Kuntz CA, Jacobson SG, Cideciyan AV, Li ZY, Stone EM, Possin D, Milam AH. Sub-retinal pigment epithelial deposits in a dominant late-onset retinal degeneration. Invest Ophthalmol Vis Sci. 1996 Aug;37(9):1772-82.

Kuryshev VY, Skryabin BV, Kremerskothen J, Jurka J, Brosius J. Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. J Mol Biol 2001;309:1049–1066.

Kwok PY, Deng Q, Zakeri H, Taylor SL, Nickerson DA. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics. 1996 Jan 1;31(1):123-6.

Langton KP, McKie N, Curtis A, Goodship JA, Bond PM, Barker MD, Clarke M. A novel tissue inhibitor of metalloproteinases-3 mutation reveals a common molecular phenotype in Sorsby's fundus dystrophy. J Biol Chem. 2000 Sep 1;275(35):27027-31.

Lamba JK, Adachi M, Sun D, Tammur J, Schuetz EG, Allikmets R, Schuetz JD. Nonsense mediated decay downregulates conserved alternatively spliced ABCC4 transcripts bearing nonsense codons. Hum Mol Genet. 2003 Jan 15;12(2):99-109.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.

Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 2001;294:858–862.

Lazar MA, Hodin RA, Darling DS, Chin WW. A novel member of the thyroid/steroid hormone receptor family is encoded by the opposite strand of the rat c-erbA alpha transcriptional unit. Mol Cell Biol. 1989 Mar;9(3):1128-36.

Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. 2001 Sep 3;20(17):4987-97.

Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nature Genet 1999;21:400–404.

Lei XD, Chapman B, Hankinson O. Loss of cyp1a1 messenger rna expression due to nonsense-mediated decay. Mol Pharmacol. 2001 Aug;60(2):388-93.

Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A. 2003 Jan 7;100(1):189-92.

Li Y, Maher P, Schubert D. Requirement for cGMP in nerve cell death caused by glutathione depletion. J Cell Biol. 1997 Dec 1;139(5):1317-24.

Li AW, Murphy PR. Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation. Mol Cell Endocrinol 2000;162:69–78.

Lim JI. Photodynamic therapy for choroidal neovascular disease: photosensitizers and clinical trials. Ophthalmol Clin North Am. 2002 Dec;15(4):473-8, vii.

Lin Y, Brosius J, Tiedge H. Neuronal BC1 RNA: co-expression with growth-associated protein-43 messenger RNA. Neuroscience. 2001;103(2):465-79.

van Lith-Verhoeven JJ, Cremers FP, van den Helm B, Hoyng CB, Deutman AF. Genetic heterogeneity of butterfly-shaped pigment dystrophy of the fovea. Mol Vis. 2003 Apr 24;9:138-43.

Liu AY, Torchia BS, Migeon BR, Siliciano RF. The human NTT gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4þ T cells. Genomics 1997;39:171–184.

Livesey FJ, Cepko CL. Vertebrate neural cell-fate determination: lessons from the retina. Nat Rev Neurosci. 2001 Feb;2(2):109-18.

Lori S. Sullivan and Stephen P. Daiger, Inherited retinal degeneration: exceptional genetic and clinical heterogeneity, Molecular medicine today, September, 1996**.**

Losson R, Lacroute F. Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc Natl Acad Sci U S A. 1979 Oct;76(10):5134-7.

Lykke-Andersen J, Shu MD, Steitz JA. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell. 2000 Dec 22;103(7):1121-31.

Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. 2004 Feb;5(2):89-99.
MacNeil MA, Masland RH. Extreme diversity among amacrine cells: implications for function. Neuron. 1998 May;20(5):971-82.

Madamanchi NR, Hu ZY, Li F, Horaist C, Moon SK, Patterson C, Runge MS. A noncoding RNA regulates human protease-activated receptor-1 gene during embryogenesis. Biochim Biophys Acta 2002;1576:237–245.

Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. 2004 Feb;5(2):89-99.

Maquat LE, Carmichael GG. Quality control of mRNA function. Cell. 2001 Jan 26;104(2):173-6.

Maquat LE, Li X. Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. RNA. 2001 Mar;7(3):445-56.

Maquat LE, Kinniburgh AJ, Rachmilewitz EA, Ross J. Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. Cell. 1981 Dec;27(3 Pt 2):543-53.

Maquat LE. Nonsense-mediated mRNA decay. Curr Biol. 2002 Mar 19;12(6):R196-7.

Makrinou E, Fox M, Lovett M, Haworth K, Cameron JM, Taylor K, Edwards YH. Tty2: a multicopy y-linked gene family. Genome Res 2001;11:935–945.

Marquardt A, Stohr H, Passmore LA, Kramer F, Rivera A, Weber BH. Mutations in a novel gene, VMD2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (Best's disease). Hum Mol Genet. 1998 Sep;7(9):1517-25.

Marth G, Yeh R, Minton M, Donaldson R, Li Q, Duan S, Davenport R, Miller RD, Kwok PY. Single-nucleotide polymorphisms in the public domain: how useful are they? Nat Genet. 2001 Apr;27(4):371-2.

Masland RH. The fundamental plan of the retina. Nat Neurosci. 2001 Sep;4(9):877-86.

Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell. 1999 Feb 5;96(3):307-10.

Mattick JS. Introns: evolution and function. Curr Opin Genet Dev. 1994 Dec;4(6):823-31.

Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol. 2001 Sep;18(9):1611-30.

Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. 2001 Nov;2(11):986-91.

Mattick JS, 2003. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. BioEssays 25:930–939.

Mattick JS. RNA regulation: a new genetics? Nature 2004. 5:316-323.

Maw MA, Corbeil D, Koch J, Hellwig A, Wilson-Wheeler JC, Bridges RJ, Kumaramanickavel G, John S, Nancarrow D, Roper K, Weigmann A, Huttner WB, Denton MJ. A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. Hum Mol Genet. 2000 Jan 1;9(1):27-34.

McBee WL, Lindblad AS, Ferris FL 3rd. Who should receive oral supplement treatment for age-related macular degeneration? Curr Opin Ophthalmol. 2003 Jun;14(3):159-62.

McConlogue L, Brow MA, Innis MA. Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine. Nucleic Acids Res. 1988 Oct 25;16(20):9869.

Mckusick VA, Eldridge R, Hostetler JA, Ruangwit U, Egeland JA. Dwarfism in the amish. ii. cartilage-hair hypoplasia. Bull Johns Hopkins Hosp. 1965 May;116:285-326.

McLeskey SB, Dowds C, Carballada R, White RR, Saling PM. Molecules involved in mammalian sperm-egg interaction. Int Rev Cytol. 1998;177:57-113.

Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM, Dietz HC. Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Hum Mol Genet. 2001 Jan 15;10(2):99-105.

Meyers SM, Greene T, Gutman FA. A twin study of age-related macular degeneration. Am J Ophthalmol. 1995 Dec;120(6):757-66.

Michaelides M, Hunt DM, Moore AT. The genetics of inherited macular dystrophies. J Med Genet. 2003 Sep;40(9):641-50.

Milam AH, Curcio CA, Cideciyan AV, Saxena S, John SK, Kruth HS, Malek G, Heckenlively JR, Weleber RG, Jacobson SG. Dominant late-onset retinal degeneration with regional variation of sub-retinal pigment epithelium deposits, retinal function, and photoreceptor degeneration. Ophthalmology. 2000 Dec;107(12):2256-66.

Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH, Porteous DJ. Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 2000;9:1415–1423.

Misener SR, Walker VK. Extraordinarily high density of unrelated genes showing overlapping and intraintronic transcription units. Biochim Biophys Acta. 2000 Jun 21;1492(1):269-70.

Mitrovich QM, Anderson P. Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans. Genes Dev. 2000 Sep 1;14(17):2173-84.

Miyajima N, Horiuchi R, Shibuya Y, Fukushige S, Matsubara K, Toyoshima K, Yamamoto T. Two erbA homologs encoding proteins with different T3 binding capacities are transcribed from opposite DNA strands of the same genetic locus. Cell. 1989 Apr 7;57(1):31-9.

Modrek B, Lee CJ. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet. 2003 Jun;34(2):177-80.

Montoliu C, Llansola M, Kosenko E, Corbalan R, Felipo V. Role of cyclic GMP in glutamate neurotoxicity in primary cultures of cerebellar neurons. Neuropharmacology. 1999 Dec;38(12):1883-91.

Morelli C, Magnanini C, Mungall AJ, Negrini M, Barbanti-Brodano G. Cloning and characterization of two overlapping genes in a subregion at 6q21 involved in replicative senescence and schizophrenia. Gene. 2000 Jul 11;252(1-2):217-25.

Moriarty PM, Reddy CC, Maquat LE. Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. Mol Cell Biol. 1998 May;18(5):2932-9.

Morrison M, Harris KS, Roth MB. smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans. Proc Natl Acad Sci U S A. 1997 Sep 2;94(18):9782-5.

Mousa SA, Lorelli W, Campochiaro PA. Role of hypoxia and extracellular matrix-integrin binding in the modulation of angiogenic growth factors secretion by retinal pigmented epithelial cells. J Cell Biochem. 1999 Jul 1;74(1):135-43.

Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci. 1998 Jun;23(6):198-9.

Nakamura A, Amikura R, Mukai M, Kobayashi S, Lasko PF. Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. Science 1996;274:2075–2079.

Nemes JP, Benzow KA, Koob MD. The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). Hum Mol Genet 2000;9:1543–1551.

Nichols BE, Drack AV, Vandenburgh K, Kimura AE, Sheffield VC, Stone EM. A 2 base pair deletion in the RDS gene associated with butterfly-shaped pigment dystrophy of the fovea. Hum Mol Genet. 1993 Aug;2(8):1347.

Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? Bioessays. 2003 Dec;25(12):1147-9.

Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nat Biotechnol. 2001 May;19(5):434-9.

Normark S, Bergstrom S, Edlund T, Grundstrom T, Jaurin B, Lindberg FP, Olsson O. Overlapping genes. Annu Rev Genet. 1983;17:499-525.

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 2002;420:563–573.

Patel MJ, Wypij DM, Rose DA, Rimele TJ, Wiseman JS: Secretion of cyclic GMP by cultured epithelial and fibroblast cell lines in response to nitric oxide. J Pharmacol Exp Ther 1995, 273:16–25

Penotti FE. Human pre-mRNA splicing signals. J Theor Biol. 1991 Jun 7;150(3):385-420.

Petrukhin K, Koisti MJ, Bakall B, Li W, Xie G, Marknell T, et al. Identification of the gene responsible for Best macular dystrophy. Nat Genet. 1998 Jul;19(3):241-7.

Pirrotta V. Trans-splicing in Drosophila. Bioessays 2002;24:988–991.

Player MR, Torrence PF. The 2-5A system: modulation of viral and cellular processes through acceleration of RNA degradation. Pharmacol Ther. 1998 May;78(2):55-113.

Poetsch A, Molday LL, Molday RS. The cGMP-gated channel and related glutamic acid-rich proteins interact with peripherin-2 at the rim region of rod photoreceptor disc membranes. J Biol Chem. 2001 Dec 21;276(51):48009-16.

Poulopoulou C, Nowak LM. Extracellular 3',5' cyclic guanosine monophosphate inhibits kainate-activated responses in cultured mouse cerebellar neurons. J Pharmacol Exp Ther. 1998 Jul;286(1):99-109.

Pulak R, Anderson P. mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev. 1993 Oct;7(10):1885-97.

Rajavel KS, Neufeld EF. Nonsense-mediated decay of human HEXA mRNA. Mol Cell Biol. 2001 Aug;21(16):5512-9.

Rattner A, Sun H, Nathans J. Molecular genetics of human retinal disease. Annu Rev Genet. 1999;33:89-131.

Rechtman E, Ciulla TA, Criswell MH, Pollack A, Harris A. An update on photodynamic therapy in age-related macular degeneration. Expert Opin Pharmacother. 2002 Jul;3(7):931-8.

Reenan RA, Hanrahan CJ, Barry G. The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing. Neuron. 2000 Jan;25(1):139-49.

Richard EG, Benjamin PL, Hillman T, Blanchette M, Lareau FL, Garnett AT, Rio DC and Brenner SE. Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. Bioinformatics. Vol. 19 Suppl. 1 2003, pages i118–i121

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996 Sep 13;273(5281):1516-7.

Ridanpaa M, van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, vanVenrooij W, Pruijn G, Salmela R, Rockas S, Makitie O, Kaitila I, de la Chapelle A. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. Cell. 2001 Jan 26;104(2):195-203.

Ridanpaa M, van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, vanVenrooij W, Pruijn G, Salmela R, Rockas S, et al. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. Cell 2001;104:195–203.

Roberts RG, Gardner RJ, Bobrow M. Searching for the 1 in 2,400,000: a review of dystrophin gene point mutations. Hum Mutat. 1994;4(1):1-11.

Romeo T. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. Mol Microbiol. 1998 Sep;29(6):1321-30.

Rychlik W, Rhoads RE. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res. 1989 Nov 11;17(21):8543-51.

Sachidanandam, R. et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001 Feb 15;409(6822):928-33.

Sager G, Orbo A, Pettersen RH, Kjorstad KE: Export of guanosine 3_,5_-cyclic monophosphate (cGMP) from human erythrocytes characterized by inside-out membrane vesicles. Scand J Clin Lab Invest 1996, 56:289–293 of MOAT-C and MOAT-D, new members of the MRP/cMOAT subfamily of transporter proteins. J Natl Cancer Inst 1998, 90:1735–1741

Scherer SW, Cheung J, MacDonald JR, Osborne LR, Nakabayashi K, Herbrick JA, Carson AR, Parker-Katiraee L, et al. Human chromosome 7: DNA sequence and biology. Science 2003;300:767–772.

Schork NJ, Cardon LR, Xu X. The future of genetic epidemiology. Trends Genet. 1998 Jul;14(7):266-72.

Schmidt S, Klaver C, Saunders A, Postel E, De La Paz M, Agarwal A, Small K, Udar N, Ong J, Chalukya M, Nesburn A, Kenney C, Domurath R, Hogan M, Mah T, Conley Y, Ferrell R, Weeks D, de Jong PT, van Duijn C, Haines J, Pericak-Vance M, Gorin M (2002) A pooled case-control study of the apolipoprotein E (APOE) gene in age-related maculopathy. Ophthalmic Genet 23:209–223

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell. 2000 Jun 9;101(6):671-84.

Shintani S, O'hUigin C, Toyosawa S, Michalova V, Klein J. Origin of gene overlap: the case of TCP1 and ACAT2. Genetics. 1999 Jun;152(2):743-54.

Seddon JM, Rosner B, Sperduto RD, Yannuzzi L, Haller JA, Blair NP, Willett W. Dietary fat and risk for advanced age-related macular degeneration. Arch Ophthalmol. 2001 Aug;119(8):1191-9.

Smith W, Assink J, Klein R, Mitchell P, Klaver CC, Klein BE, Hofman A, Jensen S, Wang JJ, de Jong PT (2001). Risk factors for age-related macular degeneration: Pooled findings from three continents. Ophthalmology. Apr;108(4):697-704.

Snow KK, Seddon JM. Do age-related macular degeneration and cardiovascular disease share common antecedents? Ophthalmic Epidemiol. 1999 Jun;6(2):125-43.

Snyder M, Gerstein M. Defining genes in the genomics era. Science 2003;300:258–260.
Soboleva G, Geis B, Schrewe H, Weber BH. Sorsby fundus dystrophy mutation Timp3 (S156C) affects the morphological and biochemical phenotype but not metalloproteinase homeostasis. J Cell Physiol. 2003 Oct;197(1):149-56.

Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M. Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. Mol Cell Biol. 2001 Jan;21(1):354-66.

Stohr H, Mah N, Schulz HL, Gehrig A, Frohlich S, Weber BH. EST mining of the UniGene dataset to identify retina-specific genes. Cytogenet Cell Genet. 2000;91(1-4):267-77.

Stone EM, Sheffield VC, Hageman GS (2001) Molecular genetics of age-related macular degeneration. Hum Mol Genet 10:2285–2292

Stone EM, Lotery AJ, Munier FL, Heon E, Piguet B, Guymer RH, et al. A single EFEMP1 mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. Nat Genet. 1999 Jun;22(2):199-202.

Sulisalo T, Sistonen P, Hastbacka J, Wadelius C, Makitie O, de la Chapelle A, Kaitila I. Cartilage-hair hypoplasia gene assigned to chromosome 9 by linkage analysis. Nat Genet. 1993 Apr;3(4):338-41.

Sureau A, Gattoni R, Dooghe Y, Stevenin J, Soret J. SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. EMBO J. 2001 Apr 2;20(7):1785-96.

Sutherland HF, Wadey R, McKie JM, Taylor C, Atif U, Johnstone KA, Halford S, Kim UJ, Goodship J, Baldini A, Scambler PJ. Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. Am J Hum Genet. 1996 Jul;59(1):23-31.

Suzuki T, Nishio K, Sasaki H, Kurokawa H, Saito-Ohara F, Ikeuchi T, Tanabe S, Terada M, Saijo N. cDNA cloning of a short type of multidrug resistance protein homologue, SMRP, from a human lung cancer cell line. Biochem Biophys Res Commun. 1997 Sep 29;238(3):790-4.

Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet. 2001 Dec;2(12):930-42.

Tabata Y, Isashiki Y, Kamimura K, Nakao K, Ohba N. A novel splice site mutation in the tissue inhibitor of the metalloproteinases-3 gene in Sorsby's fundus dystrophy with unusual clinical features. Hum Genet. 1998 Aug;103(2):179-82.

Takeda K, Ichijo H, Fujii M, Mochida Y, Saitoh M, Nishitoh H, Sampath TK, Miyazono K. Identification of a novel bone morphogenetic proteinresponsive gene that may function as a noncoding RNA. J Biol Chem 1998;273:17079–17085.

Tam W. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a non-coding RNA. Gene 2001;274:157–167.

Taylor JG, Choi EH, Foster CB, Chanock SJ. Using genetic variation to study human disease. Trends Mol Med. 2001 Nov;7(11):507-12.

Timmons L, Fire A. Specific interference by ingested dsRNA. Nature. 1998 Oct 29;395(6705):854.

Touyz, R. M., Picard, S., Schiffrin, E. L., and Deschepper, C. F. (1997) J. Neurochem. 68, 1451–1461

Travis GH, Sutcliffe JG, Bok D. The retinal degeneration slow (rds) gene product is a photoreceptor disc membrane-associated glycoprotein. Neuron. 1991 Jan;6(1):61-70.

Tuo J, Bojanowski CM, Chan CC. (2004) Genetic factors of age-related macular degeneration. Prog Retin Eye Res. 23(2):229-49.

Tupler R, Perini G, Green MR (2001) Expressing the human genome. Nature 409:832-833.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002 Jun 18;3(7).

Velleca MA, Wallace MC, Merlie JP. A novel synapse-associated noncoding RNA. Mol Cell Biol 1994;14:7095–7104. Lin Y, Brosius J, Tiedge H. Neuronal BC1 RNA: co-expression with growth-associated protein-43 messenger RNA. Neuroscience 2001; 103:465–479.

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.

Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature. 2001 Sep 27;413(6854):432-5.

Vuong TM, Chabre M. Deactivation kinetics of the transduction cascade of vision. Proc Natl Acad Sci U S A. 1991 Nov 1;88(21):9813-7.

Wagner EJ, Curtis ML, Robson ND, Baraniak AP, Eis PS, Garcia-Blanco MA. Quantification of alternatively spliced FGFR2 RNAs using the RNA invasive cleavage assay. RNA. 2003 Dec;9(12):1552-61.

Wang Y, Crawford DR, Davies KJ. adapt33, a novel oxidant-inducible RNA from hamster HA-1 cells. Arch Biochem Biophys 1996;332:255-260.

Wang Y, Newton DC, Robb GB, Kau CL, Miller TL, Cheung AH, Hall AV, VanDamme S, Wilcox JN, Marsden PA. RNA diversity has profound effects on the translation of neuronal nitric oxide synthase. Proc Natl Acad Sci U S A. 1999 Oct 12;96(21):12150-5.

Weber BH, Vogt G, Stohr H, Sander S, Walker D, Jones C. High-resolution meiotic and physical mapping of the best vitelliform macular dystrophy (VMD2) locus to pericentromeric chromosome 11. Am J Hum Genet. 1994 Dec;55(6):1182-7.

Wijnholds J, Mol CA, van Deemter L, de Haas M, Scheffer GL, Baas F, Beijnen JH, Scheper RJ, Hatse S, De Clercq E, Balzarini J, Borst P. Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs. Proc Natl Acad Sci U S A. 2000 Jun 20; 97(13):7476-81.

Wilkinson MF, Shyu AB (2002) RNA surveillance by nuclear scanning? Nat Cell Biol. Jun;4(6):E144-7.

Williams T, Fried M. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends. Nature. 1986 Jul 17-23;322(6076):275-9.

Wilson GM, Sun Y, Sellers J, Lu H, Penkar N, Dillard G, Brewer G (1999) Regulation of AUF1 expression via conserved alternatively spliced elements in the 3' untranslated region. Mol Cell Biol 19:4056-4064

Winter J, Lehmann T, Krauss S, Trockenbacher A, Kijas Z, Foerster J, Suckow V, Yaspo ML, Kulozik A, Kalscheuer V, Schneider R, Schweiger S. (2004) Regulation of the MID1 protein function is fine-tuned by a complex pattern of alternative splicing. Hum Genet. 114(6):541-52.

Wolf S, Mertens D, Schaffner C, Korz C, Dohner H, Stilgenbauer S, Lichter P. B-cell neoplasia associated gene with multiple splicing (BCMS): the candidate B-CLL gene on 13q14 comprises more than 560 kb covering all critical regions. Hum Mol Genet. 2001 Jun 1;10(12):1275-85.

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell. 2004 Jan 16;13(1):91-100.

Woods M, Houslay MD: Desensitization of atriopeptin stimulated accumulation and extrusion of cyclic GMP from a kidney epithelial cell line (MDCK). Biochem Pharmacol 1991, 41:385–394

Sun X, Li X, Moriarty PM, Henics T, LaDuca JP, Maquat LE. Nonsense-mediated decay of mRNA for the selenoprotein phospholipid hydroperoxide glutathione peroxidase is detectable in cultured cells but masked or inhibited in rat tissues. Mol Biol Cell. 2001 Apr;12(4):1009-17.

Yabuuchi H, Shimizu H, Takayanagi S, Ishikawa T. Multiple splicing variants of two new human ATP-binding cassette transporters, ABCC11 and ABCC12. Biochem Biophys Res Commun. 2001 Nov 9;288(4):933-9.

Dorsett Y, Tuschl T. siRNAs: applications in functional genomics and potential as therapeutics. Nat Rev Drug Discov. 2004 Apr;3(4):318-29.

Yang D, Lu H, Erickson JW. Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in Drosophila embryos. Curr Biol. 2000 Oct 5;10(19):1191-200.

Yoshida A, Yoshida M, Yoshida S, Shiose S, Hiroishi G, Ishibashi T. Familial cases with age-related macular degeneration. Jpn J Ophthalmol. 2000 May-Jun;44(3):290-5.

Yoshizawa T. Molecular basis for color vision. Biophys Chem. 1994 May;50(1-2):17-24.

Young JM, Cheadle C, Foulke JS Jr, Drohan WN, Sarver N. Utilization of an Epstein-Barr virus replicon as a eukaryotic expression vector. Gene. 1988;62(2):171-85.

Zhuo D, Zhao WD, Wright FA, Yang HY, Wang JP, Sears R, Baer T, Kwon DH, Gordon D, Gibbs S, et al. Assembly, annotation, and integration of UNIGENE clusters into the human genome draft. Genome Res 2001;11:904–918.

8   Appendix

Abbreviations

A	adenine
aa	amino acid
acc.no	accession number
AMD	Age-Related Macula Degeneration
Amp	Ampicilin
bp	base pair
BLAST	Basic Local Aligmnent Search
BLAT	Basic Local Aligmnent Tool
BSA	Bovine Serum Albumine
C	cytosine
°C	degrees Celsius
cDNA	complementary DNA
ddH2O	double destiled water
DEPC	Diethylpyrocarbonate
DHPLC	Denaturing High Presure Lliquid Chromathography
DMSO	Dimethyl Sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleosid triphosphate
ds	double stranded
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed Sequence Tag
EtBr	Ethidium bromide
ex	exon
FCS	Fetal Calf Serum
g	gram
G	guanine
GUSB	β-Glucuronidase
Hs.	*Homo sapiens*
hrs	hours
htgs	unfinished high-throughput sequences
IPTG	Isopropyl-beta-D-thiogalactopyranoside
IVS	Intervening Sequences
LTR	long terminal repeat
M	molar
MaLR	mammalian apparent LTR
Mb	Mega base pair
min	minutes
M.musculus	*Mus musculus*

MOPS	Morpholinopropanesulfonic acid
MW	Molecular Weight
NCBI	National Center for Biotechnology Information
NMD	Nonsense-Mediated mRNA Decay
nt	nucleotide
ORF	Open Reading Frame
PBS	Phosphate-buffered saline
PCI	Phenol Chlorophorm Isoamyl alcohol
PCR	Polymerase Chain Reaction
PAA	Polyacrylamid
pfu	plaque forming units
qRT-PCR	quantitative PCR
RNA	ribonucleic acid
Rnase	ribonuclease
R.norvegicus	*Rattus norvegicus*
RT	room temperature
RT-PCR	reverse transcriptase PCR
RPE	Retinal Pigment Epithelium
SDS	Sodium dodecyl sulphate
sec	seconds
ss	single stranded
SNP	Single Nucleotide Polymorphism
S.scrofa	*Sus scrofa*
T	Thymine
Ta	annealing temperature
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tm	melting temperature
TM	transmembrane
UTR	untranslated region
U	units
V	volt
VN	virtual Northern blot
vol	volumes
v/v	volume per volume
w/v	volume per weight

Table 19. General primers

Gene / Plasmid	Primer	Sequence
GUSB	**GUS B3**	ACTATCGCCATCAACAACACACTGACC
GUSB	**GUS B5**	GTGACGGTGATGTCATCGAT
GUSB	**GUS B6**	GATCCACCTCTGATGTTCAC
GUSB	**GUS B7**	CCTTTAGTGTTCCCTGCTAG
G3PDH	**G3PDH F**	ATCGTGGAAGGACTCATGACC
G3PDH	**G3PDH R**	AGCGCCAGTAGAGGCAGGGAT
λTriplEx2	**M13F**	CGCCAGGGTTTTCCCAGTCACGAC
λTriplEx2	**M13R**	AGCGGATAACAATTTCACACAGGA
pmalcxh	**pmal E**	GGTCGTCAGACTGTCGATGAAGCC
pGEX	**pGEX F**	GGGCTGGCAAGCCACGTTTGGTG
pGEX	**pGEX R**	CCGGGAGCTGCATGTGTCAGAGG
All mRNAs	**3'-RACE AP**	GGCCACGCGTCGACTAGTACD(T)$_{25}$(A,G,C)

Table 20. RT-PCR primers

Primer	Sequence	Primer	Sequence
L01F	TGGAAGATTTGGACCGTTAG	**L09R**	TCTGTCATTTGTAGTGGCTG
L01R	GGTGGCGGTAACATAGGAG	**L10F**	TTCCTTCCCTTCAGGTACTG
L02F	AATGAGCGTGTAGAAGAAAG	**L10R**	TCCACATTGATATAGGCCAC
L02F1	GAGCTCCTGCTCTACCTTG	**L11F**	CAATGCCTAAGAGACCAGAC
L02R	ATCTTTATGTGCTTCCAATG	**L11F1**	GAGGAGACGGAGCAAATGA
L02R1	CATCTTCATCAACTGCTCCA	**L11R**	TAAGAGCCAGTGAAAGGGTC
L03F	AGACTCAACTCCTCAGCAAC	**L12F**	TGTTTGGGATGTTGCTTTA
L03F1	CCTGCCTGCTATTGGTCAC	**L12R**	TATTTATGGTCGCTTTTGTC
L03R	TGAACTCTGGGTCTTTGTAG	**L13F**	TGTTTGGGATGTTGCTTTA
L04F	ACAAGAAACTGAAAAACTCG	**L13R**	TATTTATGGTCGCTTTTGTC
L04R	GGGACCATTCTTCTTTGAG	**L14F**	TGCTGTTTGTAAGGCTGGAG
L05F	GGTTCAGAATTTTCACAGCT	**L14R**	AGCGAATATGGTAGAAGGGA
L05F1	AGATAGTGCCTTTCAGCCCT	**L14F2**	CCTTTGAGTTGGTCCTTGAA
L05F2	CCTCAGCAGATGGTATTGGA	**L14R**	TCTCTTTGATGACACGGCTC
L05F3	GAGCACTTCATCAGGTTCA	**L14R1**	TCCATCTCTCCATCTTCTGC
L05R	CTGGTGTTTGTGTCCTATTC	**L15F**	GTGATGCTCTGCTTGATAAG
L05R1	TGACTGAACTGGGAACACTG	**L15R**	TAGAACAGGAATGACTTTGC
L05R2	TCCAATACCATCTGCTGAGG	**L16F**	TGAGGAAGAGGAAGCAGAAC
L05R3	CAAAAGTCTCAGTTCCTGTGT	**L16F1**	TGATCAGCGGTCAGCAACTCC
L06F	CAGTTTGAGGAGCGACAGC	**L16F2**	GCGTTTTGTGGTAATGGA
L06R	GCCTCAGTGTATTGGTCGG	**L16R**	CTGAACCTATCCATCCCAC
L07F	CACTTCATCCTAAACACAGC	**L16R1**	CAGTATGGATTAAAACCATTG
L07R	GCTGAATGACTGGAAATGT	**L16R2**	ATGTGGCGGTAACCTTGTTG
L08F	TATGGTGCCGAGTGGTAAC	**L17F**	GATAGAAGAATAGGGGCAAG
L08R	TACTGCTATGGTCACGGTTC	**L17F1**	CTTCACCTTCCACTTCAGAC
L09F	AGTCTATGTTCAGTTGTGGC	**L17R**	GAATGGACAACACAAATAGG

Primer	Sequence	Primer	Sequence
L18F	GTAGAAGAGTTGAAGGCTGC	L34F1	TAAGAGAGCAAGGAAGTGTC
L18R	CATACAAGGAGGAAGACTGG	L34R	AGAATGAAGAGGTTGAGTCC
L19F	GGAGCACTTGGGTTTAGAG	L34R1	AGTTCTCCAAGTCTCACCAG
L19R	CCCTCAACACTCTTCACAG	L35F	ACCTTTGCCTTCTCTTGCTC
L20F	CTGATGGAAACCGCTGAAG	L35F1	CTGGCTTCGGATAGTGTTTG
L20R	CTTCAATGGCTTCCTCTGC	L35R	CAAACACTATCCGAAGCCAG
L21F	TTTACTTGCTGTGGTTGTGG	L35R1	CAGTCATTACCCCATTCTCC
L21F1	CATTCACTGTCTCTGTTGGT	L36F	TTAGCACATACATCCACTTG
L21F3	AAGTTACTGAAAAGGAGGTAGAAA	L36F2	CAAGCTGCTTTGGTTCTGGA
L21R	TCTGAGGCATAGCAAAACAC	L36R	GTTTGGAGATTACTGGATTG
L21R1	GTGAGAGCCTTTATTCCTGG	L37F	AACTGGCTACTCTTTCAACG
L21R2	ATGCAGTAACCTCAGGCTTC	L37R	CTTCATCCTCATTGTCTTCG
L21R3	CCCAAAAGAAATGAGAAGCCAA	L38F	TCCACAACGATGCTTTCTA
L22F	TTCTTGTAGTGTCTGCTTGC	L38R	AACAAACCTTCACACAAATC
L22R	AGCCTTATGTCCTTATTCAG	L38R1	AATGAGCAGTGATCTGATCC
L23F	AGTTCTCCTGATGTTGTTCC	L39F	TAACACCCCAGCCATTGATT
L23R	GCCCACACATCATAAGGTAC	L39F2	AAGAAAAGGAGATTAGGGTACAG
L24F	TACAGTCCTCCTTTGCTTGG	L39F3	TGTCATGCTAGGAAAGTCGAAC
L24F1	GTTGTGGGTGACTCTTTTGG	L39F4	CAAGGTAGGTGGGAGAAGGTG
L24R	CTGTGCTTTCGTAGGGTTTC	L39R	CAGGAGGAGCACAAGCAAG
L24R1	GTGAGGTCTGTTCCAATAGC	L39R2	GGAAGTTTATTAGGAAGAATTGG
L25F	GAAAAGACCCAGAAGGATG	L39R3	CAATACAAGATCCGCATAGCAG
L25F1	AGAAGAGCTGAATGAAGTTG	L40F	AAGAACAAAAAGAGGATGC
L25F2	GGTTGTCCTGGAGCAGGG	L40F1	CAGGATATGGAATGAATGTC
L25F3	AGTGTGAGGGAGAGAACCAG	L40F2	CACATGAAAGGTCAAGAAGC
L25R	AGCCATCTAACAGGTCATC	L40F3	TAATTGTTCTGCACGTCTGC
L25R1	TTCAATGCCCAAGTCAGTG	L40R	TCCACTATGATGTTGAAAGG
L25R2	GTCTCCGACGGTTGTTTAAC	L40R1	ATCTCACCAAGCTAAGAACA
L25R4-Mus	CTGGAGTACAATGCATCAGG	L40R2	AAGGCAGCCACTACAGCTAG
L25R5-Mus	CTGGAGTACAATGCATCAGG	L40R3	CCATCCTCCTCTTTATCTCC
L26F	GATTCCAGCCCAACAAACC	L40R4	GCTTCTTGACCTTTCATGTG
L26R	CGGATACCCTTGAGACAGC	L41F	CAGGCAATGTCTTCAGCAG
L27F	CAGGCAAGAAGGTGTTAGTG	L41R	ACCTCAAATGCCAATACCAG
L27R	GGTCTGGAATGAATGGAAAG	L42F	GTCTACTGGCACAAACCTTC
L28F	GCCCTACAGAATGAATACAC	L42R	TCCTTGTATCGTCATTCCAG
L28R	TATCATTCACAAGTTTCCCA	L43F	AGATAACAAAGGAGCGAGAG
L29F	AGAAAGAGATACAGAATGCC	L43R	CCACCTGTTCACTAATGTCC
L29R	TAATCTTTTTGGCTTTGGTG	L44F	ATTATGAGAGCCAAAGGAGC
L30F	CTGTTTGTCACCCTTCTTGC	L44R	ATCCCAGAGAAAGTGTTGAG
L30F1	GGCTGCTACCATTAAACAAC	L45F	GAGGCTGCTGAATAATGTAG
L30F2	TGGCTATGAAGATGCTGGAGA	L45R	ACTGGATGCCCTGAATGAC
L30R1	ACTGGGTTGTAAAGGCATCAT	L46F	GACCTCACAGACATCGTTTC
L30R	CAGTGAGAAGTGCCATAAGC	L46R	CTATGGGTGTGGTGAGTAAC
L31F	TCTGACACCAAATGAGGACC	L47F	CAACTTCAGAGACAACCCAC
L31R	GCTCCATTTCTGCTTCCAAC	L47R	CCGTATTTATGGAGATTGGT
L32F	GTGATCCAGAAAGTTGATGG	L48F	CCATGAAGAAGATGACATTG
L32R	CATTACCAAGACAACCCCTC	L48R	AGACAGGATCTCTGTGCTAG
L33F	CTTCCACTTCTTCCATCTGC	L49F	TCAGTTCTATGTTCCAGCAG
L33F1	CCAGGAAAGAAGGGCACAT	L49R	TAAGAAGCCAACAACTCCAG
L33R	TGCTGGTTTCACAATACTGC	L50F	GAGATGAAGGTGAAGGTGTC
L33R1	TTCTTGTGTCTTCATGTGCC	L50R	CATAGAAAGGTGAAGTGCTG
L34F	GTCTCCATTTCCATTCACAG	L50R1	GGCCATAGGCAATGAGGGCTC

Primer	Sequence	Primer	Sequence
L51F	GCTCCCAATGGAGACTTTTG	**L74F**	AAACTTGGCTTGATTGCTTC
L51R	ACAGTGATGGCTTCCTTCTC	**L74R**	AAATACCCAGGCACTTCATC
L52F	CCACTTTGATGTTGCGGAG	**L75F**	TGGAGGACTATTCACAGGAG
L52R	AGAAGCCCAAGAGTCCCTG	**L75R**	TATCCTCTTTCCAACGGGTG
L52R1	TTCCTCGTCTGTAATGGTGCC	**L76F**	TAGAAGGCACAGAGAGAATG
L53F	AAAGTAGGACACGGTTGAGG	**L76R**	CTTGTCTGGGATGGAAACTC
L53R	CCTCCAACACACATACCAC	**L77F**	ATGTTGAAAATACTCCCTGC
L54F	TTATCTGTCCTCCAAAGTGA	**L77R**	CTTGATTGAGAGCTTATTGC
L54R	TGATTGTCGTTGCTGCTTTC	**L78F**	TAAGCTCCCTAACTGCCTTC
L55F	GCCTCTTTCATCTTCCCATC	**L78F2**	CACCAGGCAGTACTTTGACAC
L55R	GAAGACGGAATCACAGGGAG	**L78R**	AGAAACATCATCCAGGGTCG
L56F	GATTCTTCACAGGCTCAGG	**L79F**	TGGGAGTATCTGGCTTTAGG
L56R	ACCTGGCTAACTCATCCTTC	**L79R**	AGGAAGAGAGCAACAGAAGG
L56F1	CATGGACGGACATCAAGCCTG	**L80F**	AATACCTGCCTGGAGAATGG
L57F	CGTTGATAGGGAGTTTGCTG	**L80R**	GGGTCATCAGGGAGAATAC
L57R	GGTTATCCCCAATGTCTTCC	**L81F**	ACAACAAGTCCACAACCTGC
L58F	TAGCAAGAGCCCAACAATCC	**L81R**	AGTGAGGGAGAACAAAGAGC
L58R	CTGATGTTGACTTGGCTGC	**L82F**	TGGAAGATAGATGATAAGCC
L59F	GGACGACAATAATGAGAACG	**L82R**	TAATCCCAAATCAAAGCCAC
L59F2	GTCTGGCATTGTTGGCTTGTG	**L83F**	TTGGAGGGCAAGTACACAGA
L59R	CCAGAGGACCAAGAAATACC	**L83R**	CATCTCTTGTTCCTTGGTTG
L59R2	GTGCTCAGGACCCGGAAGAAT	**L84F**	AGTCTGGGAGTCTTTGGAGT
L59R3	ACAAGCCAACAATGCCAGACC	**L84R**	ACTTTCTTCCTCCTCATTCC
L60F	CACCTGGAAACTGGACTGAT	**L85F**	CCTGGAAGTGAAGTTGTTAC
L60R	ATCCAACAGCATTTCATCTC	**L85R**	CCTTGAGTGCCTGGTGTAG
L61F	GGTAACAGAGGAACAAAGGC	**L86F**	GTTCAGAGCGAAGGAGGAG
L61R	AATCATCCCTCTTTACAACG	**L86F2**	CTTTTGCCTCACCCACTCCAG
L62F	GTATCACATCAACTCCGTCC	**L86R**	GCTCCAGTCTCCTATCATCC
L62R	TCCTGCTCTGGTTCCTTAC	**L86R2**	GGCCAGGTGCGGTGGTTCATG
L63F	CCAGAAGCTCCAACATTGTC	**L87F**	GCAACAAGTGGGATTATGTG
L63R	CTAACAAGACAATGAGGCAG	**L87R**	ATCAGCACCAAGTAGAAATG
L64F	ACAGGAAAGTCACAGCAAAG	**L88F**	TATACCACTGGCACAACAAC
L64R	AAGACTCACAAATGCCTCAG	**L88F2**	AGACAGTTGGGAGCATCAGCA
L65F	AAGTAGAAGTCGCTCAAGGG	**L88R**	ATGATGATGTCCCAACTCCT
L65R	GACTTAGACTTGTGTTTGCG	**L88R2**	TCATACCGTCTGCTGCCTGCT
L66F	AAACTCCTGCTTATCTCTGC	**L89F**	GCTTCACCTACACCTTCCAT
L66R	CAGGCTCTCAAGTTTCAATG	**L89R**	TCGGAACAACACTCAATGCA
L67F	ATGACTTGTGGGGAGACTTC	**L90F**	GCTAGTTTTCGTCGCTTTGG
L67R	TCACAAAGAGGCACACAAGG	**L90R**	TGGCTTATCCTTGATAGTGG
L68F	CTAAAGAGCCTGATGGGGAC	**L91F**	TCTTGCCAGTAGTCTCAGTC
L68R	CCAGTAGCATCAGAAAGAGC	**L91R**	ATCATCTTCATCCAGTTCAC
L69F	AGACAGAAGCCAATCACATC	**L92F2**	CAGGAGCGCTGAACCCAGAAC
L69R	GTGCTTCTGGTTGTCTGATG	**L92R2**	GGTGTGGGTGGAAGTGGGAGA
L70F	GCTGAAACCTGGCTTATCTC	**L92F/L53F1**	GCTTACTGAGGACTTCTTTC
L70R	AGGAAAAGAGACACCCAATG	**L92R/L53R1**	TGTTCTCCCTTTCTGTCCTG
L71F	GAAGAACAGAACAGTGAAGC	**L93F2**	CACTGGAGAAGCTCAAC
L71R	CATTCATTTCACAGCAACAC	**L93R2**	CGAGGCTGCTGAATAATGTAG
L72F	TCCGTGGTGATGAAGAGTTG	**L93F/L45R**	TGAATGACGCACTGGAGAAG
L72R	TCCCAGGTTCTCACTCTTTC	**L93R/L45F1**	ATTTACCAAGCAAGTGCAGC
L73F	ACCAGAAGACCAGGATGATG	**L94F**	CTCCAGAAATAAAGCAGGTG
L73R	GGAGCAAGAGAACAAGAGAC	**L94R**	ACACAGGATGACAGCAGAAG

Table 21. Primer sequences of clone inserts for L25 gene

Primer	Sequence	Purpose	Enzyme	Cloning vector
L25SV2-EcoRI-F1	ATCGGAATTCCGGCTTCAGTGGACCAAATTT	GST/ MBP-L25SV2 fusion proteins	EcoRI	pGEX-4T-3 / pMal-cxh
L25SV2-XhoI-R1	CGATCTCGAGTGAAGCAACTCAGTAAGATG	GST/ MBP-L25SV2 fusion proteins	XhoI	pGEX-4T-3 / pMal-cxh
L25SV3-EcoRI-F	ATCGGAATTCCCCTTCCTTTGGAGACTGCTC	GST/ MBP-L25SV2 fusion proteins	EcoRI	pGEX-4T-3 / pMal-cxh
L25SV3-XhoI-R1	CGATCTCGAGATGTGGAACTGAGTCACAG	GST/ MBP-L25SV2 fusion proteins	XhoI	pGEX-4T-3 / pMal-cxh
L25-SV2-NotIF	GCGGCCGCATGAAGGATATCGACATAGG	Protein taged with 1D4	NotI	pCEP4
L25-SV2-NotIR	GCGGCCGCTTTCTGACCGCAGAATACAGC	Protein taged with 1D4	NotI	pCEP4
L25-SV3-NotIR	GCGGCCGCTCGTGAAGCAACTCAGTAAGA	Protein taged with 1D4	NotI	pCEP4

Table 22. Primers used for quantitative Real-Time PCR

Gene	Primers	Sequence	Fragment size	MgCl$_2$	T$_a$	Plate read
ABCC5_SV2	L25qPCRF1 L25qPCRR1	5'-CAAGAAGAGCTGAATGAAGT-3' 5'-ACAGCACCAAGCAAGTGGTC-3'	147 bp	2 mM	64°C	87°C
ABCC5_SV3	L25qPCRF2 L25qPCRR2	5'-GGCAAGAAGAGCTGAATGAAGT-3' 5'-CAGCCATCCTGAAAATTTGGT-3'	151 bp	2 mM	64°C	87°C
ABCC5_SV4	L25qPCRF2 L25qPCRR3	5'-GGCAAGAAGAGCTGAATGAAGT-3' 5'-CAGTCTCCAAAGGAAGGTGGT-3'	151 bp	2 mM	64°C	72°C, 87°C
ABCC5	ABCC5qPCRF4 ABCC5qPCRR7	5'-TGGCTTCAGTGGACCAGCCT-3' 5'-AGACCGCACGATTTCCGTCA-3'	124 bp	2 mM	66°C	86°C
B2M	B2M F B2M R	5'-TAAGCAGCATCATGGAGGTTT-3' 5'-AGCAAGCAAGCAGAATTTGGA-3'	70 bp	2 mM	66°C	72°C, 83°C
TBP	TBP F TBP R	5'-GGTTTGCTGCGGTAATCAT-3' 5'-CTGGACTGTTCTTCACTCTTGG-3'	106 bp	2 mM	66°C	72°C, 83°C
SDHA	SDHA F SDHA R	5'-TGGGAACAAGAGGGCATCTG-3' 5'-CCACCACTGCATCAAATTCATG-3'	86 bp	2 mM	64°C	78°C
HPRTI	HPRTI F HPRTIR	5'-TGACACTGGCAAAACAATGCA-3' 5'-GGTCCTTTTCACCAGCAAGCT-3'	100 bp	3 mM	65°C	72°C, 82°C

Table 23. Primers for SNP analysis of L33 gene

Primer	Sequence
L33-5'UTR-F4	TGTATGTAAGGCCCAGTGGC
L33-5'UTR-R4	GAGCCAGGTTTACATTTGAC
L33-5'UTR-F3	GAGAGCCCGTGATAGCAAAG
L33-5'UTR-R3	TATGGAAGCGCTAGCCAAGT
L33-5'UTR-F2	AAGAGCATCAAGGGGAAACA
L33-5'UTR-R2	CTGTAGGAGTTGAGTTGGAG
L33-5'UTR-F1	AGCGAAACACTTCCTCCAGA
L33-5'UTR-R1	GCATGGCTGATGAATCATTG
L33Ex1F	TGTGGAGTGAGAAAGTGGGT
L33Ex1R	GGAGGCAAGAAAACAGGTCA
L33-IVS1-F1	TGACCTGTTTTCTTGCCTCC
L33-IVS1-R1	TGCTTCGTCTCAACCTACTG
L33-IVS1-F2	CAGTAGGTTGAGACGAAGCA
L33-IVS1-R2	GAAAACCTACCATCAGAGCC
L33-IVS1-F3	TCTCCTGTAGCCTTATGACC
L33-IVS1-R3	TGATTGACAGATTCGGGGAC
L33-IVS1-F4	GGCTCTGATGGTAGGTTTTC
L33-IVS1-R4	AAATACTTGCCCTCTGTGTC
L33-IVS1-F5	TGTATTCCCTTTGCCCATCT
L33-IVS1-R5	GGTCATAAGGCTACAGGAGA
L33-IVS1-F6	GTCCCCGAATCTGTCAATCA
L33-IVS1-R6	GATACCTTTTCTCTCGCTCT
L33Ex2F	TGTGTCCTCTGGCATAGCAA
L33Ex2R	TCCTTTACCCTGCAACAACC
L33Ex3F	GGTTGTTGCAGGGTAAAGGA
L33Ex3R	CATAGGGGTTCTTGGAGAGA
L33Ex4F	TTGTGCATCGTTGTCCTGTAG
L33Ex4R	GAACTGACTTTGCTGGTGAC
L33-Rev	GGCAAGTCCGAAATCCATAG
L33-IVS4-F1	GTCACCAGCAAAGTCAGTTC
L33-IVS4-R1	CTGGATGACGCACTCTGAAC
L33-IVS4-F2	GCCCAGCACCAGGAATAATA
L33-IVS4-R2	GTCCATCCATTATTGGCATG
L33Ex5F	TTGTCAACACTCCTTGGACG
L33-3'UTR-F1	GGTGACAGAGGGCAACAAGT
L33-3'UTR-R1	TAGGGCTGGATGAGGGTGAC
L33-3'UTR-F2	CCAACATTCACTGCCATTCTT
L33-3'UTR-R2	CAGCCCACATTTCTCAACCT
L33-3'UTR-F3	CACTTCCTCCTCCCAAATCA
L33-3'UTR-R3	AAACACACCAGTAGGAAAGC
L33-3'UTR-F4	GTCACCCTCATCCAGCCCTA
L33-3'UTR-R4	CAGGTATTCAGACCACATGC

Table 24. Primers for SNP analysis of L35 gene

Primer	Sequence
L35-5'UTR-F2	AGCTTTCTGAAGGTTGGGAC
L35-5'UTR-R2	GGGTCTTGCTTGAGGAGTAA
L35-5'UTR-F1	AGGAAGAAATGCCAAGACCA
L35-5'UTR-R1	TCTAGTAGATGGTGGCCAGA
L35-5'UTR-F3	ACCCTGAAACTCCAGATCCT
L35-5'UTR-R3	TTGATAAGGTAAGGGAGGAT
L35-5'UTR-F4	TTACTCCTCAAGCAAGACCC
L35-5'UTR-R4	AGGATCTGGAGTTTCAGGGT
L35Ex1F	TTGCATTGGTGTGTGTGTTG
L35Ex1R	GAGCAAGAGAAGGCAAAGGT
L35Ex1F1	GCTGCTGAAACACCTATTGC
L35Ex1R1	TGCATGTTAGTACACAGTAG
L35-IVS1-F1	GTCCTCCTGCTTCCCACATA
L35-IVS1-R1	ATCCCTCTCCCACCTCAATG
L35-IVS1-F2	TACGTGACTAAGCCCAGACC
L35-IVS1-R2	GTTGAGCTTCTGAGCATTAC
L35-IVS1-F3	CATGTCACTGGGGATTCTTG
L35-IVS1-R3	GAGTTCCATTGCTTCATCCT
L35Ex2F	GACAAATCAAGCCTCCTCTA
L35Ex2R	GGCAGGGCATTTAGTCAAGA
L35Ex3F	GCCAAGAGGTGACAATAGTA
L35Ex3R	AAGTCAGAGTTTAGGCATCC
L35Ex4F1	CACCTTAGTCATTTTCTTGC
L35Ex4R1	GTAGAGAGGACAATGGCAAG
L35Ex4F2	CTTGCCATTGTCCTCTCTAC
L35Ex4R2	TTTCATGGGTCATGAGTGGG
L35Ex4F3	CCCACTCATGACCCATGAAA
L35Ex4R3	GCTTAATTCTCAACTGGCCA
L35Ex4F4	TGGCCAGTTGAGAATTAAGC
L35Ex4F5	GACAAGCTACCTCAACAGAC
L35Ex4F6	TCAATGTAGCCTATGCAAGC
L35Ex4R5	TCCATTTGTGCTCATCTCAG
L35-3'UTR-F1	GGAGAATGGGGTAATGACTG
L35-3'UTR-R1	GCGATTTGGGGAGGTAGACT
L35-3'UTR-F2	AGTCTACCTCCCCAAATCGC
L35-3'UTR-R2	AAATCCAGGGTAAAGTCAGG
L35-3'UTR-F3	AGCTCATTCTAAATCCCCAG
L35-3'UTR-R3	ATTCCTTCACCATTCACAGC

Table 25. Primers for SNP analysis of L38 gene

Primer	Sequence
L38-5'UTR-F3	GGCTGTATACCTGAGTCACC
L38-5'UTR-R3	ATCATCTTCTTCCCCTTCTG
L38-5'UTR-F2	AAAAGATTGCTCGGTGACAG
L38-5'UTR-R2	TTTGCTTGCTTTGTGATGAC
L38-5'UTR-F1	ATTCCTTGCCTGATGAGTTC
L38-5'UTR-R1	CCTGAAATGTAACCACTGAT
L38-5'UTR-R4	GCCTTATCTTCCACTACCCA
L38Ex1F	TTGGTGACTGAAGAAAGGTG
L38Ex1R	ATCAAAATCCCTCCCTGGTT
L38-IVS1-F1	ATGCCTATCTTTGTCTTTCC
L38-IVS1-R1	GTGAACACAGGGAATTCTGG
L38-IVS1-F2	CGTGGTGTAAGTGTTTTCCC
L38-IVS1-R2	GACTTCAACTTTTCCCTAAC
L38Ex2F	AGAGGGCGGTTTTTAGGTTC
L38Ex2R	AAAGCCCTAGATAAGAAGAG
L38Ex2F1	GCAGGTAGGCCAGTGATTTC
L38Ex2R1	AGCCTCCATCCTATTTATCC
L38-IVS3-F1	CCCATCTACTCCTTTGCTTA
L38-IVS3-R1	CCCACTATCACTTTCTCCTG
L38-IVS3-F2	TTCTCACAACGACTATGCTG
L38-IVS3-R2	CACATGAACAAACACACCAT
L38Ex3F	TTGCCATGTAGGCTGGTAGG
L38Ex3R	TAAGCAAAGGAGTAGATGGG
L38Ex4F	ATGGTGTGTTTGTTCATGTG
L38Ex4R	CTCTCCTAAATACTCGGCAG
L38Ex5F	GTATGTCTCCAGTAAGCACC
L38Ex5R	ATGGAAGCTCAGGGATCTGT
L38-3'UTR-F1	TCTTCCCAAAACATCCCTGC
L38-3'UTR-R1	GCAATCCACCAAACTTTAAG
L38-3'UTR-F2	TTCTTTTGGTTAGGGTCTGG
L38-3'UTR-R2	GCATTTGTCTCTTCCAGGCT
L38-3'UTR-F3	GGGACTGAAAGATACAATGC

9 List of publications

Thesis-related publications

Stojic J, Schulz HL, Weber BH: Nonsense-mediated mRNA decay regulating expression of ABC transporters preferentially expressed in human retina. (Manuscript in preparation)

Schulz H, Rahman FA, Fadl El Moula FM, **Stojic J,** Gehrig A, Weber BH: Identifying differentially expressed genes in the mammalian retina and the retinal pigment epithelium by suppression subtractive hybridization. Cytogenet Genome Res. 2004;106(1):74-81.

Stohr H**, Stojic J**, Weber BH: Cellular localization of the MPP4 protein in the mammalian retina. Invest Ophthalmol Vis Sci.2003 Dec; 44(12):5067-74.

List of poster presentations

Stojic J, Weikert M, Gehrig A, Stöhr H, Weber BH. 2001. Evaluation of a cDNA library enriched for retinal transcripts by suppression subtractive hybridization. German Human Genome Meeting, The Genome and Beyond. Braunschweig, Germany.

Stojic J, Gehrig A, Schulz HL, Wagner M, Weber BH. 2002. Identifying novel retina-specific genes by characterising suppression subtracted cDNA library highly enriched for retinal genes. 13. German Human Genome Meeting. Leipzig, Germany.

Schulz HL, Stöhr H, Fröhlich S, Berger C, **Stojic J**, Weber BH. 2002. Identification of gene preferentially expressed in the human retina using an expressed sequence tag (EST) approach. 13. German Human Genome Meeting. Leipzig, Germany.

Schulz HL, Stöhr H, **Stojic J**, Fröhlich S, Berger C, Weber BH. 2002. Towards a comprehensive characterization of the human retinal transcriptome. 1st Symposium of the National Genome Research Net, Berlin, Germany.

Stojic J, Schulz HL, Gehrig A, Wagner M, Weber BH. 2003. Cloning and characterization of three putative non-coding RNA genes exclusively expressed in retina. 14. German Human Genome Meeting. Marburg, Germany.

Schulz HL, Stöhr H, **Stojic J**, Weber BH. 2003. Towards a comprehensive characterization of the human retinal transcriptome. 53rd Annual Meeting of the American Society of Human Genetics, Los Angeles, USA.

10 CURRICULUM VITAE

Personal information

Date of birth	August 21st, 1974
Place of birth	Belgrade, Yugoslavia
Nationality	Serbian
Marital status	Single

Education

Elementary school
1981-1989

Belgrade, Yugoslavia

Gymnasium
1989-1993

Belgrade, Yugoslavia

University studies
1993-2000

Diploma of Science, Molecular Biology, Faculty of Biology, University of Belgrade, Yugoslavia.
Diploma project: "Isolation, identification and antibiotic activity of novel Actinomycetes soil strains".

Ph.D. Thesis
January 2001
-June 2004

Research Doctorate in Human Molecular Genetics, University of Würzburg, Germany.
Project: "Cloning and functional characterization of novel genes expressed preferentially in the human retina".

Professional experience

2001-present

Training in the Laboratory of Molecular Ophthalmology and Oncology Institute of Human Genetics, University of Würzburg, Germany.

1999-2000

Training in the Laboratory for Molecular Genetics of Actinomycetes, Institute of Molecular Genetics and Genetic Engineering, Belgrade, Yugoslavia.

1992-2000

Active member, and one of the founders of Youth of JAZAS (Yugoslav Association against AIDS), working in HIV/AIDS prevention in Serbia.