Studies on Transcriptional Regulation in the Human Retina: Mapping of Transcriptional Start Sites of Retinal Expressed Genes
and
Functional Characterization of the *CNGA3* promoter

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

der Fakultät für Biologie
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Ronald Erick Carpio Farro
aus Callao

September-2009

Tag der mündlichen Prüfung:

Dekan der Fakultät für Biologie:	Prof. Dr. Hanspeter Mallot
Dekan der Medizinischen Fakultät:	Prof. Dr. Ingo B. Autenrieth
1. Berichterstatter:	Prof. Dr. Bernd Wissinger
2. Berichterstatter:	Prof. Dr. Olaf Rieß
Prüfungskommission:	Prof. Dr. Bernd Wissinger
	Prof. Dr. Eberhart Zrenner
	Prof. Dr. Frank Schaeffel
	Prof. Dr. Thomas Gasser
	Prof. Dr. Nikolaus Blin

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale."

Marie Curie

to my family

GENERAL ACKNOWLEDGMENTS

It takes a long time to write a PhD thesis, though not as long as it takes to do the experiments for it, surprisingly. I would like to acknowledge and extend my heartfelt gratitude to the following persons who have made the completion of this thesis possible:

Prof. Dr. Bernd Wissinger, our Chair, for his understanding and assistance during the achievement of this thesis at the Molecular Genetics Laboratory, Institute for Ophthalmic Research at the Centre for Ophthalmology.

Our Director, Prof. Dr. Eberhart Zrenner and Prof. Dr. Theo van Veen, for their vital encouragement and support.

Dr. Simone Schimpf-Linzenbold and Dr. Susanne Kohl, thank you for the supervision, constant reminders and much needed motivation.

Dr. Katja Koeppen and Dipl. Biol. Tanja Grau, for the help, inspiration and friendship they extended during this last four years.

Dr. Valeria Roni, for introducing me to the field of eye research.

Thanks to Dipl. Biol. Peggy Reuter for the discussions during this thesis period.

Special thanks to the great technical assistance provided by Monika Papke, Norman Rieger and Britta Bauman during the experimental course of this thesis.

Thanks to all Molecular Genetics Laboratory and Department for Experimental Ophthalmology members and Staff, of particular Dr. Francois Paquet-Durand, Dr. Blanca Arango Gonzales, Dr. Stefan Kustermann and Javier Sancho Peluz, thank you for the support, friendship and good sense of humor.

Special thanks to the Histolab Staff to Silvia Bolz and Katja Dengler for their support with the experiments.

Finally I thank Bodil Gesslein for her love, support, stubbornness, patience and courage during the last two years. Also thanks for the unstinting reminder that there are more important things in life than a PhD thesis (namely enjoying walks, cooking and travelling).

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project, most especially to my family and friends.

And to God, who made all things possible...

TABLE OF CONTENTS

7

INTRODUCTION

The Retina

The retina is a light sensitive neural tissue at the innermost surface of the eye (Figure 1). The eye projects an image onto the retina where the light initiates a cascade of chemical and electrical events that ultimately trigger nerve impulses. Through the optic nerve these impulses are sent to various visual centers of the brain. In vertebrate embryonic development, the retina and optic nerve originate as outgrowths of the developing brain, so the retina is considered part of the central nervous system (CNS).
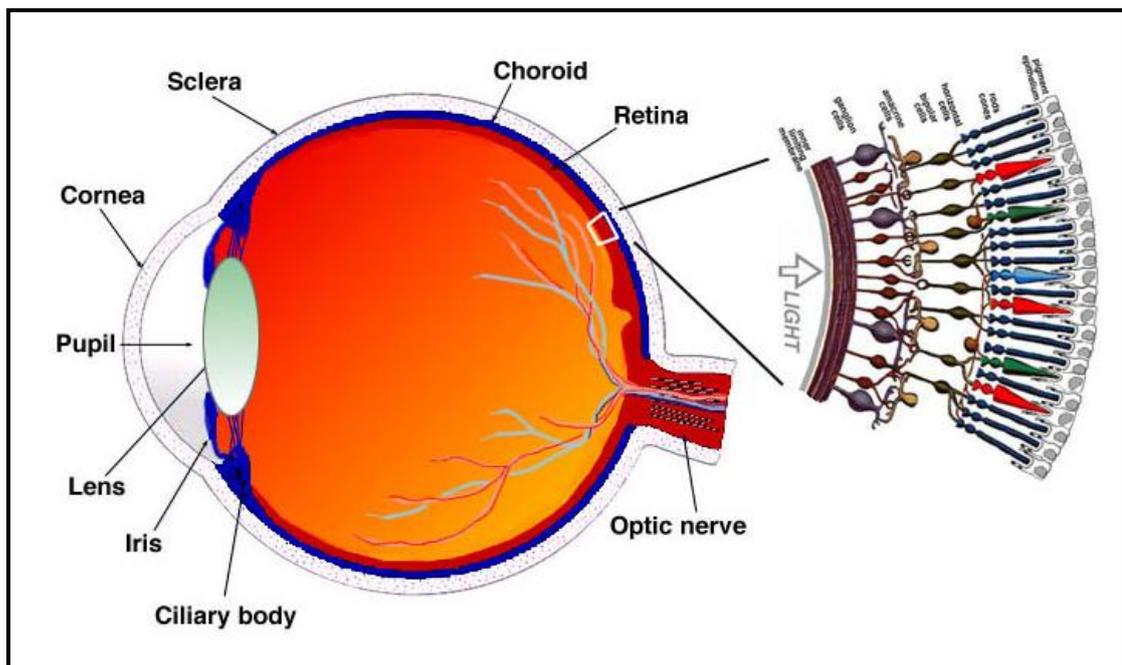
Figure 1*: Drawing of a section through the human eye with a schematic enlargement of the retina (reproduced from www.webvision.med.utah.edu).*

Structure of the Retina

In vertebrates the retina is structured in different layers (Figure 2). The **Retinal Pigment Epithelium Layer (RPE)** is the outermost layer which has vital functions visual cycle, maintenance of photoreceptors and vitamin A cycle. Moreover, the RPE cells are responsible for the phagocytosis of the outer segments of rods and cones as they are daily renewed. Another and even more important function of the RPE is the storage and metabolization of trans-retinal, as part of the visual cycle. The subsequent cell layer, the **Photoreceptor Layer,** in the human eye consists of two different types of photoreceptors: the rods (~120 million) with its visual pigment rhodopsin (absorbance maximum of ~500 nm), responsible for the scotopic vision, and the cones (~6 million). The human eye has three different types of cones: blue or short-wavelength, green or middle-wavelength, and red or long-wavelength sensitive cones with a maximal absorbance of 420 nm, 534 nm and 563 nm, respectively. This difference results from the expression of three types of photopigments (opsins) that

constitutes the functional specialization of these three cone photoreceptor types. Photoreceptors are not equally distributed in the retina. The macula contains the highest concentration of cones. The *fovea centralis*, located at the center of the macula, contains only cones and no rods. In contrast, the highest concentration of rods is in the peripheral retina, decreasing in density up to the *macula*. The nuclei and cell bodies of the rod and cone photoreceptors form the **Outer Nuclear Layer (ONL)**. The following layer, the **Outer Plexiform Layer (OPL)**, is the site of numerous synapses between rods and cones and the dendrites of various integrator neurons (horizontal, bipolar and amacrine cells). The next layer, the **Inner Nuclear Layer (INL),** consists of cell bodies and nuclei of integrator neurons, bipolar, amacrine and horizontal cells. Following the INL, the **Inner Plexiform Layer (IPL)** is the second region of synapses, including the axons of the bipolar and amacrine as well as the dendrites of the ganglion cells. Here the bipolar cells forms synapses with dendritic processes of ganglion cells. The somata of the ganglion cells form the **Ganglion Cell Layer (GCL)**, which has far fewer nuclei than the inner or outer nuclear layers. The **Nerve Fiber Layer (NFL)** is formed by the axonal fibers of the ganglion cells. These axons run radially to the papilla, where they are bundled together and leave the eye as the optic nerve.
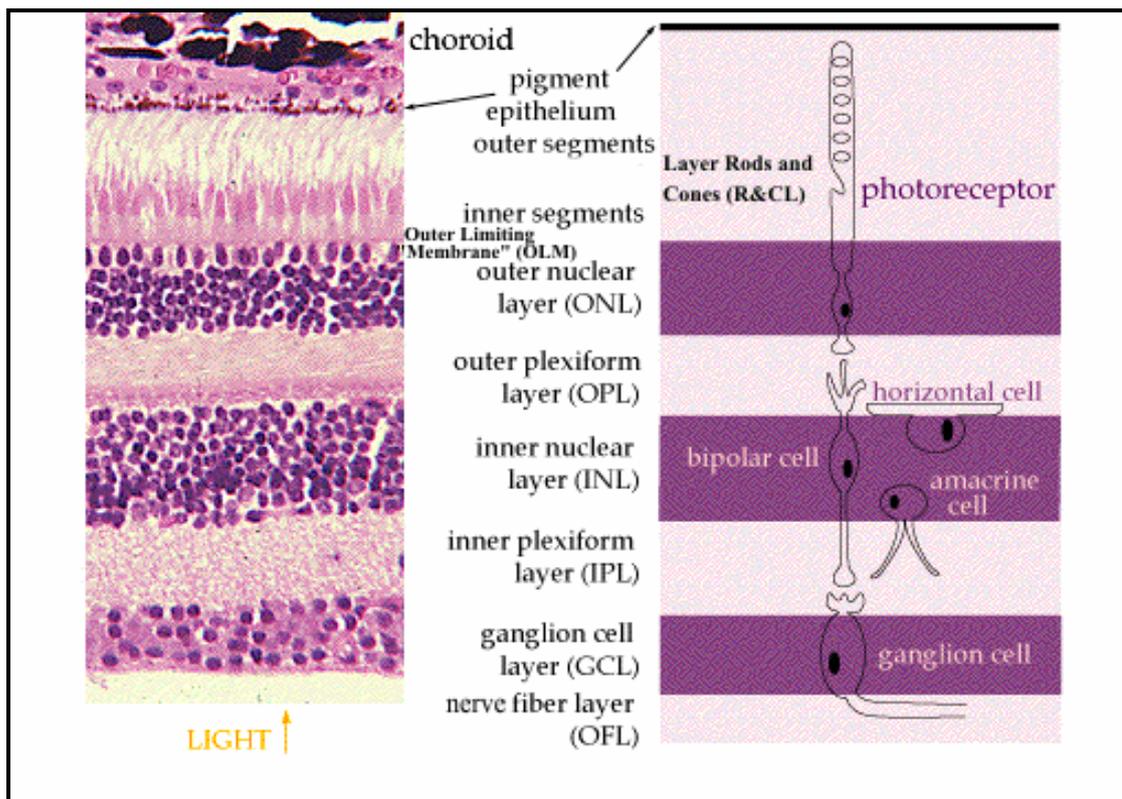
Figure 2: Diagram of the organization of the vertebrate retina in comparison to a transversal section of the tissue (adapted from http://thalamus.wustl.edu/ course/eyeret.html).

Phototransduction

Cones respond to bright light and mediate spatial high-resolution together with color vision. Rods respond to dim light and mediate lower-resolution, monochromatic and night vision.

cGMP is the second messenger in photoreceptor function and is crucial for the phototransduction (Figure 3) (Kawamura and Tachibanaki, 2008). In the dark, the intracellular concentration of cGMP is high. cGMP is the principal ligand of the cGMP-gated channels. Elevated levels of cGMP leads to the opening of channels and allow the influx of ions into the outer segment. In the dark the membrane potential of the photoreceptor cells is slightly depolarized and exerts a constant release of glutamate from the synaptic terminal. Upon illumination, a photon is absorbed by the photopigment in the membrane disks in the outer segment of photoreceptor. The activated photopigment activates a G-protein, called transducin, which in turn activates a phosphodiesterase that hydrolyzes cGMP. This lowering of the concentration of cGMP throughout the outer segment, reduces the number of cGMP molecules that are available for binding to the cyclic nucleotide gated channels, leading to channel closure (Shuji Tachibanaki, 2007). Channel closure induces membrane hyperpolarization and subsequently glutamate release stops. This change in neurotransmitter release is registered by bipolar cells, which transmit this signal onto ganglion cells. The ganglion cells then relay this information to the brain.
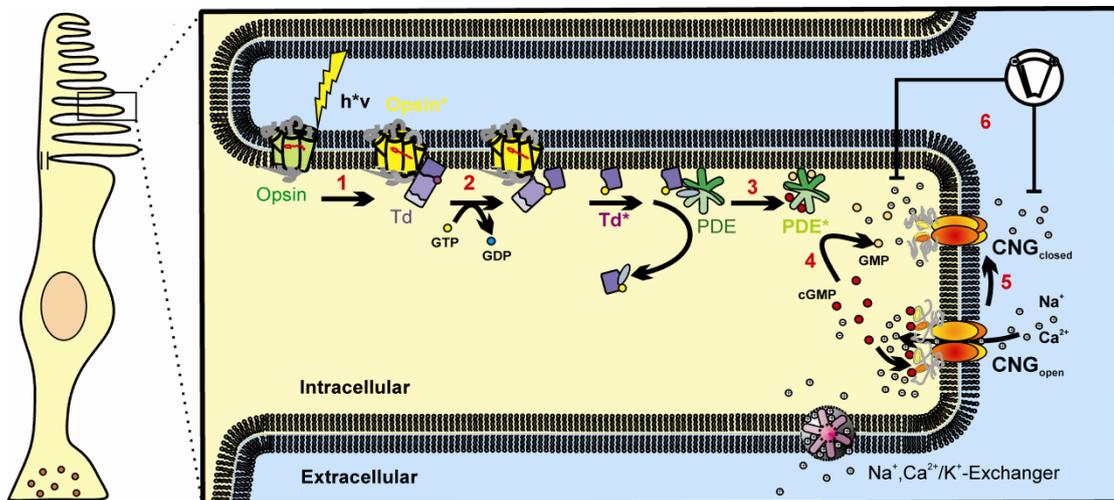
Figure 3: *Scheme of phototransduction in human cone photoreceptors (adapted from Wissinger, 1998).*

Achromatopsia: loss of color vision

Achromatopsia (rod monochromacy or total color blindness) is an autosomal recessive stationary cone dystrophy with a prevalence of less than 1 in 30,000 (Sharpe, 1990). The disorder is congenital and characterized by poor visual acuity, severe sensitivity to

light, nystagmus, and a complete lack of color perception. To date, mutations in three genes that play essential roles in the cone phototransduction pathway have been identified. The first two genes discovered were *CNGA3* and *CNGB3*, which encode the α- and β-subunits of the cGMP-gated channel present in cone photoreceptors, respectively (Kohl et al., 2000; Kohl et al., 1998). The other gene identified is *GNAT2,* which encodes the cone photoreceptor-specific α-subunit of transducin, a G-protein of the phototransduction cascade (Kohl et al., 2002).

Due to the fact that people with achromatopsia have no functional cones, they are left with the monochromatic extremely light sensitive, low acuity rod system. Figure 4 gives an idea of an achromats perception.
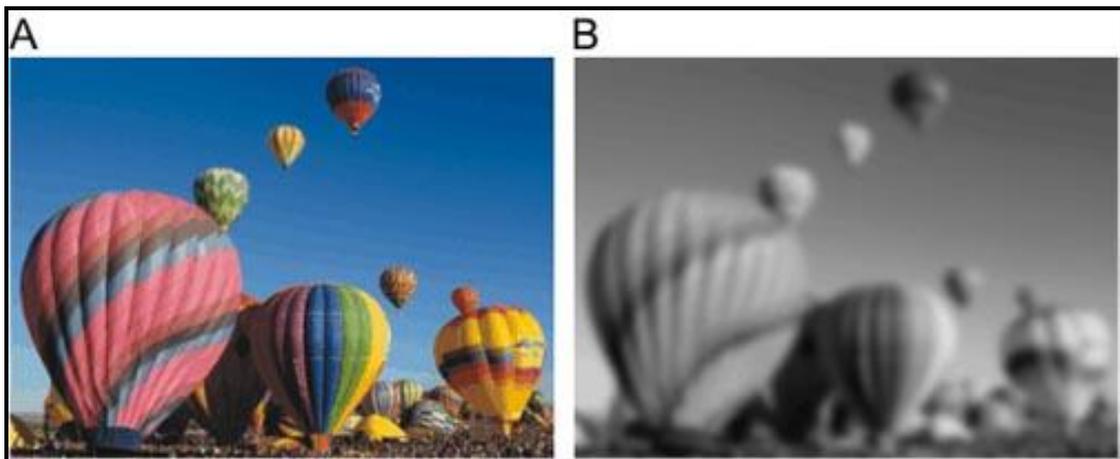
Figure 4: *Comparison of representations of normal vision field (A) to an achromat (B) (Adapted from http://www.fz-juelich.de/isb/isb-1/Achromatopsia/)*

Gene Regulation

Gene regulation includes the processes that cells use to turn the information of genes into gene products. Although a functional gene product may be RNA or a protein, the majority of known regulatory mechanisms are on protein coding genes. Any step of the gene's expression may be modulated, from DNA-RNA transcription to the post-translational modification of a protein. Gene regulation is essential for viruses, prokaryotes and eukaryotes as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed. Furthermore, gene regulation drives the processes of cellular differentiation and morphogenesis, leading to the creation of different cell types in multicellular organisms where the different types of cells may possess different gene expression profiles though they all possess the same genome sequence.

Transcription Start Sites

Transcription start site (TSS) is the location within a gene where transcription of RNA begins; in consequence TSSs are located in the core promoter region of a gene.

The transcription start site of a metazoan gene remains poorly understood, mostly because there is no clear signal present in all genes. mRNA sequence data stored in public databases, lack information about their 5' ends because of the difficulty in obtaining full-length cDNAs. Thus, even after the completion of human genome sequencing, it is not easy to locate TSSs.

TSSs are usually determined by aligning complete mRNA or 5'end EST sequences (for instance obtained by 5' RACE) against the genome. Note that protein sequences and other ESTs are not sufficient for this task, since they typically start downstream of the TSS. For some species including human, large scale sequencing projects of complete mRNAs have been undertaken, but many low copy genes still evade being sequenced.

When multiple TSSs for a gene are found, it enables to examine the dynamic nature of the transcriptional initiation events. Loose specification of TSS may reflect slippery interaction between the promoter and the transcription machinery, whereas tight specification may reflect the rigid interaction. In any case, it is intriguing to find correlations between distribution patterns of TSSs and their upstream sequences.

Also, since TSS marks the 5' end limit of the cDNA, the amino acid sequence of the exact N-terminus of the encoding protein could be deduced, which is essential to examine the presence of protein sorting signals.

The RNA Polymerase II Core Promoter

A promoter is the region on the DNA that enables the initiation of the transcription of a particular gene. Promoters are usually located proximal to the corresponding genes, which they regulate, and upstream of the transcription start site. In Eukaryotes three different RNA polymerases named I, II and III transcribe genes encoding ribosomal RNA, messenger RNA and transfer RNAs, respectively. The RNA polymerase II core promoter is therefore the sequence of DNA that directs the initiation of mRNA transcription for protein coding genes (Juven-Gershon et al., 2006; Muller et al., 2007; Sandelin et al., 2007; Smale and Kadonaga, 2003; Thomas and Chiang, 2006). This core promoter could be as short as a single motif (as the INR, defined as YYANWYY in humans (Purnell et al., 1994), or as complex as a set of binding sites for transcription factors for a particular promoter. Since the combinatorial binding of transcription factors can activate different promoters, its properties and binding factors are not strictly absolute as a single model, therefore little is known about specific gene mechanisms of transcription.

Promoters can be divided in two main types: focused and dispersed promoters (Figure 5). Most research has been done on focused or single-peak core promoters where there is either a single transcription start site or a distinct cluster of start sites within a short region.
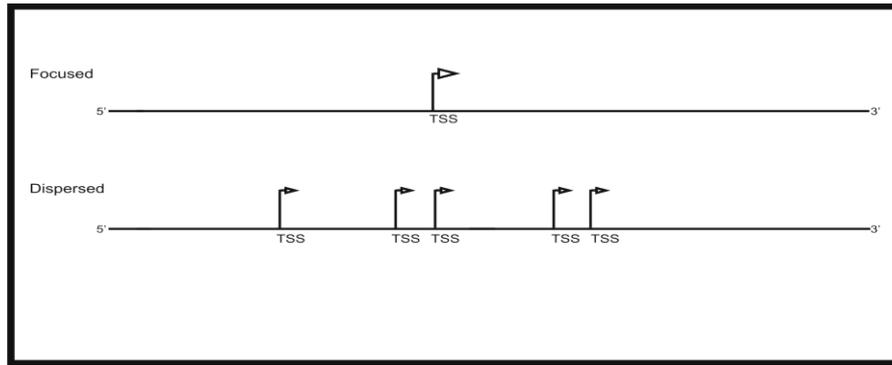
Figure 5: *Focused versus dispersed core promoters (adapted from Juven-Gershon et al., 2008).*

In vertebrates, only about one-third or less of core promoters are focused promoters; instead, the majority of genes appear to contain dispersed core promoters in which multiple transcription start sites are distributed over a broad region that range from 50 to 100 nucleotides (Juven-Gershon et al., 2008) (Figure 5 and 6). The structure of the distribution of these multiple TSS can be further subdivided in different classes: i) a broad (BR) shape with multiple, low strength TSS, ii) a bimodal/multimodal (MU) shape class which implies multiple well-defined TSSs within one cluster and iii) a combination of a well-defined TSS surrounded by weaker TSSs resulting in a broad shape with a dominant peak (PB) (Carninci et al., 2006; Figure 6).
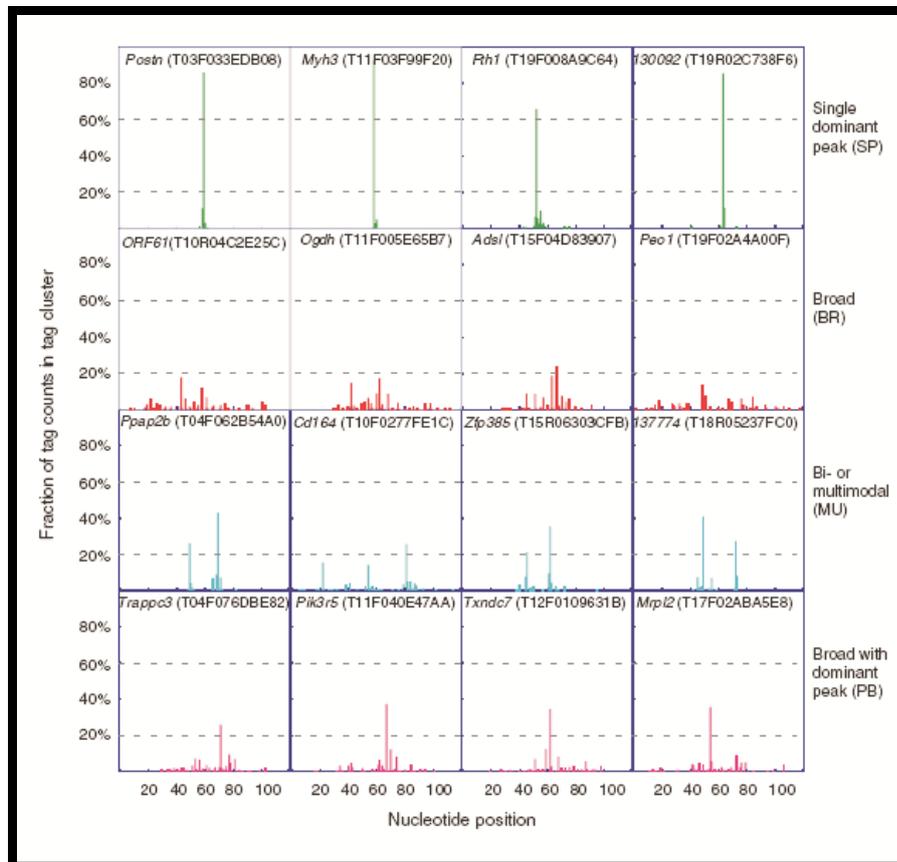
Figure 6: *Types of dispersed core promoters taken from Carninci et al., 2006.*

Core Promoter Elements

Certain core promoter elements are typically found in focused core promoters (Figure 7). Nevertheless, those elements are not universal and some are present in only a subset of them. Moreover, some core promoters appear to lack all of the known core promoter elements.
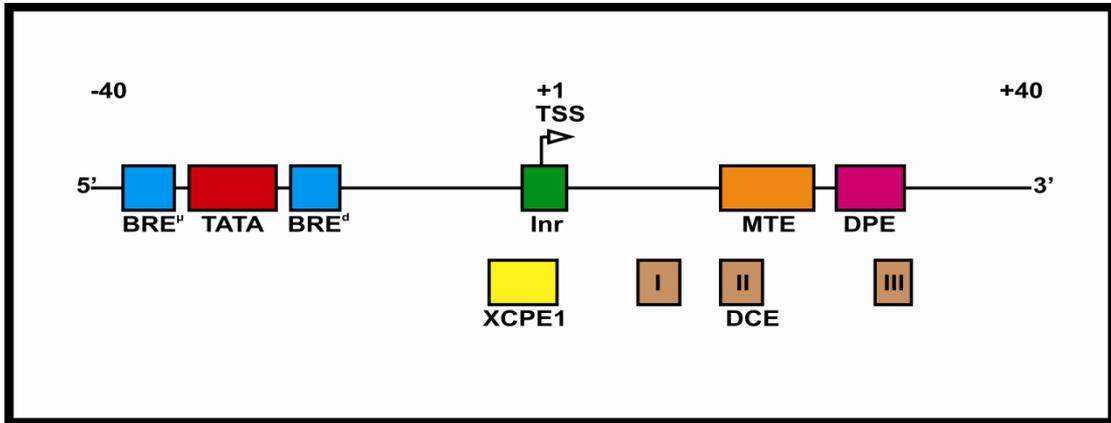
Figure 7: Core promoter motifs (adapted from Smale and Kadonaga, 2003).

The **initiator (Inr)** is the most commonly occurring sequence motif in focused core promoters (around 66.5%) (FitzGerald et al., 2006; Gershenzon et al., 2006; Ohler et al., 2002) and is usually around the transcription start site (Smale and Baltimore, 1989; Smale and Kadonaga, 2003). In humans its consensus sequence is YYANWYY (Figure 8) and it is a recognition site for the binding of the Transcription Factor IID RNA polymerase preinitiation complex (TFIID) (Purnell et al., 1994). It is worth to mention that three other proteins recognize Inr elements: RNA polymerase II, TFII-I, and YIN-YANG-1 (YY1) (Carcamo et al., 1991; Roy et al., 1993; Roy et al., 1991; Seto et al., 1991). YY1 is a zinc finger protein that has been identified as a transcription factor that binds a distal element in the adeno-associated virus (AAV) P5 promoter (Shi et al., 1991). YY1 can repress transcription through this element in the absence of the adenovirus E1A protein, but it activates it in the presence of E1A (Shi et al., 1991). This duality on its function serves as a complex mechanism of gene regulation.

The **TATA box**, which was the first eukaryotic core promoter element identified, is the most ancient and most widely used core promoter motif throughout nature. The TATA box has the consensus sequence TATAWAAR (Wong and Bateman, 1994; Figure 8), where the upstream 5' or proximal T nucleotide is most commonly at position −31 or −30 relative to the TSS (Carninci, 2006; Ponjavic et al., 2006). The TATA box is recognized and bound by the TATA-binding protein (TBP), which is a subunit of the TFIID complex in eukaryotes. The TATA box is present in a subset of focused core promoters and is rather uncommon in vertebrates (Carninci, 2006; Cooper et al., 2006; Kim et al., 2005).

The **BRE (TFIIB recognition element)** was originally identified as a TFIIB-binding sequence that is located immediately upstream of a subset of TATA boxes (Lagrange et al., 1998). It can be located upstream or downstream of the TATA box and is denoted as the BRE^u or BRE^d sequences (Deng and Roberts, 2005; Deng and Roberts, 2007; Figure 8). The BRE^u consensus is SSRCGCC (Lagrange et al., 1998) and the BRE^d consensus is RTDKKKK (Deng and Roberts, 2005). Depending on the promoter context, the BRE^u and BRE^d can act in either a positive or negative manner (Deng and Roberts, 2005; Deng and Roberts, 2007; Lagrange et al., 1998).
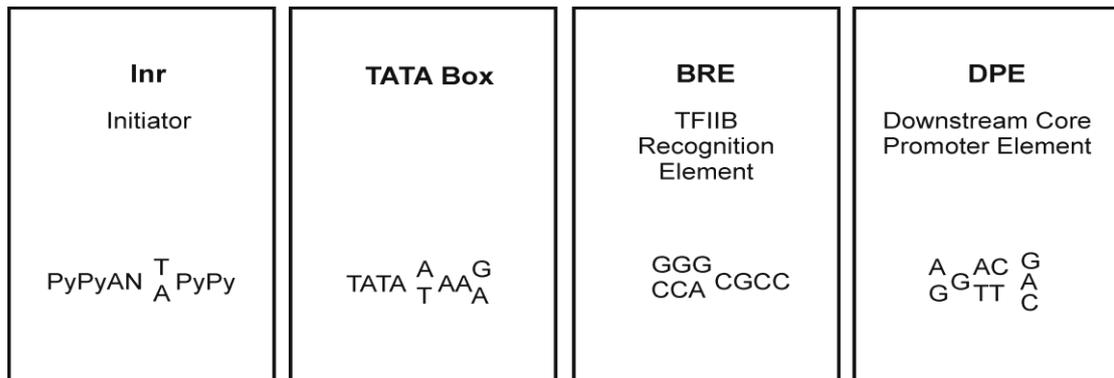
Figure 8: Core Promoter motifs (adapted from Smale and Kadonaga, 2003). Degenerate nucleotides are indicated according to the IUPAC nucleotide code.

The **DPE (downstream core promoter element)** is important for basal transcription activity of a subset of TATA-less promoters (Burke and Kadonaga, 1996; Burke et al., 1998; Butler and Kadonaga, 2002; Kadonaga, 2002). The DPE consensus is RGWYVT in *Drosophila* (Kutach and Kadonaga, 2000) and is located +28 to +33 bp upstream of the Inr where it works cooperatively with it.

The **MTE (motif ten element)** has the consensus sequence CSARCSSAAC and is typically located from +18 to +27 bp upstream of the Inr. The MTE is present in humans (Lim et al., 2004); however it's not an abundant motif and does not emerge as an overrepresented sequence in computational analyses of mammalian promoter databases (FitzGerald et al., 2006; Frith et al., 2008).

The **DCE (downstream core element)** frequently occurs together with the TATA box, and appears to be distinct from the DPE. The DCE consists of three sub elements: S_I, CTTC at +6 to +11; S_{II}, CTGT at +16 to +21; and S_{III}, AGC from +30 to +34 (Juven-Gershon et al., 2008).

The **XCPE1 (X core promoter element 1)** is present in about 1% of human core promoters, most of which are TATA-less. The XCPE1 motif is located from −8 to +2 relative to transcription start site and has the consensus sequence DSGYGGRASM (Tokusumi et al., 2007). XCPE1 may work along with sequence-specific activators in CpG islands to direct transcription initiation since it exhibits little activity by itself and it acts in conjunction with sequence-specific activators, such as NRF1, NF-1, and Sp1.

CpG Islands

CpG islands contain a high frequency of CG dinucleotides. The CpG islands are genomic regions associated with approximately half of the promoters for protein-coding genes (Antequera and Bird, 1993; Matsui et al., 1980). It is estimated that in mammals, the human genome contains approximately 29 000 CpG islands of 0.5–2 kbp length (Bird, 2002). Despite the prevalence of promoters associated with CpG islands, the elements that are responsible for their function remain limited defined. One common feature of them is the presence of multiple binding sites for the transcription factor Sp1 (Blake et al., 1990; Brandeis et al., 1994; Macleod et al., 1994), but they usually lack consensus TATA boxes, DPE or Inr elements (Blake et al., 1990). Promoters associated with CpG islands usually contain multiple transcription start sites that span a region of 100 bp or more often located 40–80 bp downstream of the Sp1 sites. Core promoter recognition within CpG islands relies on the same general activation process as for other classes of promoters discussed before. The main difference is that binding of basal factors is more strongly dependent on activator proteins bound to distal promoter elements.

In silico analysis of promoter regions

Computational studies have revealed the occurrence of various combinations of core promoter motifs (FitzGerald et al., 2006; Gershenzon and Ioshikhes, 2005; Gershenzon et al., 2006; Jin et al., 2006; Ohler, 2006), which are essential to understand the full range of mechanisms that are used in the process of basal transcription and to characterize all of the sequence elements that contribute to core promoter activity. Positional weight matrices are the main algorithm of transcription factor binding prediction. Experimentally determined binding sites for a transcription give the weights in the weight matrix. There are two major databases for eukaryotic transcription factor weight matrices: TRANSFAC (Matys et al., 2003) and JASPAR (Vlieghe et al., 2006), which have matrices based on experimental evidence. The false-positive problem, however, is inherent to any predictive algorithm like the ones linked to databases. False-positive predictions arise by chance of single-motifs matching in a DNA sequence without being functional in that region in the *in vivo* context. To solve this, additional type of information is needed. Therefore, the most powerful workaround for the many false-positive annotations is the analysis of evolutionary conservation of promoter sequences across several different species ("phylogenetic footprints"; Tagle et al., 1988). Where selective pressure exerts on regulatory *cis*-acting elements, it is likely to conserve them more than the surrounding non-coding sequences. Other approaches to eliminate "false-positives" are usually provided through better motif descriptions and combination of motifs into *cis*-regulatory modules.

Transcriptional Regulation in Photoreceptors

At the molecular level, development and maintenance of photoreceptors require precisely regulated gene expression. Photoreceptor cells preferentially express a set of genes that are essential for their function, so called photoreceptor-specific genes. Increased or decreased expression levels of a photoreceptor gene can lead to photoreceptor degeneration (Humphries et al., 1997; Olsson et al., 1992). Furthermore, precisely regulated photoreceptor gene expression is also a driving force for photoreceptor development/differentiation (Turner and Cepko, 1987; Wetts and Fraser, 1988). This regulation is mediated by a network of photoreceptor transcription factors centered on Crx, an Otx-like homeodomain transcription factor. Crx and Otx2 are trans-activator for many photoreceptor genes which activates transcription by directly binding to the promoter and/or enhancer regions of the target genes in photoreceptor cells (Peng and Chen, 2005). Data indicate that Otx2 and Crx have redundant but indispensable roles in photoreceptor development and maintenance that might be contributed by protein-protein or protein-promoter interactions between the two factors (Martinez-Morales et al., 2003; Nishida et al., 2003; Peng and Chen, 2005). The cell type specificity of this network is governed by factors that are preferentially expressed by rods, cones or both, including the rod-determining factors neural retina leucine zipper protein (Nrl) and the orphan nuclear receptor Nr2e3; and cone-determining factors, mostly nuclear receptor family members.

Most vertebrates have at least two different cone subtypes, producing opsins with different spectral sensitivities: S-opsin (blue) and M-opsin (green) cones. The regulatory factors that are responsible for triggering cone progenitors to begin differentiating are still not well understood. Retinoic acid or related compounds are tempting candidates for cone differentiation (Hyatt and Dowling, 1997). Retinoic acid related compounds work by diffusing through cellular membranes and binding to nuclear receptors, which associate with specific regulatory elements of gene promoter and enhancer regions. Cone gene regulation include transcription factors like: thyroid hormone receptor β2 (Trβ2), retinoid related orphan receptor Rorβ and retinoid X receptor Rxrγ. Rxrs are unique among nuclear receptors because they heterodimerize with members of several other nuclear receptor families, including Nr2e3 (Chen et al., 1999) and thyroid hormone receptor Trβ2. Rxrγ receptor is of particular interest because it is localized to developing cone photoreceptors in a number of species. Retinoid-related orphan receptors (Ror) are another family of genetically-related receptors with homology to the receptors for retinoids whose actual ligands have not been identified. One isoform in particular, Rorβ2, has also been shown to synergize with Crx in vitro to activate the *S-opsin* gene (Srinivas et al., 2006).

The M-cone differentiation is mostly established by thyroid hormone (TH), a splice variant of the thyroid hormone receptor B (*Thrb*) gene (Szanto et al., 2004), which acts as an extrinsic signal responsible (Harpavat and Cepko, 2003). The thyroid hormone receptor Trβ2 is implicated in photoreceptor development in chick and mouse (Ng et al., 2001; Roberts et al., 2005; Roberts et al., 2006; Yanagi et al., 2002). Trβ2 binds

directly to the *S-opsin* promoter and has been reported to interact with the basal transcription machinery as well as co-activators and co-repressors to exert complex regulatory effects on target genes (Eckey et al., 2003). TH nuclear receptors are reported to exert their effects as heterodimers in combination with retinoid (usually Rxr) receptors (Mangelsdorf and Evans, 1995); its likely role in S-cones is inhibitory rather than inductive (Roberts et al., 2005).

Cone photoreceptor promoters

The present research done on gene regulatory sequences of cone photoreceptors is limited on the extent of it applications. Specific cone-directed therapy is of high priority in the treatment of human hereditary retinal diseases. Cone specificity had been targeted in some animal models and now the next step is to reach cone specificity in humans.

Regulatory sequences and nuclear factors governing tissue-restricted expression of the arrestin gene in mouse and frog have been investigated. Its proximal promoter binds three putative retina-specific nuclear factors through overlapping sequences centered between positions -25 and -15 (Kikuchi et al., 1993). These CRX-binding elements together with TATA elements are crucial for driving both basal transcriptional activity and tissue specificity to cone photoreceptors and pinealocytes (Pickrell et al., 2004; Zhu et al., 2002).

Cis-acting DNA elements involved in the expression of the human M and S opsin genes have been correlated with transcription factor binding sites. The expression of these genes in a cell-specific fashion seems to be controlled mainly by positive-acting elements in the region between -130 and the TATA box. The higher activity of the M opsin gene promoter could have evolved in someway to compensate for its longer distance from an activating locus control region compared to the one of the red pigment gene promoter (approximately 34 versus 3.5 kb) (Shaaban and Deeb, 1998). Also, using transgenic *Xenopus*, (Babu et al., 2006) identified a highly conserved extended region (-725 to -173) in the L opsin promoter necessary for cone-specific expression. They further identified a short element (5'-CCAATTAAGAGAT-3') highly conserved amongst tetrapods, including humans, necessary to restrict expression to cones in the retina. Three other versions of the human red cone opsin promoter have been evaluated for their specificity and robustness. Experiments targeting GFP expression to subclasses of cones in the canine retina using recombinant adeno-associated viral vectors (Komaromy et al., 2008; Li et al., 2008) suggests that L/M cones, the predominant class of cone photoreceptors in the retinas of dogs and most mammalian species, can be successfully targeted using the human cone L opsin promoter. Interestingly, the human S opsin promoter preferentially targets reporter gene expression to S-cone photoreceptors on experiments conducted in rats (Glushakova et al., 2006).

Also, it have been shown that cone expression of the GNAT2 gene is controlled in mouse by a strong silencer region, a weak upstream cell-specific promoter, and a strong downstream element (Morris et al., 1997; Ying et al., 1998).

PURPOSE OF THE WORK

Identification of the Transcription Start Sites (TSSs) in human retinal expressed genes

The first part of this work was dedicated to the identification of TSSs of human retinal expressed genes and to improve and complete gene annotation concerning the 5'-UTR. This work was based on the following goals:

- *In silico* analysis of putative TSSs for human retinal expressed genes.
- Comparison of whole available pull of mRNA transcripts in order to understand and predict potential alternative TSSs for human retinal expressed genes.
- Mapping of the TSSs of selected human retinal expressed genes.
- Analysis of the selectivity of genes for TSSs in human retinal tissue.

Identification and Characterization of the human *CNGA3* gene promoter

Based on the knowledge generated with the first part, the second major task of this work was to identify and characterize the basic promoter of the human *CNGA3* gene. This was done by means of:

- *In silico* analysis of the putative promoter region of *CNGA3*.
- Interspecies comparison of the TSS and the putative promoter region of *CNGA3*.
- Generation of reporter constructs containing the putative promoter regions of *CNGA3*.
- Generation of individual deletions for putative transcription factor binding on reporter that contain promoter activity.
- Cotransfection of specific plasmids for transcription factors together with the reporter constructs containing promoter activity.
- Localization of different *CNGA3* transcripts within the human retina.

Experimental Outline

Chapter I: Identification of the Transcription Start Sites (TSSs) in human retinal expressed genes

- *In silico* analysis of ESTs and known algorithm predictions of genes expressed in human retina to select the best candidate genes for Cap-finer RACE experiments.
- Experimental validation via Cap-finder RACE of genes expressed in human retina.
- Report of TSSs of 54 retina expressed genes: new sequences for 41 genes.
- Results analysis and classification into five categories (i) TSS located in new first exons, (ii) splicing variation of the second exon, (iii) extension of the annotated

first exon, (iv) shortening of the annotated first exon, (v) confirmation of previously annotated TSS.

Chapter II: Identification and Characterization of the human *CNGA3* gene promoter

- 5'-RACE experiments performed with human, mouse and zebrafish retinal mRNA. In human, data indicates that transcription of the *CNGA3* gene can start within two alternative exons: TSS.1 located in exon U1 and TSS.2 located in exon 1.
- Luciferase reporter gene constructs containing the putative human *CNGA3* promoter transfected into ARPE-19, HEK 293, Y79, WERI-Rb1 and 661W measured by luminometry. Only the upstream region of exon 1 exhibits promoter activity in reporter gene assays.
- Interspecies sequence comparison and predictions of putative transcription factor binding sites were done to identify potential regulatory candidates.
- Identification of potential regulatory regions with sequential deletions experiments suggests regulatory targets for transcription factors. Verification of those results with directed deletions experiments via mutagenesis of the predicted transcription factor binding sites and overexpression of some cotransfected transcription factors. Data suggests promoter negatively regulated by the transcription factor YY1.
- Subtractive *in situ* hybridization experiments on human retinal tissue with specific probes for exons U1 and 1 suggest a differential cell specific expression of the *CNGA3* transcript that start at TSS.2

CHAPTER I: MAPPING OF TRANSCRIPTION START SITES OF HUMAN RETINA EXPRESSED GENES

Introduction

The spatial and temporal regulation of gene transcription is primarily determined by its flanking promoter (*cis*-regulatory DNA elements) through interaction with *trans*-acting regulatory proteins (transcription factors) (Hahn, 2004; Smale and Kadonaga, 2003). The start of transcription is accomplished by the formation of a pre-initiation complex on the DNA, yet our knowledge of transcriptional initiation sequences in the human genome is still limited despite the availability of the complete genome sequence (Lander et al., 2001; Venter et al., 2001). Therefore one of the main remaining challenges is to locate these gene sequences, defined as the transcription start site (TSS), in order to explore core promoter and *cis*-regulatory elements that direct the start of every transcript. Genomic structure and full length cDNA sequences aligned on the genome provide opportunities to locate TSSs. Conventional methods for determining exact TSSs, such as 5′ RACE or primer extension are laborious and are not selective for the complete transcript. Consequently, many mRNA sequences stored in public databases lack information about their genuine 5' ends, mainly due to the difficulties in obtaining full-length cDNA. Several bioinformatic and experimental approaches have been developed to explore full-length cDNAs and the human transcriptome (Trinklein et al., 2003). Computational predictions may represent a powerful tool to localize first exons and TSSs on an averaged genome-wide scale (Davuluri et al., 2001; Sonnenburg et al., 2006), however they may fail at the level of individual genes or in genes with complex regulatory patterns (e.g. multiple or tissue-specific TSS).

Recently a number of experimental approaches to compile TSSs on a genome-wide scale have been established including the Database of human Transcriptional Start Sites (DBTSS) (Suzuki et al., 2004), whole genome tilling array analysis (Cheng et al., 2005), and the exploration of mouse and human CAGE tag libraries (Carninci, 2006). To enable future progress we need to complete and revise these catalogues with an accurate annotation of the 5′and 3′end, and include splice isoforms of the transcripts. In addition to genome wide approaches, there is a need for more specific studies, which cover tissue specific genes, expressed in a restricted manner. Identification of potential transcription signals that are tissue specific relies on the correct determination of transcriptional start sites.

In this work we describe an experimental approach to identify the TSSs of a selected group of genes, which are predominantly expressed in retina. We focused our attention on the human retina, due to its unique and specialized function. This complex tissue, composed of multiple, highly differentiated and specific cell types (e.g. rod and cone photoreceptors, amacrine cells, Mueller glial cells), expresses a large number of

specific genes. Mutations in many of these genes result in blinding disorders. The subset of genes expressed in human retina has been partially elucidated (Blackshaw et al., 2004; Schulz et al., 2004), with a number of studies defining genes that are either highly expressed in retina or which pose a crucial target of transcription factors in this tissue (Blackshaw et al., 2001; Chowers et al., 2003; Lord-Grignon et al., 2004; Yu et al., 2003). We selected a pool of retina expressed genes and employed Cap selective RACE to ensure amplification and subsequent cloning of genuine TSSs. We describe herein the results of this analysis, reporting the correct TSSs within this group of retinal transcripts.

Methods

In silico analysis

We used available public database information (RefSeq, NCBI and Ensemble predictions covered by at least one EST, Unigene EST database) to perform *in silico* assembly and analysis of 5´transcript termini. The procedure involves downloading the sequences from public databases, clustering them to obtain a consensus and designing a gene model with the most complete 5' transcript termini. Sequences were downloaded and assembled using the SeqMan program (DNAstar). In our analysis we consider retina ESTs as well as non-retinal ESTs to obtain the most complete information about the different alternative start sites already mapped. Additionally, the information about putative TSS was assessed using Promoser (Halees et al., 2003), Eukaryotic Promoter Database Current Release 87 (http://www.epd.isb-sib.ch), and Database of Transcriptional Start Sites DBTSS (http://dbtss.hgc.jp). First exon boundaries were determined by aligning the predicted sequence to the genome using BLAT (Kent, 2002). The gene model was then deduced considering the clustering of all the collected sequences giving the priority to the most accurate database in the 5' termini (EPD and DBTSS), but considering also gene prediction (GENSCAN, implemented in the UCSC Genome Browser) and gene annotation that are confirm by at least one EST (AceView, http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/). To evaluate first exons conservation we used BioEdit pairwise and multiple alignments (www.mbio.ncsu.edu/BioEdit/bioedit.html).

Primer design

Gene specific reverse primers were selected within exons other than the first exon to obtain spliced products. Two primers were chosen for each gene, in order to perform a nested PCR, which allows to enhance specificity, and to obtain a sufficient amplification product for rare transcripts. Primers were designed using the Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) software and checked for uniqueness by querying against the human genome.

RACE protocol

We applied the RNA-ligase–mediated RACE (RLM–RACE) system from Ambion (Austin, Texas). RNA sample from adult human retina was treated with DNAse I. After DNAse treatment and inactivation, 10 µg of total RNA was dephosphorylated for 60 min at 37°C with 10U Calf Intestinal Phosphatase to remove the 5'-phosphate from all RNA species except those that have a cap structure (present on all Pol II transcripts). RNA was then phenol/chloroform extracted, precipitated and re-suspended in water. Dephosphorylated RNA was then digested for 60 min at 37°C in a 10-µL reaction with 10U tobacco acid pyrophosphatase. Subsequently the RNA was incubated for 60 min at 37°C with 1 U T4 RNA ligase and 0.3 µg of an RNA adapter (5' RACE Adapter 5'-GCU-GAU-GGC-GAU-GAA-UGA-ACA-CUG-CGU-UUG-CUG-GCU-UUG-AUG-AAA-3'). After ligation, 180 ng of RNA was incubated for 2 min at 75°C in the presence of 5 µM random decamers in RT buffer. Single-stranded cDNA was generated by the addition of 100 U M-MLV RT and incubation at 42°C for 60 min.

Amplification of 5' RACE cDNA was carried out using nested gene-specific primers and adapter specific primers and with 1 µl of the first-strand cDNA reaction. PCR reactions were done in 50 µl volume including 5 µl 10X PCR Buffer supplied in the RLM kit, 4 µl dNTP Mix 10 mM, 2 µl 5' RACE gene-specific outer primer (10 µM), 2 µl 5' RACE Outer or Inner Primer (10 µM), and 1 U thermostable DNA polymerase. Cycling conditions were: 5 min initial denaturation at 94°C PCR followed by 35 cycles of 95°C for 30 s, 60-55 °C (empirically determined) for 30 s (annealing), 72°C for 30 s (extension) and a final extension at 72°C for 7 min. Amplified products were analyzed on a 3% agarose gel and visualized by ethidium bromide staining.

Cloning and sequencing of RACE products

5' RACE products were cloned into pCR-2.1 vector (Invitrogen, Carlsbad, California). 2 µL of PCR reaction were incubated with 0.1 ng of vector at 16°C over-night. Aliquots of the ligation were used to transform library efficiency chemically competent *E. coli* DH5α (Invitrogen, Carlsbad, California) 8 to 48 clones of each transformation were subjected to colony PCR and the inserts sequenced with standard M13 forward and reverse primers applying Big Dye Terminator v3.0 chemistry (Applied Biosystems, Carlsbad, California). Sequencing products were separated and analyzed on an ABI 3100 DNA sequencer.

Sequence analysis

Gene sequences and cDNA sequences, obtained using the RACE, were aligned to the human genome using BLAT. cDNA sequence was considered informative only if the following criteria were met: (I) the spliced sequence mapped to the same region of the genome as the gene sequence; (II) the product could be mapped uniquely to the genome with >95% identity and (III) presence of the gene specific primer sequence and the 5' RACE Adapter primer sequence.

Primer extension

A fluorescein phosphoramidites (FAM)-labelled reverse primer was added to 10 µg of DNAse I-treated total RNA to a final concentration of 5 nM. Samples were heated at 70 °C for 5 min followed by 20 min incubation at 58 °C and then allowed to cool for 15 min to room temperature. First strand cDNA synthesis was performed using Avian Myeloblastoma Virus (AMV) reverse transcriptase and 5x AMV-RT buffer (Promega) according to the manufacturer's instructions in a final volume of 60 µl. The primer extension reaction was done in two repeated reaction cycles. After an initial reverse transcription step (60 min at 42°C in a total volume of 50 µl), enzyme was replenished and the samples underwent a second extension reaction (60 min at 42°C, adjusting the buffer to a total volume of 60 µl). Finally 3µl of 5M NaOH was added to each cDNA sample, the reaction incubated at 37 °C for 15 min and then neutralized with 16µl of 2M HEPES free acid. Extension products were purified using the columns AutoSeq G-50 (Amersham Pharmacia Biotech). Samples were separated on 50-cm capillary columns in the POP4 acryl amide polymer (Applied Biosystems, Carlsbad, California) on an ABI PRISM 3100 Genetic Analyzer (Applied-Biosystems, Carlsbad, California) Sequencer with GENESCAN 500-ROX added as size standard (Applied-Biosystems, Carlsbad, California). The sequences of the gene specific reverse primers were: *RDH12* 5' tgagcagcagcctgactctgagcagagcccaga 3', *CNGA3* 5'atcttctcggtttgtcacatttagc 3', with 5' ends modified with the fluorescent molecule 6-FAM.

Results

Genes Selection and in silico assembly

76 annotated genes were selected for analysis. The selection was done based on the following criteria: (i) specific or high levels of expression in retina, (ii) a role in retina specific physiological processes or retinal development, and (iii) involvement in retinal disease. A compilation of all tested genes including gene symbol, definition, chromosomal location and tissue/cell type of expression is shown on Table 1.

cDNA and transcript sequences available in public databases (RefSeq, NCBI and Ensemble gene predictions covered by at least one EST, Unigene ESTs database) were downloaded and new assemblies generated using SeqMan.

We found that 5' transcript termini represented in public datasets can be readily identified by clusters of cDNA ends in the assemblies. Additionally, the information about putative TSSs was assessed in The Eukaryotic Promoter Database (Schmid et al., 2006) and Database of Transcriptional Start Sites (DBTSS) (Suzuki et al., 2004). These data were compiled to create a preliminary gene model which was used to design primers for the subsequent Cap-finder RACE experiments.

Table 1: List of the genes selected for Cap-finder RACE procedure. *The list shows the Gene Symbol, Gene Name, Chromosomal location, tissue/cell type of expression, associated disease of the genes selected for the study.*

ABCA4	ATP binding cassette, subfamily A (ABC1), member 4	1p22.1p21	Photoreceptors	Stargardt disease, FF, CRD, MD age related, RP
AIPL1	aryl hydrocarbon receptor interacting protein like 1	17p13.1	Photoreceptors, pineal gland	recessive Leber congenital amaurosis; dominant CRD
ANKRD33	ankyrin repeat domain 33	12q13.13	Retina	n/a
AOC2	amine oxidase, copper containing 2 (retina specific)	17q21	Retinal ganglion cell layer	Choroideremia
C14ORF2	chromosome 14 open reading frame 2	14q32.33	Ubiquitous	n/a
C1orf32	chromosome 1 open reading frame 32	1q24.1	Eye	n/a
C1QL2	complement component 1, q subcomponent like 2	2q14.2	Eye, Brain	n/a
C7orf9	chromosome 7 open reading frame 9	7p21p15	Retina	n/a
CHM	choroideremia (Rab escort protein 1)	Xq21.2	Ubiquitous	Choroideremia
CLUL1	clusterin like 1 (retinal)	18p11.32	Retina (cones)	n/a
CNGA3	cyclic nucleotide gated channel alpha 3	2q11.2	Retina (cones), pancreas, brain, testis	Achromatopsia 2
CNGB3	cyclic nucleotide gated channel beta 3	8q21q22	Retina (cones), skeletal muscle, pineal gland, brain	Achromatopsia 3
CRB1	crumbs homolog 1 (Drosophila)	1q31q32.1	Photoreceptors (inner segment)	Leber congenital amaurosis
CRB2	crumbs homolog 2 (Drosophila)	9q33.2	Retina	n/a
CRX	cone rod homeobox	19q13.3	Photoreceptors	CRD 2
DHRS3	dehydrogenase/reductase (SDR family) member 3	1p36.1	Ubiquitous, high expression in retina	n/a
ELOVL4	elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)like 4	6q14	Photoreceptors (inner segment)	MD, autosomal dominant, chromosome 6 linked
ELOVL5	elongation of long chain fatty acids (FEN1/Elo2, SUR4/Elo3like, yeast) like 5	6p21.1p12.1	Ubiquitous	n/a
EYA3	eyes absent homolog 3 (Drosophila)	1p36	Ubiquitous	n/a
FSCN2	fascin homolog 2, actin binding protein, retinal (*Strongylocentrotus purpuratus*)	17q25	Eye, tyroid, placenta, ovary	RP 30
GNAT1	guanine nucleotide binding protein (G protein), alpha transducin activity polypeptide	3p21	Retina (rods)	Night blindness, congenital stationary

GNG13	guanine nucleotide binding protein (G protein), gamma 13	16p13.3	Retina Bipolar Cells	n/a
GNGT1	guanine nucleotide binding protein (G protein), gamma transducin activity polypeptide	7q21.3	Retina (rods)	n/a
GRK7	G protein coupled receptor kinase 7	3q21q23	Retina, brain	n/a
GUCA1A	guanylate cyclase activator 1A (retina)	6p21.1	Retina, brain and testis	Cone dystrophy 3
GUCA1B	guanylate cyclase activator 1B (retina)	6p21.1	Retina (photoreceptors)	n/a
GUCY2D	guanylate cyclase 2D, membrane (retinaspecific)	17p13.1	Retina	CRD 6
GUCY2F	guanylate cyclase 2F, retinal	Xq22	Retina	n/a
HPCA	Hipocalcin	2p25.1	Ubiquitous	n/a
Hs.221513	Clone c222389 mRNA sequence	6q21	Retina, testis	n/a
IMPDH1	IMP (inosine monophosphate) dehydrogenase 1	7q31.3q32	Ubiquitous adult, restricted to liver and eye in fetus	RP 10
IMPG1	Inter photoreceptor matrix proteoglycan1	6q14.2-q15	Interface between photoreceptors and the RPE	n/a
IMPG2	Inter photoreceptor matrix proteoglycan2	3q12.2q12.3	Interface between photoreceptors and the RPE	n/a
KIFC3	kinesin family member C3	16q13q21	Ubiquitous	n/a
LHX3	LIM homeobox 3	9q34.3	Retina, brain	Pituitary hormone deficiency with rigid cervical spine
LRRC21	leucine rich repeat containing 21	10q23	Photoreceptors (outer segment), brain	n/a
MPP4	membrane protein, palmitoylated 4 (MAGUK p55 subfamily member 4)	2q33.2	Retina	n/a
NR2E3	nuclear receptor subfamily 2, group E, member 3	15q22.32	Photoreceptors	Enhanced Scone syndrome
NRL	neural retina leucine zipper	14q11.1q11.2	Retina (rods)	RD, autosomal recessive, clumped pigment type
OPN1SW	opsin 1 (cone pigments), shortwave sensitive (color blindness, tritan)	7q31.3q32	Cones	Color-blindness, tritan
OPN4	opsin 4 (melanopsin)	10q22	Retina (ganglion cells)	n/a
OPTC	Opticin	1q32.1	Eye	n/a
OTX2	orthodenticle homolog 2 (Drosophila)	14q21q22	Retina, brain	n/a
PCDH21	protocadherin 21	10q22.1q22.3	Sensory neurons	n/a

PDC	phosducin	1q25.2	Rods	n/a
PDE6B	phosphodiesterase 6B, cGMP specific, rod, beta	4p16.3	Retina (photoreceptor cells)	Night blindness, congenital stationary, type3
PDE6G	phosphodiesterase 6G, cGMP specific, rod, gamma	17q25	Ubiquitous, in retina restricted to rods	n/a
PDE6H	phosphodiesterase 6H	12p13	Retina (cones)	n/a
PRPF31	pre mRNA processing factor 31 homolog (yeast)	19q13.42	Ubiquitous	RP11
RAX	retina and anterior neural fold homeobox	18q21.32	Eye	n/a
RBP3	retinol binding protein 3, interstitial	10q11.2	Interface between photoreceptors and the RPE	n/a
RCV1	recoverin	17p13.1	Retina and brain	n/a
RdCVF	thioredoxinlike 6	19p13.11	Rods	n/a
RDH11	retinol dehydrogenase 11 (all-trans and 9cis)	14q24.1	Ubiquitous and Muller cells	n/a
RDH12	retinol dehydrogenase 12 (all-trans and 9cis)	14q24.1	Retina (photoreceptor cells)	Leber congenital amaurosis, type III
RDH5	retinol dehydrogenase 5 (11cis and 9cis)	12q13q14	Retinal Pigment Epithelium	Fundus albipunctatus
RDH8	retinol dehydrogenase 8 (all-trans)	19p13.2p13.3	Retina (photoreceptor cells)	n/a
RDS	retinal degeneration, slow	6p21.2p12.3	Retina (photoreceptor cells)	Butterfly dystrophy, retinal
RGR	retinal G protein coupled receptor	10q23	RPE and Muller cells, muscle and brain	RP, autosomal dominant
RHO	rhodopsin (opsin 2, rod pigment) (retinitis pigmentosa 4, autosomal dominant)	3q21	Retina (rods)	Night blindness, congenital stationery
RLBP1	retinaldehyde binding protein 1	15q26	RPE	Bothnia retinal dystrophy
ROM1	retinal outer segment membrane protein 1	11q13	Retina (photoreceptor cells)	RP, digenic
RP1	retinitis pigmentosa 1 (autosomal dominant)	8q11q13	Retina, testis, muscle, trachea	RP
RP1L1	retinitis pigmentosa 1like 1	8p23	Retina (photoreceptor cells)	n/a
RPE65	retinal pigment epithelium specific protein 65kDa	1p31	Retina (cones) and RPE	Leber congenital amaurosis 2
RPGR	retinitis pigmentosa GTPase regulator	Xp11.4	Retina (photoreceptor cells)	Cone dystrophy 1
RPGRIP1	retinitis pigmentosa GTPase regulator interacting protein 1	14q11	Ubiquitous, in retina restricted to photoreceptors	CRD 9

RPL14	ribosomal protein L14	3p22	Ubiquitous	n/a
RS1	retinoschisis (X linked, juvenile) 1	Xp22.2p22.1	Secreted photoreceptor protein	Retinoschisis
SAG	S antigen; retina and pineal gland (arrestin)	2q37.1	Retina, pineal gland	Oguchi disease1
SLC1A7	solute carrier family 1 (glutamate transporter), member 7	1p32.3	Ubiquitous high expression in retina	n/a
SLC24A1	solute carrier family 24 (sodium/potassium/calcium exchanger), member 1	15q22	Retina (rods)	n/a
SLC24A2	solute carrier family 24 (sodium/potassium/calcium exchanger), member 2	9p22p13	Retina (cones)	n/a
TULP1	tubby like protein 1	6p21.3	Retina, testis, muscle, larynx	RP 14
VAX2	ventral anterior homeobox 2	2p13	Ventral portion of the retina (development)	n/a
WDR17	WD repeat domain 17	4q34	Retina, testis	n/a

RP = retinitis pigmentosa, n/a = not available, MD = macular degeneration, CRD = cone-rod dystrophy, RD = retinal degeneration,

FF: fundus flavimaculatus.

Experimental examination of TSSs

Cap-finder RACE cDNA fragments were cloned and a variable number of clones were sequenced for each gene, depending on the number and the sizes of the colony PCR products detected on the gel. We obtained products for 54 genes out of the 76 genes analyzed. A summary of the results obtained with Cap-finder RACE is shown in Table 2. Genes for whom the promoter and TSS were already known (*RHO* and *OPN1SW*) served as internal positive controls. For each gene we detected at least one splice variant that agreed with one or more RefSeq annotated exons. Our strategy relies on the location of gene specific primers within internal exons. We obtained those cDNA products that covered at least one exon-exon junction and thus ruled out the possibility of amplification of genomic contamination. This strategy has enabled us to identify alternatively spliced 5' ends that arise from tissue specific gene expression and regulation. Table 2 lists the results of the Cap-finder RACE experiments for 54 retinal expressed genes and the corresponding RefSeq entry (database release 18). These results can be grouped into five categories with reference to RefSeq; (i) new TSSs within novel exons (8 genes), (ii) alternative splice form of the second exon (2 genes: *IMPG1, SAG*, (iii) extension of the annotated first exon (27 genes), (iv) length shortening of the annotated first exon (4 genes), (v) confirmation of previously annotated TSSs (13 genes). Table 2 provides the exact nucleotide positions of 5' termini of the Cap-finder RACE cDNA clones referring to the UCSC Human Genome Browser (March 2006 assembly). In defining the interval where TSSs are located we report the start, and when present the internal frequent start and end nucleotide position of each TSS. Sequences from this study have been submitted to GenBank under the accession numbers: DQ067456-DQ067464, DQ426859-DQ426897, and DQ980599-DQ980621.

Table 2: Results of the RACE experiment to determine TSS of retina transcripts. The Table provides Gene Symbols of the processed genes, RACE results referring to the RefSeq entry (database release 18), nucleotide position of the TSS (start, internal frequent start and end) referring to the UCSC Human Genome Browser (March 2006 assembly), Chromosomal location (Chr.), shape of the TSS (Shape) according to the classification of (Carninci et al., 2006): single peak [SP], broad [BR], bimodal/multimodal [MU], broad with dominant peak shape [PB], comparison with Cage TSS database (+ for correspondence, - no correspondence). The genes are listed according to the type of results that was obtained according to the description in the text: (i) new TSSs within novel exons (8 genes), (ii) alternative splice form of the second exon (2 genes:, (iii) extension of the annotated first exon (27 genes), (iv) length shortening of the annotated first exon (4 genes), (v) confirmation of previously annotated TSSs (13 genes).

Gene Symbol	RACE results		TSS location			Chr.	Shape	CAGE
			Start	Int. peak	End			
C1orf32	Two new first exons	Isoform a	165,208,791			1q24.1	SP	+
		Isoform b	165,269,346				-	-
CNGA3	Extension exon 0 of 180 bp	Isoform a	98,329,050		98,329,083	2q11.2	BR	+
		Isoform b	98,329,288				SP	-
		Isoform c	98,329,371				SP	+
DHRS3	New first exon	Isoform a	12,578,722			1p36.1	SP	-
	Shorter exon 1 of 508 bp	Isoform b	12,599,903		12,599,935		SP	-
ELOVL5	New first Exon		53,320,946			6p21.1-p12.1	-	-
KIFC3	Two new first exons	Isoform a	56,437,970			16q13-q21	SP	-
		Isoform b	56,370,953				SP	-
RCV1	Extension exon 1 of 203 bp	Isoform a	9,749,613	9,749,402	9,748,934	17p13.1	PB	-
	New first Exon	Isoform b	9,745,910				-	-
	Lack of exon 1. exon extended	Isoform c	9,745,244		9,745,221		MU	-
RDH12	Two new exons		67,254,268	67,254,276	67,257,284	14q24.1	MU	-
SLC24A2	Two new first exons	Isoform c	19,778,103			9p22-p13	SP	-
		Isoform a	19,778,808				-	-
		Isoform b	19,778,609				-	+
		Isoform d	19,776,949		19,777,002		PB	+
IMPG1	Lack of exon 2		76,839,115		76,839,078	6q14.2-q15	SP	-
	Lack of exon 1. shorte exon2		76,808,496				SP	-
	Complete in the first 3 Exon		76,839,060				-	-
SAG	Lack of exon 2		233,998,462		233,998,52?	2q37.1	BR	-
AIPL1		61bp	6,279,208		6,279,306	17p13.1	-	-
AOC2		10bp	38,250,126			17q21	SP	+
CLUL1		31 bp	606,669	606,693	606,698	18p11.32	PB	-
CNGB3	Extension exon 1 of	43 bp	88,404,286		88,404,354	8q21-q22	-	-
EYA3		14 bp	28,287,732		28,287,684	1p36	BR	+
FSCN2		66 bp	77,109,947		77,109,997	17q25	-	-
GNAT1		44 bp	50,204,027		50,204,047	3p21	-	-

Gene		bp				Cytoband		
GRK7		40 bp	142,979,632		142,979,72	3q21-q23	BR	-
GUCA1B		19 bp	42,270,691	42,270,689-42,270,672	42,270,649	6p21.1	PB	-
GUCY2D		30 bp	7,846,687			17p13.1	-	-
HPCA		15 bp	33,124,670		33,124,680	2p25.1	MU	+
IMPDH1		206 bp	127,837,736			7q31.3-q32	SP	+
IMPG2		23 bp	102,522,132		102,522,10	3q12.2-q12.3	BR	-
LRRC21		121 bp	85,991,318	85,991,280	85,991,194	10q23	PB	-
MPP4		30 bp	202,271,692	202,271,616	202,271,60	2q33.2	PB	-
OPN4		67 bp	88,404,287			10q22	SP	-
PDC		17 bp	184,696,879		184,696,86	1q25.2	MU	-
PDE6H		7 bp	15,017,238		15,017,243	12p13	SP	-
PRPF31		117 bp	59,310,532		59,310,609	19q13.42	MU	-
RdCVF		37 bp	17,432,763	17,432,753-17,432,735	17,432,716	19p13.11	MU	-
RDH5		39 bp	54,400,449			12q13-q14	SP	-
RDH8		63 bp	9,984,862	9,984,925	9,984,991	19p13.2-p13.	MU	-
RDS		61 bp	42,798,348		42,798,296	6p21.2-p12.3	BR	-
RLBP1		83 bp	87,566,008		87,565888	15q26	-	-
ROM1		215 bp	62,136,883		62,137,111	11q13	BR	-
RPGR		7 bp	38,071,739		38,071,704	Xp11.4	BR	-
TULP1		70 bp	35,588,693		35,588,651	6p21.3	PB	+
CRB2		52 bp	125,158,322		125,158,32	9q33.2	-	-
CRX	Shortening exon 1 of	60 bp	53,016,971	53,017,001	53,017,005	19q13.3	SP	-
RP1		53bp	55,691,206		55,691,233	8q11-q13	BR	+
WDR17		77 bp	177,224,132	177,224,169-177,224,185	177,224,20	4q34	MU	+
ABCA4			94,359,290		94,359,245	1p22.1-p21	BR	-
C14ORF2			103,457,619			14q32.33	-	+
C1QL2			119,632,934			2q14.2	-	-
C7orf9			25,427,912			7p21-p15	-	-
CHM	Previous 5' end confirmed		85,189,194		85,189,222	Xq21.2	BR	+
ELOVL4			80,714,034			6q14	-	+
OPN1SW			128,203,084			7q31.3-q32	SP	-
RAX			55,091,570			18q21.32	SP	-
RBP3			48,010,989		48,011,000	10q11.2	-	-

RDH11		67,232,201		67,232,186	14q24.1	**SP**	**-**
RHO		130,730,174			3q21	**SP**	**-**
RPL14		40,473,831			3p22	**SP**	**+**
RS1		18,600,144			Xp22.2-p22.1	**SP**	**-**

Retinal expressed genes with new 5' exons

For 8 genes (*C1orf32, CNGA3, DHRS3, ELOVL5, KIFC3, RCV1, RDH12, SLC24A2*) we have identified a new exon composition at the 5' end of the transcript and in some cases new untranslated 5' exons that locate the TSS several kilobases upstream or downstream from the annotated one.

C1orf32: This transcribed locus in chromosome 1 was selected for its retinal expression. For this gene, whose function is still not characterised, we retrieved two new isoforms lacking the first annotated exon found in RefSeq. These isoforms contain TSSs in two new exons. One form displays a new first exon located 3 kb downstream from the previous TSS. The other form presents a first exon located 58.4 kb upstream of the former TSS, generating a new first intron spanning a locus transcribed in the opposite strand, the gene *MAEL* (Figure 1).
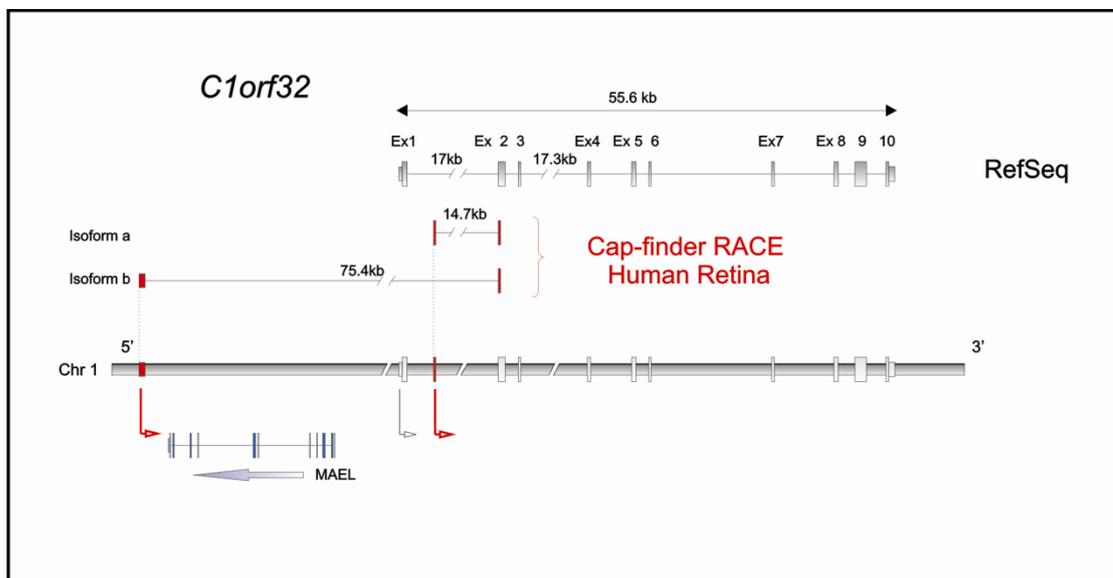
Figure 1: Schematic gene structure of the C1orf32 gene. *Transcripts in human retina: isoform a contains a new first exon located 2.7 kb downstream from previous TSS, isoform b presents a first exon 58,5 kb upstream the previous TSS. This second transcript defines an intron containing another transcribed locus in the opposite strand, the gene MAEL.*

Schema of the RefSeq, ESTs, exonic structure of new isoforms identified with the Cap-finder RACE of human retina mRNA and the genomic structure containing the TSSs indicated by arrows. Red Arrows indicate new retina TSSs.

CNGA3 (cyclic nucleotide gated channel alpha 3): Codes for the α-subunits of the cone photoreceptor cGMP-gated channel, a crucial component of the cone phototransduction cascade in color vision. Mutation in this gene causes achromatopsia. The RACE experiment confirmed the presence of 4 isoforms, all containing a splicing of untranslated exon 0 localized 23,4 kb upstream of exon1 (Wissinger et al., 2001) (Figure 2).
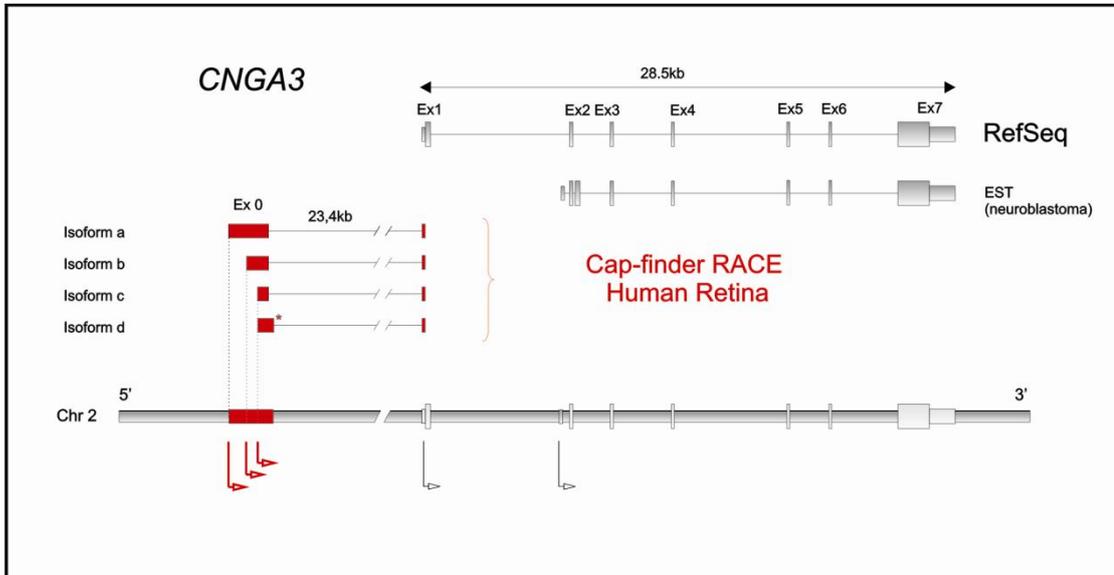
Figure 2: Schematic gene structure of CNGA3. *Cap-finder RACE of human retina mRNA confirmed the presence of an untranslated exon localized 23.4 kb upstream of exon1. One isoform contains alternatively spliced variant of exon 0 (asterisk).*

RDH12 (Retinol dehydrogenase 12): Is an enzyme with dual-specificity retinol dehydrogenases that metabolize both all-*trans*- and *cis*-retinols, reported to be expressed in photoreceptors (Haeseleer et al., 2002). Mutations within *RDH12* cause both recessive early onset Retinitis Pigmentosa and Lebers Congenital Amaurosis (Janecke et al., 2004; Perrault et al., 2004). In human retinal mRNA we retrieved two forms of the transcripts containing a new first exon located upstream of the RefSeq TSS and a differentially spliced second exon. The *in silico* assembly and experimental pipeline allowed us report three putative TSSs for this gene; the first is defined by the RefSeq annotation, the second was deduced from the most upstream transcript represented by ESTs from pooled colon and the third is a new TSS displayed by retinal transcripts (Figure 3).
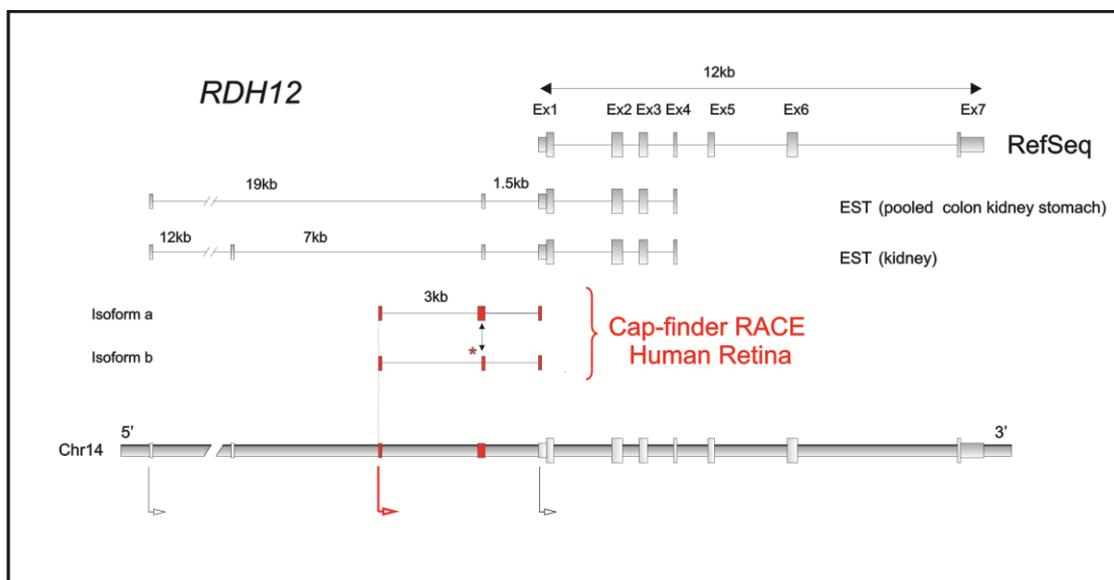
Figure 3: Schematic gene structure of RDH12. *The schematic representation of EST from human retina shows two forms of the transcripts starting in a new retina TSS. These two isoforms contain two alternatively spliced variants of the second exon (asterisk).*

DHRS3 (dehydrogenase/reductase, SDR family, member 3): Codes for an enzyme catalyzing the reduction of all-*trans*-retinal to all-*trans*-retinol in the presence of NADPH (Haeseleer et al., 1998; Haeseleer et al., 2002). The gene was included in our study for its high expression in retina. Cap-finder RACE confirmed the previous first exon and TSS. We also detected an alternative TSS in a new first exon downstream from the annotated one which was predicted with FirstEF (Davuluri et al., 2001) (Figure 4).
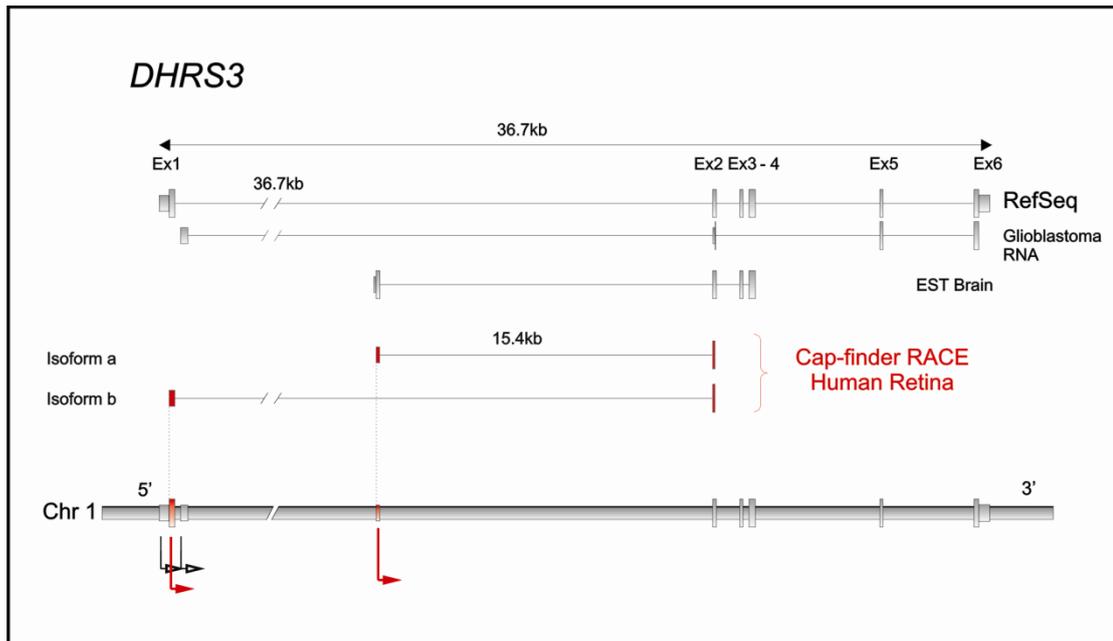
Figure 4: Schematic gene structure of DHRS3*. Schema of the RefSeq, ESTs, exonic structure of new isoforms identified with Cap-finder RACE of human retina mRNA and a schema of the new genomic structure containing new TSSs (indicated by red arrows). The Cap-finder RACE allow us to confirm the first exon and TSS of this gene. We also show the presence in retina of an alternative form of transcript containing a first exon 21 kb downstream the annotated TSS.*

ELOVL5 (elongation of long chain fatty acids, including docosahexanoic acid (DHA), family member 5): This gene was recently annotated as a retinal expressed gene (Yoshida et al., 2004) and a target of mutation studies in retinitis pigmentosa (Barragan et al., 2005). We detected a new form of the transcript with a new first exon that was not previously annotated or described for retina (Figure 5).
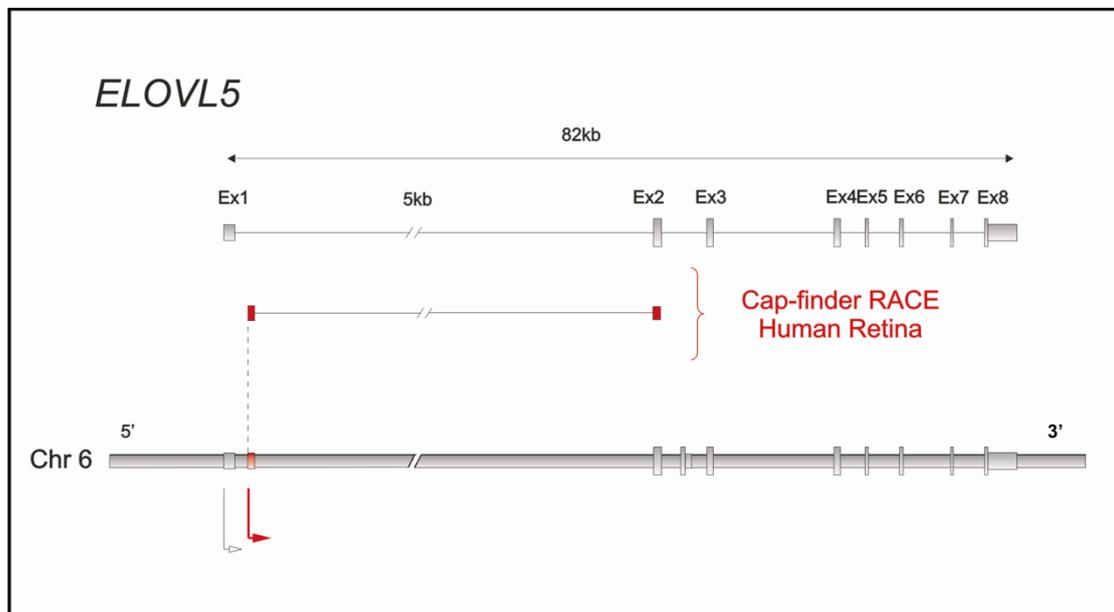
Figure 5: Schematic gene structure of ELOVL5. *Schema of the RefSeq, ESTs, exonic structure of new isoforms identified with Cap-finder RACE of human retina mRNA and a schema of the new genomic structure containing new TSSs (indicated by red arrows): we show the presence in retina of an alternative transcript with a first exon downstream from the annotated one.*

KIFC3 (Kinesin family member C3): codes for a retina specific microtubule-associated force-producing protein that may play a role in intracellular transport (Hoang et al., 1999). We have characterized two new isoforms of *KIFC3* retinal transcripts which lack the first 3 exons annotated in RefSeq. Both transcripts include a new first exon that localizes these new TSSs 44 kb upstream and 27 kb downstream respectively from the TSS referenced in the RefSeq database. The more upstream start site locates the gene in proximity to another retina specific gene (*CNGB1*) (Figure 6).
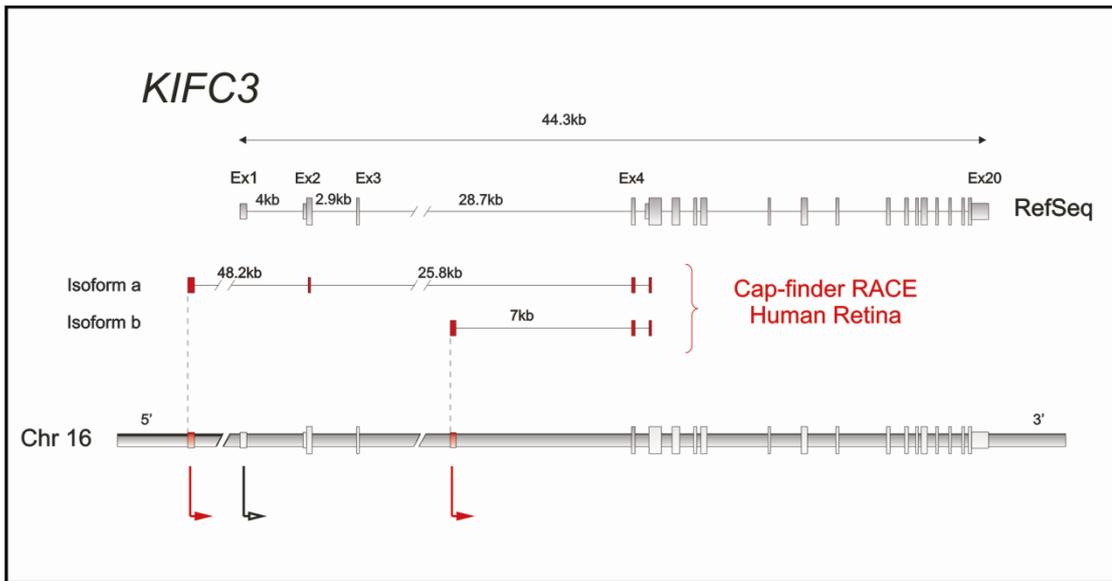
Figure 6: Schematic gene structure of KIFC3. Schema of the RefSeq, ESTs, exonic structure of new isoforms identified with Cap-finder RACE of human retina mRNA and a schema of the new genomic structure containing the TSSs (indicated by red arrows): figure shows two new isoforms lacking the 3 first exons annotated in the RefSeq. Both transcripts let us to define new TSSs located respectively 44 kb upstream and 27 kb downstream from the previous TSS.

RCV1 (recoverin): Inhibits rhodopsin kinase activity in retinal photoreceptors by reducing the binding of arrestin to rhodopsin. Deregulation of recoverin expression in certain types of cancer demonstrates a pathological role in cancer-associated retinopathy (Maeda et al., 2000). Although a previous study of the promoter was performed (Wiechmann and Howard, 2003), no clear evidence of the TSS have been described. For this gene we detected three alternative transcripts; the first with the same 5' end as the previously annotated TSS (first exon length may vary from 203 bp longer to 444 bp shorter), the second with a more frequent isoform lacking the first exon and starting 80 bp upstream from the second exon of the RefSeq and the third form has a new first exon located downstream from the annotated one (Figure 7).
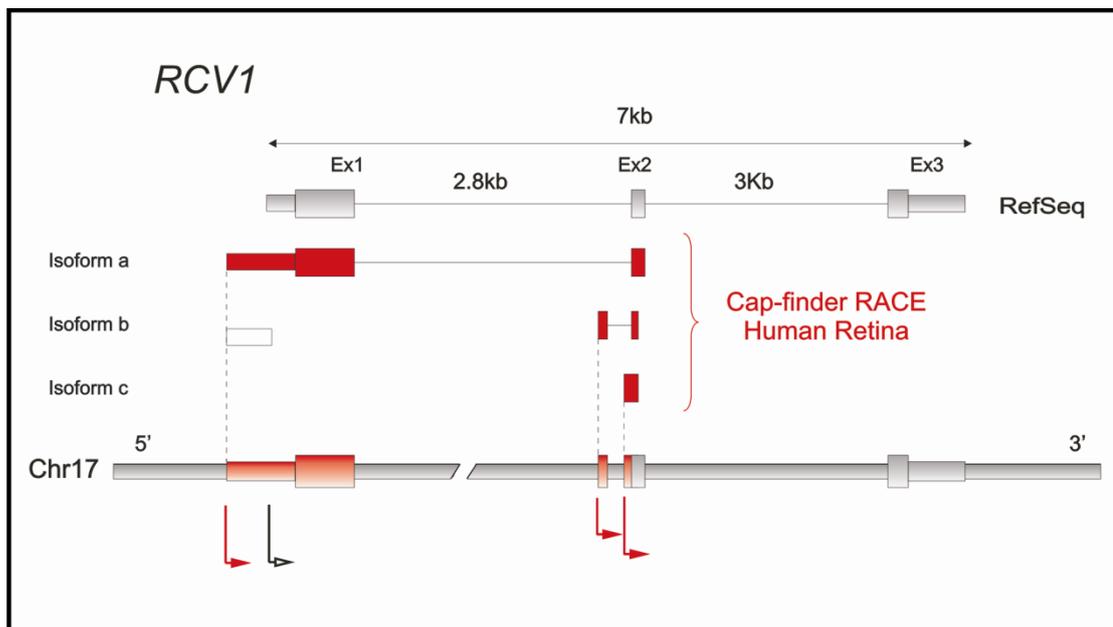
Figure 7: Schematic gene structure of RCV1. Schema of the RefSeq, ESTs, exonic structure of new isoforms identified with Cap-finder RACE of human retina mRNA and a schema of the new genomic structure containing new TSSs (indicated by red arrows). The figure shows three isoforms that we obtained with RACE experiments: one form extending the first exon by 203 bp and the other two forms lacking the first annotated exons of the RefSeq and, respectively, containing one new first exon that splices with the second annotated one and one starting at the second exon but extending it by 80 bp. For both transcripts the TSS is downstream from the annotated one.

SLC24A2 (solute carrier family 24, sodium/potassium/calcium exchanger, member 2): Codes for a potassium-dependent sodium-calcium exchanger in cone photoreceptor (Li et al., 2002). Although variant alleles of the cone *SLC24A2* gene have been identified, none of them are definitively associated with a specific retinal disease (Sharon et al., 2002). The new model we present for *SLC24A2* predicts three putative TSSs located in two new additional exons that are alternatively spliced (Figure 8).
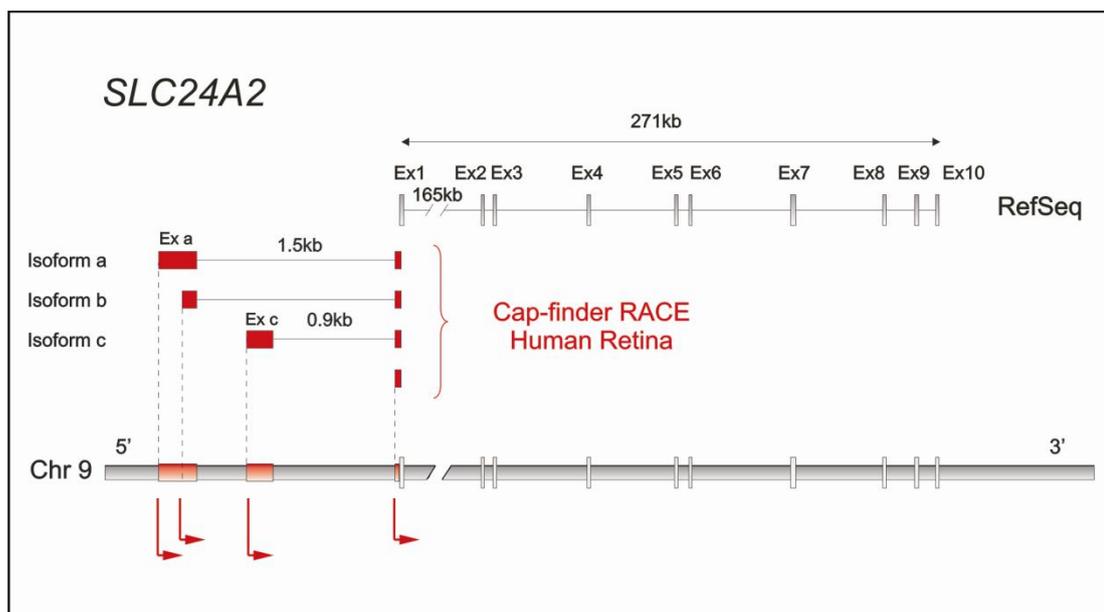
Figure 8: Schematic gene structure of SLC24A2. Schema of the RefSeq, ESTs, exonic structure of new isoforms identified with Cap-finder RACE of human retina mRNA and a schema of the new genomic structure containing new TSSs (indicated by red arrows): 4 new cDNAs clones from human retina identify 4 new TSSs for this gene. The transcripts contain two additional exons, a and b, that are alternatively spliced.

We also investigated whether the new exons that extend the 5´ end of the transcript may introduce new potentially protein coding sequences. We didn't observe in any case an extension of the open reading frame beyond the annotated start codon. However short alternate open reading frames of at least 40 codons were observed for *C1orf32* (nucleotide position 18-290 from the TSS in isoform a, and position 164-400 in isoform b, *CNGA3* (position 166-315), *DHRS3* (position 166-315), *KIFC3* (position 4-195), and *SLC24A2* (position 55-183 isoform a, 44-289 isoform b). Yet the translated sequences of these short ORFs do not have homology with any protein in public databases.

Detection of novel splicing variants and shorter transcripts

Our experimental procedure described alternatively spliced isoforms for two genes *IMPG1, SAG*, which lack exon 2 of the RefSeq. These forms have not been annotated in the RefSeq database. We confirmed these alternatively spliced isoforms by regular

RT-PCR (Data not shown). The second exon of the gene *SAG* contains the TSS and the presence of this alternative form, lacking the regular start site, may play a role in the regulation and further processing of the transcript. For 4 genes (*CRB2, CRX, RP1, WDR17*), we detected shorter transcripts that lack the annotated start codon. Since these experiments were done with the same adapter ligated first-strand cDNA we assume that these short transcripts are derived from true alternative TSSs. These transcripts may be preferentially amplified in the RT-PCR and may be translated from an internal initiation codon. We report in Table 2 the detailed results for these genes.

Confirmation of results with primer extension

To provide an experimental validation of our results we undertook primer extension experiments. We performed reverse transcription of mRNA with a sequence-specific FAM-labelled primer for two genes (CNGA3, RDH12). The length of the FAM-labelled cDNA primer extension product can be analyzed on ABI-DNA Genetic Analyzer using GeneScan software. As a result of the analyses we detected a fragment of 350 bp for *CNGA3* (Fig 9A) and a fragment of 215 bp for *RDH12* (Figure 9B). The size of these fragments confirms the presence of the transcripts that we detected with RACE.
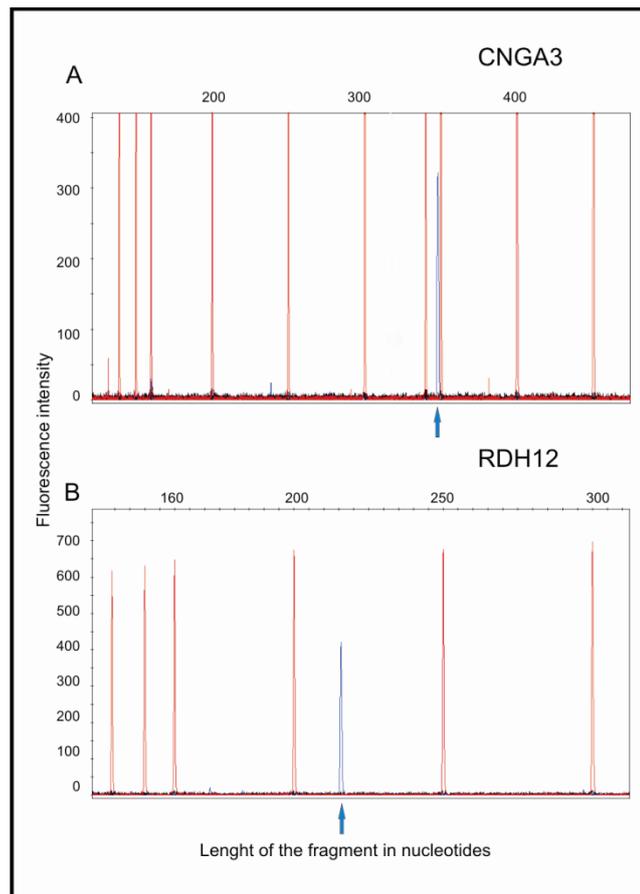
Figure 9: Primer extension results from WERI-RB1 retina cell lines mRNA. Primer extension products obtained with the gene specific primer for CNGA3 (A) and for RDH12 (B). The blue peaks in each panel correspond to the primer extension product

Comparison with existing annotations and databases

To assess the quality of current annotations of the 5' end of genes expressed in human retina, the sequences obtained by 5' RACE were compared with the corresponding gene annotation/prediction. Overall, RACE experiments detected 15 exons that were neither annotated nor predicted for retina transcripts; 8 exons did not have any matching experimental evidence in GenBank, while the other 7 showed different boundaries or alternative splice sites. Of these 15 un-annotated exons, 12 are first exons and can be considered the new first exon for the retinal transcripts. Of the 54 genes successfully amplified, 41 (76%) delivered 5' RACE sequences different from the annotation. Results of a parallel project, DBTSS (Suzuki et al., 2004), supported our results concerning 3 of these genes (*CNGA3, ELOVL5* and *SLC24A2*) although the source of mRNA was not human retina. We extended the annotated first exon of 27 RefSeq genes by an average of 60 transcribed bases. We compared our results with genome wide mapping of TSSs using CAGE tags (Carninci, 2006). We found perfect correspondence for 13 transcript isoforms; for another 6 transcripts the start site retrieved in the CAGE database is located less than 400 bp away. For 35 transcript isoforms the TSS is located in a different position (See Table 2). This discrepancy in the results may be due to the fact that the CAGE database doesn't include retina amongst the panel of analyzed tissues and therefore lack specific and rare transcript isoforms present in that tissue.

Shape of TSSs and conservation

After analyzing the distribution of RACE clones we could define the shape of TSSs according to the classification previously reported (Carninci et al., 2006). The different clones were clustered and depending on the start base position of each clone within a cluster we divided the start sites into four shapes. In the single dominant peak class (SP) the majority of clones are concentrated to no more than four consecutive start positions with a single dominant TSS. The clusters spanning a broader region are grouped in a general broad distribution (BR), a broad distribution with a dominant peak (PB) and a bi- or multimodal distribution (MU): 22 genes showed a single dominant peak, 11 a broad distribution, 8 a bi-multi peak distribution and 6 a broad distribution with a dominant peak. For some transcripts we could not make a classification because the number of clones was less than 5. We report the results of this analysis Table 2 (TSS shape). Figure 10 shows a graphical view of TSSs identified for *AOC2*, *ABCA4*, *RDH12*, and *LRRC21* as an example of the different distributions observed. The classification of the shape of TSSs defined by distribution of 5' end RACE clones within a cluster is useful for the further characterization of

expression regulation. The distribution of the clones defines different elements of the core promoter and gives insights on the start of transcription. Even if broad promoters are the major class in mammals (Carninci et al., 2006), 36 of the analyzed transcripts present a dominant peak highlighting the possibility that those transcripts are tightly regulated.
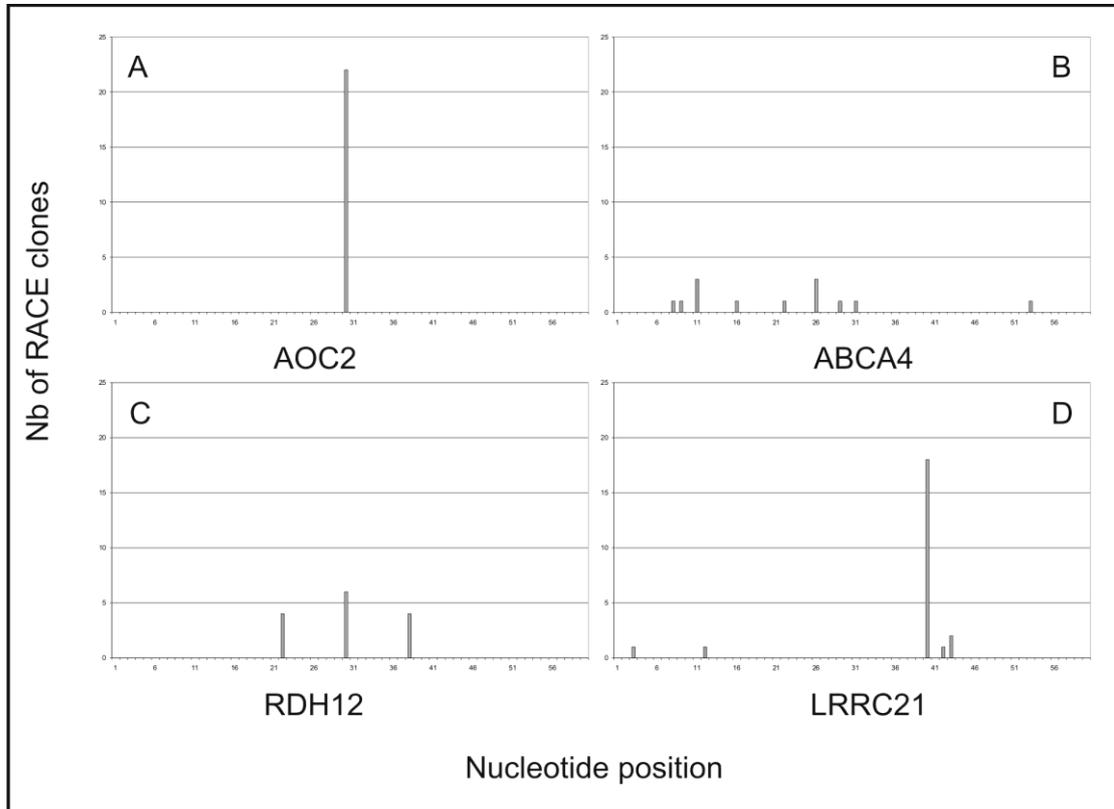
Figure 10: TSSs present different shapes. Histograms indicate the number of RACE clones mapping at each nucleotide position. Examples show the different pattern that we observed during the analysis of the Cap-finder RACE. A) Clones distribution for AOC2 (single peak class: SP). B) Clones distribution for ABCA4 (broad: BR). C) Clones distribution for RDH12 (multimodal: MU). D) Clones distribution for LRRC21 (broad with dominant peak shape: PB).

Although TSSs of orthologous genes do not necessary reside on equivalent locations because of evolution of mammalian TSSs (Frith et al., 2006), we analyzed sequence conservation of the first new exons among a set of mammals (mouse, dog, cow): the range of conservation varies between 42 and 89 %. We report the pairwise alignment percentage of identity Table 3.

Table 3: Sequence conservation of new first exons: Analysis of sequence conservation of the first new exons of the listed human transcript in comparison with a set of mammals (mouse, dog, and cow). Numbers indicate percentage of identity. (Numbers indicate percentage of identity)

Transcripts	Dog	Mouse	Cow
C1orf32 Isoform a	42	45	61
C1orf32 Isoform b	73	73	n.a.
CNGA3	62	51	71
DHRS3	n.a.	44	89
ELOVL5	71	48	63
KIFC3 Isoform a	50	54	64
KIFC3 Isoform b	44	43	47
RCV1	47	53	58
RDH12	45	44	48
SLC24A2 Isoform a	56	58	45
SLC24A2 Isoform c	77	66	45

Sequences residing upstream and downstream from the boundaries of new defined exons are regions displaying high regulatory potential calculated by a computational algorithm (King et al., 2005) integrated in the UCSC genome browser. The regulatory potential (RP) scores computed for the 500bp sequence upstream the TSS shows that in 9 new first exons out of 11 the RP value exceeds an arbitrary threshold of 0.2 (data not shown). Considering 500bp downstream the splice site of the first exon the RP value is > 0.2 at least for 7 first exons out of 11. This observation confirms the importance of the new described exons to locate new regulatory elements that are important for transcription in retina.

A high level conservation was observed for splice donor sites of the new first exon. 5 genes show an average conservation of at least 75% in the region -3 /+5 spanning the splice donor site. For example, we report an *inter*-species alignment of the 3' end of exon 0 (*CNGA3*). The sequence conservation at the level of the splice donor site highlights the possibility of a particular role for this splicing (Figure 11).
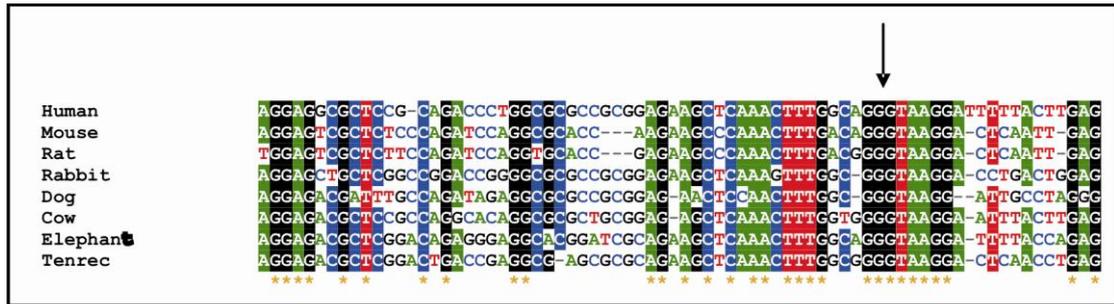
Figure 11: Inter-species alignment of exon-intron boundaries of the exon 0 of CNGA3. Conserved nucleotides are labelled with colors and with the star in the bottom those conserved in all the analyzed species (human, mouse, rat, rabbit, dog cow, elephant and tenrec). Arrow highlights the splice donor site.

Discussion

Now that the information pertaining to the genomes of human and other animals is available, the next challenge for genetic studies is to map the TSSs and regulatory sequences of the genes. Here, we carried out a study to determine the genuine TSSs of a pool of retinal expressed genes. The results give correct information about the complete 5' end of transcripts and this data will be useful to locate the respective core and proximal promoter elements. We chose Cap-finder RACE to map the TSS of retinal expressed genes because this technique is selective for the complete transcript (Dike et al., 2004; Suzuki et al., 2004). For 41 out of 54 successfully amplified genes, the Cap-finder RACE experiments detected transcripts which are different from the current gene annotation in most cases the RefSeq was incomplete. Transcripts were missing part of the first exon or even complete exons at the 5' end. This experimental determination of TSSs shows that the current gene annotation was in most cases obtained from data sources that are not strictly selective for the complete transcribed form, and need to be updated. This procedure led us to discover several transcript isoforms that were un-annotated and to locate retina specific TSSs. Proteins encoded by these genes are essential for retina function and stability. A mutation in the *cis*-regulatory elements may influence the level of transcription and have a strong effect due to sensitivity of photoreceptors for high level transcription of genes involved in phototransduction. The new described *cis*-regulatory regions and untranslated exons are possible targets for mutation studies in retinal disorders. Therefore new isoforms give a more complete picture of alternative start sites use in retina genes. The 5' untranslated region may contain important transcriptional and post transcriptional regulatory sites (Oyama et al., 2004; Rogozin et al., 2001; Sachs and Geballe, 2006) and therefore only the complete 5' UTR provides the opportunity to study the potential regulatory role of these non-coding sequences. New reported TSSs contribute to the identification of regulatory elements active in tissue specific gene regulation (Hochheimer and Tjian, 2003; Zhang et al., 2004). Moreover, bioinformatics tools that identify common regulatory elements rely on the correct determination of TSSs within

a particular tissue (Bortoluzzi et al., 2005); therefore these computational approaches will only be effective after experimental validation of the 5' end of transcripts.

Conclusions

We herein report the TSSs for 54 retina expressed genes. Our results define new and more precise locations of TSSs for 76% of the analyzed genes; moreover in 15% of the genes we found new exons in the 5' end of the transcripts. Thus, this analysis of TSSs in human retina was essential to define the complex pattern of transcripts present in this tissue.

Our results highlight the importance of applying a tissue specific approach with a systematic program of Cap-finder RACE using the known gene structures as a starting point and/or gene predictions to complete the existing gene annotation. The new TSSs and transcribed sequences provide crucial information for further exploration of the promoter and other *cis*-regulatory sequences at the 5´end of the gene, and in particular for the study those elements that are functionally active in human retina.

Authors' contributions

This chapter has been taken from the article: "Mapping of transcription start sites of human retina expressed genes" by Valeria Roni, Ronald Carpio, and Bernd Wissinger, published in BMC Genomics. 2007; 8: 42. From a total of 54 results of the analyzed genes my contributions (RC) have been involved in acquisition, analysis and interpretation of data for 48 genes of them. VR contributed to acquisition, analysis and interpretation of data of 6 genes; BW conceived and coordinated the study and contributed materials and resources.

CHAPTER II: IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF THE HUMAN *CNGA3* GENE PROMOTER

Introduction

The vertebrate neural retina contains a limited number of different cell types generated in a well-characterized order during development to create a laminate structure (Adler, 1993). High-acuity vision and color vision in humans relies on the presence and functional integrity of cone photoreceptor (Gegenfurtner and Kiper, 2003). Located in the photoreceptor outer segments, cyclic nucleotide-gated (CNG) channels play an important role in the visual phototransduction process (Kaupp and Koch, 1992; Yau and Baylor, 1989). Photoreceptor CNG channels are non selective cation channels that are open at high cGMP levels in the dark (dark current) and are closed with dropped cGMP levels upon light stimulation, causing hyperpolarization of the plasma membrane and decrease of the synaptic glutamate release. The native cone CNG channel consists of two CNGA3 and two CNGB3 subunits (Peng et al., 2005; Weitz et al., 2002; Zheng et al., 2002; Zhong et al., 2002). In humans, the 694 amino acids CNGA3 protein (Biel M, 1999; Finn et al., 1996; Zagotta and Siegelbaum, 1996) is coded by the human *CNGA3* gene located at chromosome 2 (2q11.2). Mutations in the *CNGA3* gene cause achromatopsia, which is a stationary retinal disorder with autosomal recessive inheritance characterized by loss of color discrimination, low visual acuity, photophobia and nygtasmus (Goto-Omoto et al., 2006; Sharpe, 1990). Analysis of CNGA3 deficient mice showed that the loss of cone function correlates with a progressive degeneration of cone photoreceptors but not of other retinal cell types (Biel et al., 1999). Up to now little is known about the regulatory mechanisms of the *CNGA3* gene expression. In previous studies expression of *CNGA3* was detected in retina, sperm, kidney, heart, cerebral cortex, cerebellum and other tissues (Ahmad and Barnstable, 1993; Ahmad et al., 1990; Biel M, 1999; Cassar et al., 2004; Finn et al., 1996; Zagotta and Siegelbaum, 1996). In addition, we have previously investigated the position of some TSS of *CNGA3* as part of a larger study (Roni et al., 2007).

In this study we have characterized the promoter of the *CNGA3* gene in order to analyze its transcription regulation. We identified the presence of *cis*-acting elements on the promoter which could respond to activation or repression, thus providing an explanation to its reporter activity. Our data show that transcription of the *CNGA3* gene can start within two alternative exons, where only one of them exhibit promoter activity in reporter gene assays. Interestingly, this promoter is regulated by the transcription factor Yin Yang 1 (YY1), suggesting a potential role as both an activator and repressor of transcription initiation according to the cellular context.

Material and Methods

Bioinformatics

Sequences were retrieved using the pureGene database of iDOUBT websuite (http://idoubt.fr) at IGBMC. For all studies on the *CNGA3* gene, the reference mRNA sequences (RefSeq Build 32, www.ncbi.nlm.nih.gov/RefSeq) of human (NM_001298; Wissinger et al., 1997) and mouse (NM_009918; Gerstner et al., 2000) were taken into consideration. Analysis of the gene structure was based on the public version of the human genome (GenBank release 168.0). In order to map the transcription start sites (TSS) of the *CNGA3* gene, 5′-EST's and transcript tags were identified by *in silico* analysis using the UCSC genome browser database (http://genome.ucsc.edu), the database for transcription start sites (DBTSS; http://dbtss.hgc.jp/; Suzuki et al., 2002) and the CAGE database (Shiraki et al., 2003) with the help of the CAGE basic viewer (http://fantom3.gsc.riken.jp/).

To identify promoter regions and putative transcription factor binding sites, an analysis of the upstream region of human *CNGA3* was performed using the ElDorado, Promoter Inspector and MatInspector Professional program (http://www.genomatix.de; Quandt et al., 1995) relying on the version 7.0 and later version 8.0 of its transcription factor database MatBase. We took special attention for eye-specific and core promoter transcription factors response elements using a matrix–specific optimized threshold as cut-off. CpG-rich regions were searched via iDOUBT suite fully exploiting cpgIslandExt table at UCSC.

A prediction/filtering approach was performed for transcription factor identification relying on profile scan and phylogenetic filtering. Prediction filtering was done using multiple sequence alignment of 500-1000 bp upstream of the different TSS from human, dog, mouse and rat. The similarity hits were obtained using the MultiZ program (http://genome.ucsc.edu/; Blanchette et al., 2004). Shading of the multiple sequence alignment was carried out using the program BioEdit version 7.0.5.3 (http://www.ch.embnet.org/software/BOX_form.html; Hall, 1999).

BLAT software (Kent, 2002) was used to query human genomic sequences of the *CNGA3* promoter region against the whole mouse genome. Fragments with size larger than 19 bp, corresponding to a single limit of a given fragment in the genome, were subsequently localized with respect to their nearest gene using GPS Localizer software. Only fragments located on the vicinity of a TSS or inside an intronic region were considered for further analysis.

Isolation of genomic DNA and retinal RNA

Genomic DNA was extracted from venous blood samples and HEK-293 cells using standard salting-out procedure. Total RNA was isolated from human, mouse and zebrafish retina using Trizol® Reagent (Invitrogen, Carlsbad, California). After isolation, RNA was treated with the DNA-free™ DNase treatment kit (Ambion, Austin, Texas) and 1-5 µg RNA was used to generate cDNA using SuperScript® First-

Strand Synthesis System (Invitrogen, Carlsbad, California) for RetroTranscriptase-Polymerase Chain Reaction (RT-PCR).

5'-RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE)

The 5'-ends of transcripts of *CNGA3* in human, mouse and zebrafish were obtained from retinal total RNA using the RLM-RACE Kit (Ambion, Austin, Texas) as described previously (Roni et al., 2007). Briefly, several gene-specific reverse primers were designed for each orthologous (Table 1) in order to amplify different TSS products within the first and the second exon. PCR products were cloned into the pCR2.1 vector (Invitrogen, Carlsbad, California) and sequenced using Big Dye Terminator chemistry on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Carlsbad, California).

Table 1: Primers list used to generate RLM-RACE products, luciferase reporter constructs, mutant-deletion constructs and in situ hybridization probes. All primers listed are on their respective 5'- to 3'-orientation.

5'-RLM-RACE
Human
`hCNGA3-Ex1-revA-CCTCACTTGACGAGTGGGCT`
`hCNGA3-Ex1-revB-TTTACCTTGAGGTGGGTCCT`
Mouse
`mCNGA3-EX4-revA-CAGCCAGCGGTAGTAGATGTT`
`mCNGA3-EX6-revA-CAAAGATCAGGATACCCACCA`
`mCNGA3-EX1b-revB-GCCCAGAGTAGCCAGTAGATT`
`mCNGA3-EX1-revA-TCAGAGAGGGAGATGATGAGA`
`mCNGA3-EX2-revA-GTGTTCACCTTTGCCATCTCT`
Zebrafish
`zfXM678842-EX3-revA-GCTACGATAGTCAGCCAGTGAT`
`zfXM678842-EX5-revA-CCCACAGCTTCTTGGATTCT`
`zfXR029476EX3-revA-ACCTCTCTGTGAAGCCTGTGT`
`zfXR029476-EX4-revA-GGGATTGGCTTTTTTCCTCTTA`
`zfXR029476-EX5-revA-TGGAGGAGGAGCATCTTTCT`
`zfXM690534-EX3-revA-ATGGCCCCAGCACCTGTGAA`
`zfXM690534-EX4-revA-ACAGTTGTTCATATTGTAAACTGCCAGA`
`zfXM690534-EX5-revA-TCATCCTTTTCATCTTTTTTCTCATCC`
`zfXM1340931-EX1-revA-GATGAAGCAGTTGGTGATCCT`
`zfXM1340931-EX4-revA-TGTCATCTGTGTTGTTGCAGTTGTTC`
`zfXM1340931-EX5-revA-TGTCATCCTTCTTTTCGTCCT`

Luciferase Reporter Constructs

```
hCNGA3-ExU1-A-forNHEI-TTTACACACTGCTAGCACTCGATGAACTTAAGCCTTCT

hCNGA3-ExU1-A-revHINDIII-TTTACACACTAAGCTTGTGTAATCCACACTGGATGAGTA

hCNGA3-ExU1-B-revHINDIII-TTTACACACTAAGCTTCCTCTCCGTGGGAAAGTAAG

hCNGA3-Ex1-A-forNHEI(prom508NHEI)TTTACACACTGCTAGCGAATCGTGTATTGGGTTTAGC

hCNGA3-Ex1-A-revHINDIII-TTTACACACTAAGCTTAATGTGAACACAAAGCAGACAC

hCNGA3-Ex1-B-revHINDIII-TTTACACACTAAGCTTATCTTGGCCATCTTCTCGGT

hCNGA3-Ex2-A-forNHEI-TTTACACACTGCTAGCGTTAAGAAGGCTAATTGGGTTG

hCNGA3-Ex2-A-HINDIII-TTTACACACTAAGCTTCAGCACTGACGATGTCTCCT

hCNGA3-Ex1-A-prom996-NHEI-TTTACACACTGCTAGCCTCAAAAGGCCGATCTTAGG

hCNGA3-Ex1-A-prom988-NHEI-TTTACACACTGCTAGCCCGATCTTAGGCTCACAATA

hCNGA3-Ex1-A-prom638-NHEI-TTTACACACTGCTAGCCCACATTGACAGGTGGATCT

hCNGA3-Ex1-A-prom463-NHEI-TTTACACACTGCTAGCTGTCACATATGGTCCTTTTT

hCNGA3-Ex1-A-prom372-NHEI-TTTACACACTGCTAGCAAACAAGAAAACCAGGGTCA

hCNGA3-Ex1-A-prom292-NHEI-TTTACACACTGCTAGCCCAACTTGGCCCTTCATTTA

hCNGA3-Ex1-A-prom131-NHEI-TTTACACACTGCTAGCACAAGAGACAGCAGAGGGTGT

hCNGA3-Ex1-A-prom034-NHEI-TTTACACACTGCTAGCTCCTGATGACGTGTCTGCTT
```

Site Directed Mutagenesis Constructs

```
hCNGA3-Ex-1-YY1for- ATGTGACAAACCGAGAAGCAAGATAAGCTTGGCATTCCGGTACTGTT

hCNGA3-Ex-1-YY1rev- AACAGTACCGGAATGCCAAGCTTATCTTGCTTCTCGGTTTGTCACAT

hCNGA3-Ex-1-REV-ERBafor-CCTGTCCTGAGTCTGCTTTGTGTTCACATTTTAGC

hCNGA3-Ex-1-REV-ERBarev-GCTAAAATGTGAACACAAAGCAGACTCAGGACAGG

hCNGA3-Ex-1-VDR/RXRfor-GCGGTAGCCCTTGCCCAGCTGGGTTTGCAGTTACTTTCC

hCNGA3-Ex-1-VDR/RXRrev-GGAAAGTAACTGCAAACCCAGCTGGGCAAGGGCTACCGC

hCNGA3-Ex-1-SRYfor-GCATGGTAGCATAAGAGTAAAAGACTTTGTGCCTTATGC

hCNGA3-Ex-1-SRYrev-GCATAAGGCACAAAGTCTTTTACTCTTATGCTACCATGC

hCNGA3-Ex-1-NANOGfor-CCCTTCATTTATTTATTCATTCACAGTACTTAACTCTGAGC

hCNGA3-Ex-1-NANOGrev-GCTCAGAGTTAAGTACTGTGAATGAATAAATAAATGAAGGG

hCNGA3-Ex-1-HBP1for-GGCCCTTCATTTATTTATTCTCATTCAGTACTTAACTCTGAGC

hCNGA3-Ex-1-HBP1rev-GCTCAGAGTTAAGTACTGAATGAGAATAAATAAATGAAGGGCC

hCNGA3-Ex-1-HOXB9for-GGTCCTTTTTTTACATGTTGTCCAAGATTTGAGGGAGCAGTGG

hCNGA3-Ex-1-HOXB9rev-CCACTGCTCCCTCAAATCTTGGACAACATGTAAAAAAAGGACC

hCNGA3-Ex-1-BRN2for-GGGTTTAGCATTATAAGGTTTTCTGAATGTCACATATGGTCC

hCNGA3-Ex-1-BRN2rev-GGACCATATGTGACATTCAGAAAACCTTATAATGCTAAACCC

hCNGA3-Ex-1-PTX1for-GGATTAATGAGATGAATCGTGTATTGGCATTATATTGTAAAGG

hCNGA3-Ex-1-PTX1rev-CCTTTACAATATAATGCCAATACACGATTCATCTCATTAATCC
```

Human CNGA3 in situ hybridization probes

```
hCNGA3-ExU1-fwdA-T7-TTCTAATACGACTCACTATAGGGGGATTACACTACCAGCAAAACA

hCNGA3-ExU1-fwdB-T7-TTCTAATACGACTCACTATAGGGAAACATCCCAGGAAGGAGTC

hCNGA3-ExU1-fwdA-SP6-ATTTAGGTGACACTATAGAACCTGCCAAAGTTTGAGCTTCT

hCNGA3-ExU1-fwdB-SP6-ATTTAGGTGACACTATAGAATGCCAAAGTTTGAGCTTCTC

hCNGA3-Ex1-3-fwdA-T7-TTCTAATACGACTCACTATAGGGGGGGGCTAAATGTGACAAAC

hCNGA3-Ex1-3-fwdB-T7-TTCTAATACGACTCACTATAGGGAATGTGACAAACCGAGAAGATG

hCNGA3-Ex1-3-fwdA-SP6-ATTTAGGTGACACTATAGAACTTTCTTGGCTGGACACCTC

hCNGA3-Ex1-3-fwdB-SP6A-TTTAGGTGACACTATAGAACTTGGCTGGACACCTCCTTA
```

Generation of reporter gene constructs

To verify regulatory activity of the human *CNGA3* promoter, different fragments of
the *CNGA3* locus from 48 to 1028 bp in length were amplified from genomic DNA by

PCR, used as a template to amplify ExU1(-628/+23) and ExU1(-628/+233) with respect to TSS.1; and Ex1(-508/-3), Ex1(-508/+49), Ex1(-996/+49), Ex1(-988/+49), Ex1(-638/+49), Ex1(-463/+49), Ex1(-372/+49), Ex1(-292/+49), Ex1(-131/+49), Ex1(-34/+49), Ex2 (+6763/+7791) with respect to TSS.2. PCRs were done with forward primers containing *Nhe*I site and reverse primers containing a *Hind*III or *Bgl*II site near their respective 5′-ends (shown in Table 1). The PCR was performed using Pfu DNA polymerase (Stratagene, La Jolla, California) as follows: 94 °C for 4 min, followed by 35 cycles of 94 °C for 20 sec, 58 °C for 20 sec, and 72 °C for 0.5-2 min, with a final extension at 72 °C for 8 min. The PCR products were gel-purified, digested with *Nhe*I and *Hind*III or *Bgl*II and cloned into the pGL3-enhancer luciferase vector (Promega, Mannheim, Germany). Resultant hCNGA3-pGL3Enhancer clones were sequenced to verify identity and sequence integrity. Sequencing was also performed from *CNGA3* PCRs from HEK293 cells to verify the same nucleotide sequence to the corresponding human sample (data not shown). Transfection quality DNA was isolated using the GeneJET™ Plasmid Miniprep Kit (Fermentas, St. Leon Rot, Germany) for the pGL3 luciferase constructs.

Generation of deletions and extensions in the *CNGA3* promoter

To generate deletion or extension constructs for promoter characterization, genomic fragments were amplified by PCR and cloned into pGL3-Enhancer as described above.

Generation of site-directed mutagenesis in the *CNGA3* promoter

Site-directed mutagenesis was carried out on single putative transcription factor binding sites in the *CNGA3* promoter using the QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, California). Constructs with 5 bp deletions in the core region of several transcription factor binding were generated for inserts Ex1(-508/-3), Ex1(-508/+49), Ex1(-372/+49) and Ex1(-34/+49) by repeating primer-directed mutagenesis with different set of primers (Table 1). The PCR conditions were as follows: 95 °C for 2 min, followed by 16 cycles of 95 °C for 50 sec, 60 °C for 50 sec, and 68 °C for 6 min, with a final extension at 72 °C for 7 min. The PCR products were digested with *Dpn*I at 37 °C for 1 hour and transformed into DHα5 supercompetent cells (Invitrogen, Carlsbad, California). All mutations were confirmed by DNA sequencing.

Cell culture

Human retinoblastoma cell lines Y79 (Reid, 1974) and WERI-Rb1 (McFall et al., 1977) were cultured in RPMI-1640 medium supplemented with 20% and 10% fetal calf serum, respectively, and 1% penicillin/streptomycin (Sigma-Aldrich, Munich, Germany). Mouse cone photoreceptor-derived 661W cells (Otterson et al., 1997), human retinal pigment epithelial ARPE19 (Dunn et al., 1996) and human embryonic kidney HEK293 (Graham et al., 1977) cell lines were cultured in Dulbecco's modified Eagle medium, supplemented with 10% fetal calf serum and 1% penicillin/streptomycin (Sigma-Aldrich, Munich, Germany).

Transient transfection and luciferase reporter assay

A total of 2×10^4 cells/well (ARPE-19, WERI-Rb1, Y79, 661W, HEK293) were seeded into a 96-well culture plate in antibiotic-free medium one day prior to transfection. 0.2 µg of each reporter construct was transfected into the cells with Lipofectamine 2000 (Invitrogen, Carlsbad, California) according to manufacturer's instructions and transfected cells cultured for 48 h. The pRL-TK vector (Promega, Mannheim, Germany) harboring the Renilla luciferase gene was co-transfected (0.02 µg) as an internal control to normalize for unequal transfection efficiency. Each transfection was done in triplicate; three independent experiments were performed for each construct except when specified.

Luciferase activity was assayed for the reporter constructs using the Dual Luciferase® Assay System (Promega, Mannheim, Germany). Cells were washed in phosphate buffer saline (PBS) and then lysed in reporter lysis buffer (Promega, Mannheim, Germany). The cell lysate was used for assaying the firefly luciferase activities, using an Orion Microplate Luminometer (Berthold, Bad Wildbad, Germany). Each lysate was measured twice to verify the accuracy of the instrument. Promoter activities were expressed as relative light units (RLU) normalized for the activity of *Renilla* luciferase in each extract. The data were calculated as the mean of three identical independent experiments. Cells transfected with the activity of the promoter-less pGL3-enhancer vector were tested in each experiment to serve as a baseline.

RNA *in situ* hybridization on human eye cryosections

Digoxigenin antisense probes and sense control probes for RNA *in situ* hybridization (ISH) experiments were obtained by *in vitro* transcription of PCR amplicons of human retinal cDNA (DIG RNA labeling kit; Roche Diagnostics, Manheim, Germany) with primers containing T7 or SP6 promoters (shown in Table 1). Human eye cryosections were obtained from cornea donors collected by the Italian Eye Bank (Fondazione Banca degli Occhi del Veneto, Venice, Italy) as described previously (Trifunovic et al., 2008). Briefly, twenty micrometer cryosections were fixed in 4% paraformaldehyde (PFA) in PBS for 15 minutes. After being bleached with 6% H_2O_2 in PBS with 0.1% Tween 20, the sections were treated with 10 µg/ml proteinase K for 15 minutes, followed by postfixation with 4% PFA and 0.2% glutaraldehyde. Hybridization with digoxigenin-labeled probes (2 µg/ml) was performed overnight at 60°C. The hybridized sections were washed with 50% formamide, 4x buffer sodium citrate (SSC), and 1% SDS at hybridization temperature and with 50% formamide and 2x SSC at temperature 5°C below the hybridization temperature. Sections were blocked for 1 hour with 1% blocking reagent (Roche Diagnostics, Manheim, Germany) in MABT (100 mM maleic acid, 150 mM NaCl, 0.1% Tween 20; pH 7.5), containing 10% sheep serum, and incubated with alkaline phosphatase (AP)-labeled anti-digoxigenin antibody (1:2000; Roche, Mannheim, Germany) in MABT with 1% blocking reagent overnight at 4°C. After several washes with Tris-buffered saline containing 0.1% Tween 20 (TBST), the sections were exposed to a solution containing the substrate for AP, nitroblue tetrazolium and 5-bromo-4-chloro-3-indoyl phosphate

(NBTBCIP; Sigma-Aldrich, Munich, Germany) at pH 9.5. The reaction was blocked by washes with PBS (pH 5.5) followed by postfixation in 4% PFA for 20 minutes. The slides were coverslipped with 70% glycerol in PBS or dehydrated and mounted with Eukitt mounting medium (Sigma-Aldrich, Munich, Germany).

Calculation and statistics

Statistical analysis was performed using One-Way ANOVA with Dunnett's Multiple Comparison Test. GraphPad 5.0 software was used for statistical analyses. Significance is defined as $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***) and $p>0.05$ was considered non significant. Values are presented as means ± the standard error of the mean (SEM).

Results

Alternative transcription start sites in the human *CNGA3* gene

Since the knowledge of the transcription start site (TSS) of the *CNGA3* gene is a prerequisite for the identification and characterization of its promoter, RLM-RACE-PCR was performed on cDNA from human retina. Sequencing of cloned PCR products revealed two types of cDNA. Longer cDNA (TSS.1 in Figure 1A) representing transcripts that starts at 23,4 kb bp upstream from exon 1 (Roni et al., 2007; Wissinger et al., 2001), now denominated as a standard exon U1 (Johan T. den Dunnen, 2000) and shorter cDNAs that map to exon 1 (TSS.2 in Figure 1A). The ratio of shorter and longer RACE-clones containing TSS.1 or TSS.2 was approximate 1:1. Our experimental design was not selective for a third putative TSS.3 located downstream of exon 1 which is deduced from a cDNA of a retinoblastoma cell line (human transcript AK131300, Figure 2).

We also searched the Database of Transcription Start Sites (DBTSS, http://dbtss.hgc.jp/) and the CAGE database (http://fantom31p.gsc.riken.jp/cage/) for tags of *CNGA3* cDNAs. DTBSS confirmed our findings while contrary to our RLM-RACE results, the CAGE database contained only tags that map to retinal exons (data not shown).

In order to compare the human TSS results with other species, RLM-RACE-PCR was also performed on RNA from mouse and zebrafish retina. Sequencing of mouse cloned RACE-PCR products also revealed two size groups of cDNAs. The shorter cDNA fragments define a novel exon U1 upstream of exon 1 (TSS.2 in Figure 1A). Second, longer cDNAs (TSS.1 in Figure 1A), contained a novel exon U2 that splices with a shorter variant of exon U1. Analysis in zebrafish using RLM-RACE revealed a single TSS in exon U1 for each of the two respective *CNGA3* loci.

Comparison of the human and mouse *CNGA3* genes (Figure 1A, dotted lines) showed sequence homology between the human exon U1 and the mouse exon U2. However, no equivalence for the mouse exon U1 was found in the human genome. Moreover, when searching for equivalence of the mouse exon U1 in other species it was only

possible to find an equivalent sequence within the rat genome in a region of the first intron of the *CNGA3* gene (data not shown).

When independently comparing the human exon U1 and the first coding exon 1 with other species, we found a robust alignment among different vertebrates (human, bovine, mouse, rat, zebrafish and chicken) (Figure 1B and 1C, respectively).
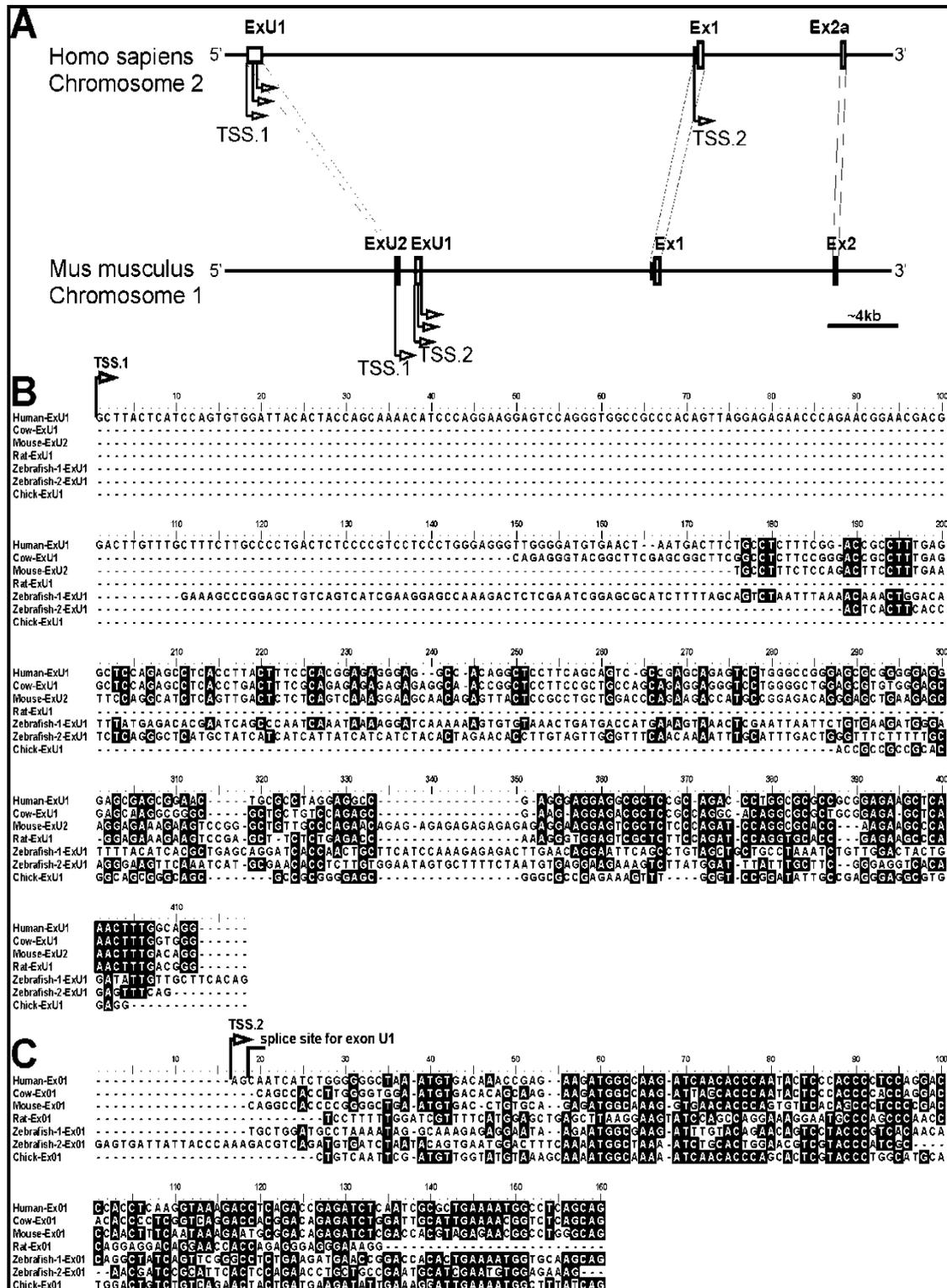
Figure 1: Interspecies comparison of the human CNGA3 gene. (A) Schematic comparison of the human and mouse CNGA3 loci which integrates data from EST/cDNA and 5'-RLM-RACE experiments (TSS shown on arrowheads). The dotted lines show homology among related exons in both species. Notice that mouse has an extra untranslated exon (Ex1) compared to human. (B) Interspecies alignment of the first untranslated exon (ExU1) from known cDNAs and 5'-RLM-RACE experiments. (C) Interspecies alignment of the first translated exon (Ex1) from known mRNA's and 5'-RLM-RACE experiments.

In *silico* analysis of the 5'-flanking regions of the human *CNGA3* gene

To identify the promoter of the human *CNGA3* gene we first performed several *in silico* approaches. CpG-rich regions were searched via iDOUBT suite exploiting cpgIslandExt table at UCSC. Only one CpG-rich region was detected immediately downstream of exon U1, comprising a CpG island of approximately 800 bp located at the beginning of the first intron (Figure 2).

Promoter predictions were also performed on the human *CNGA3* gene by using the Genomatix software programs. El Dorado (www.genomatix.de/cgi-bin/eldorado/) predicted promoters only upstream of exon 1 and upstream of exon 2 (on a region comprising TSS.3 from retinoblastoma transcript AK131300). Promoter Inspector predicted eukaryotic Pol II promoter regions only in intron 1, downstream of exon U1 (Figure 2). By using MatInspector Professional program no main core promoter elements like TATA, CAAT box or Inr sequences upstream of the exons U1 and 1 were predicted (data not shown).
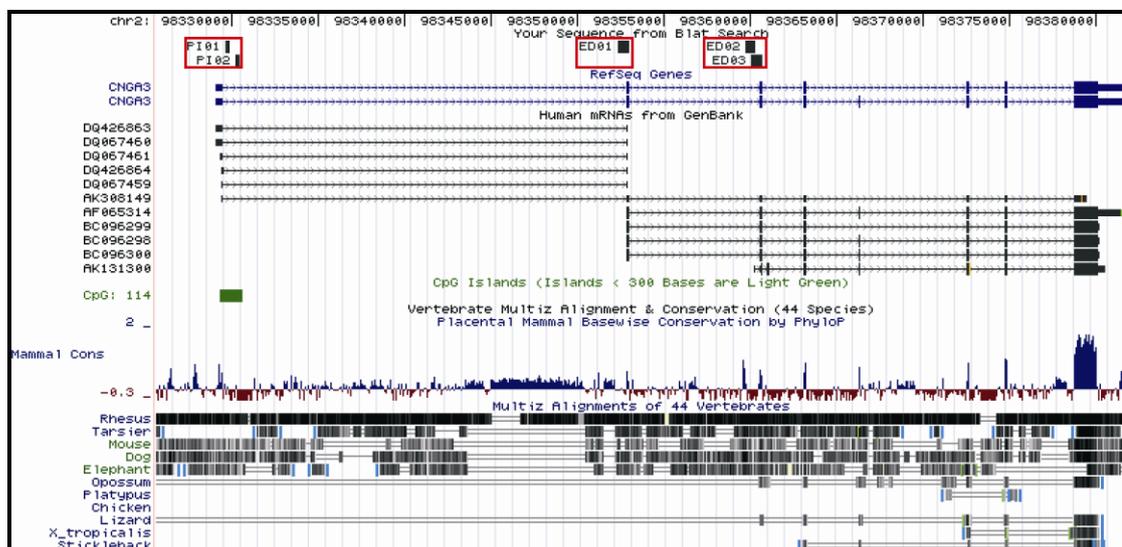
Figure 2: In silico analysis of CNGA3 promoter Schematic representation of the human CNGA3 on the UCSC Genome Browser (blue lines). The scheme integrates data from Genomatix ElDorado (ED) and Promoter Inspector (PI) promoter predictions (framed in red squares). CpG islands are represented on a green box downstream of exon U1. The browser integrates data from other cDNAs from GenBank (black lines).

In vitro characterization of the 5'-flanking regions of the human *CNGA3*

First, luciferase promoter assays were performed in human and mouse cell lines to localize the promoter region of *CNGA3*. Since RLM-RACE experiments identified TSSs in exons U1 and 1, and *in silico* analyses predicted putative promoter regions upstream of exons 1 and 2 together with a promoter region downstream of exon U1, a series of luciferase reporter constructs were constructed by cloning these *CNGA3* sequences into the pGL3-enhancer vector: a 650 and a 860 bp fragments upstream of the human exon U1 (inserts ExU1(-628/+23) and ExU1(-628/+233), in relation to the TSS.1), two regions of 508 and 556 bp upstream of exon 1 (inserts Ex1(-508/-3) and Ex1(-508/+49) in relation to TSS.2) and a 1028 bp sequence upstream of exon 2 (insert Ex2(+6763/+7791) in relation to TSS.2; Figure 3A). Subsequently these five reporter gene constructs were transiently transfected into five different human and mouse cell lines namely ARPE-19, Y79, WERI-Rb1, 661W and HEK-293. The transfection efficiency was normalized against the *Renilla* luciferase activity and standardized to the value of the positive control pGL3-control (Figure 3B). The highest and only significant promoter activity was observed for fragments upstream of exon 1 in the 661W and HEK-293 cells with a six-fold and twenty four-fold increase, respectively, above the promoterless control. To this result followed the Y79 and WERI-Rb1 cells with about five-fold and three-fold increase, respectively. The ARPE-19 cell line exhibited non-significant lower promoter activities (Figure 3B).

In contrast to the regions upstream of TSS.2, the putative promoter inserts upstream of exon U1 or exon 2 did not stimulate significant luciferase activity in any of the five cell lines (Figure 3A). Thus, these regions were not studied further in the present work.
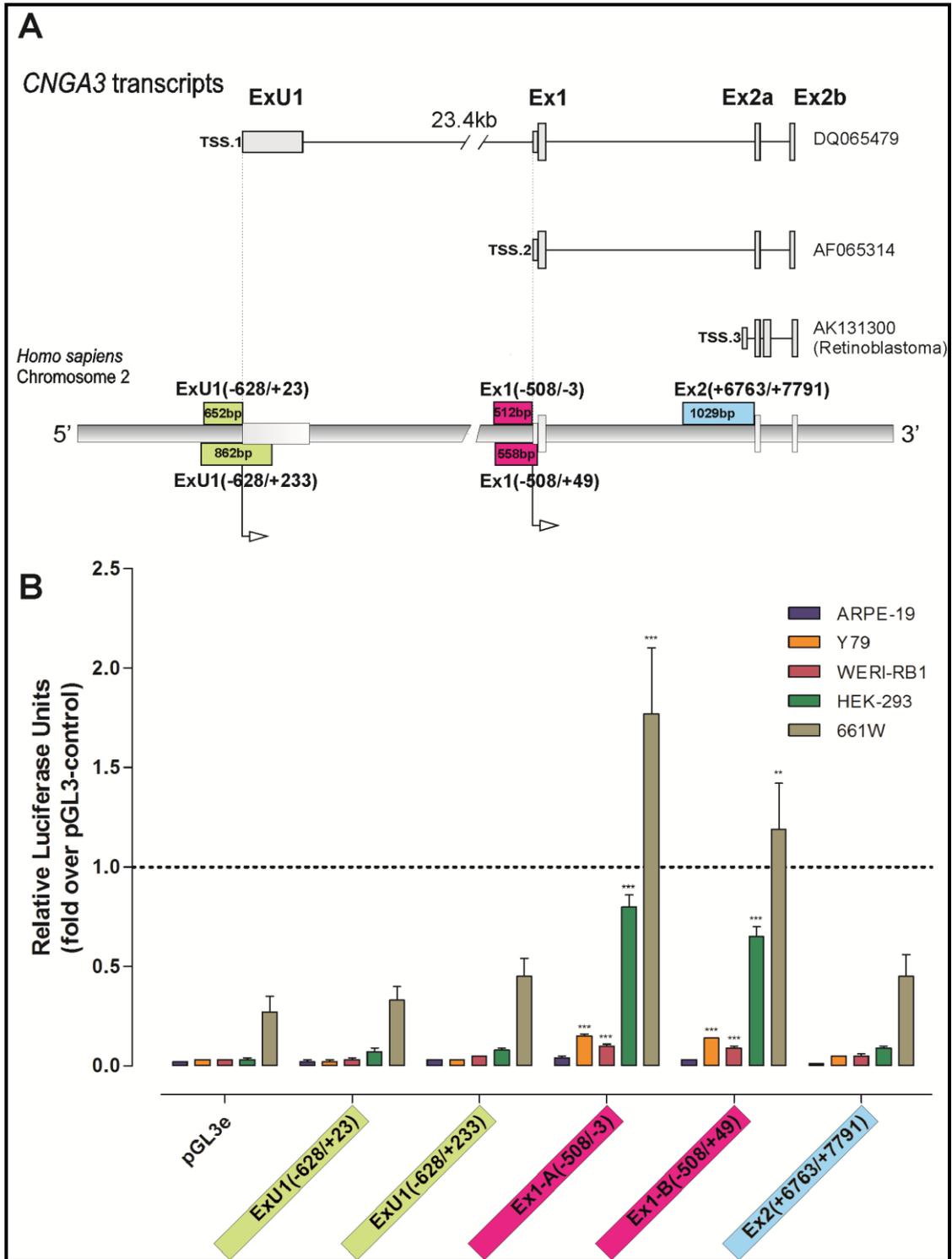
***Figure 3: Transcriptional activity of the human CNGA3 promoter in vitro**. (A) Scaled diagram of the 5'-end of the human CNGA3 locus showing the locations of the first three exons and two introns which integrates data from cDNAs and 5'-RLM-RACE experiments (verified TSSs are shown on arrowheads). The color boxes show the name and size of each corresponding insert that was tested to drive reporter gene expression by cloning them into pGL3-ehnhancer. (B) Five cell lines were cotransfected with the generated reporter gene constructs and the pRL-TK reporter*

control plasmid (ARPE-19 (n=2), Y79 (n=2), WERI-RB1 (n=3), HEK-293 (n=5) and 661W (n=5) and cultured for 48 hours. Luciferase activity was normalized to Renilla luciferase activity in relative luciferase units as a measure of transfection efficiency. Experiments were done in triplicates. The luciferase activities are plotted as fold increases over that of the pGL3-control construct which was set to 1 (dotted line). Statistical analysis was performed using One-Way ANOVA with Dunnett's Multiple Comparison Test, comparing all columns vs. pGL3-enhancer vector (pGL3e). Significance was defined as p<0.01 (**), p<0.001 (***) and p>0.05 was considered non-significant. Only significant results have been marked in the figure. Values are presented as means ± the standard error of the mean (SEM).

Deletion studies of the human *CNGA3* promoter

To determine the basal promoter upstream of exon 1/TSS.2 and to characterize *cis*-acting sequences that regulate the *CNGA3* transcription we generated a series of successive 5´-deletions and expansions of the Ex1(-508/+49) construct (Figure 4). These constructs were transfected into Y79, WERI-Rb1, HEK-293 and mouse cone-like 661W as described before. The result of the luciferase assays showed a similar pattern of luciferase activity of the different inserts tested in all four cell lines (Figure 4), with the highest promoter activity in 661W and HEK-293 cells followed by a weak luciferase activity in Y79 and WERI-RB1 cells.

Deletion of the segments between 996 and 372 bp yielded an increase in luciferase activity. Subsequent deletion of sequences between 292 and 131 bp showed strong decrease of activity (Figure 4), suggesting the presence of an activator. However, a short segment of 34 bp upstream of TSS.2 was still sufficient to induce luciferase expression as strong as the fragment of 372 bp, and thus can be considered as the minimal promoter region of TTS.2.

Interestingly, when we deleted the 5'-transcribed region of the *CNGA3* gene from the construct Ex1(-508/+49) [resulting in the construct Ex1(-508/-3)] we observed an increase of the luciferase activity when compared to its original Ex1(-508/+49) insert (Figure 3 and 4).
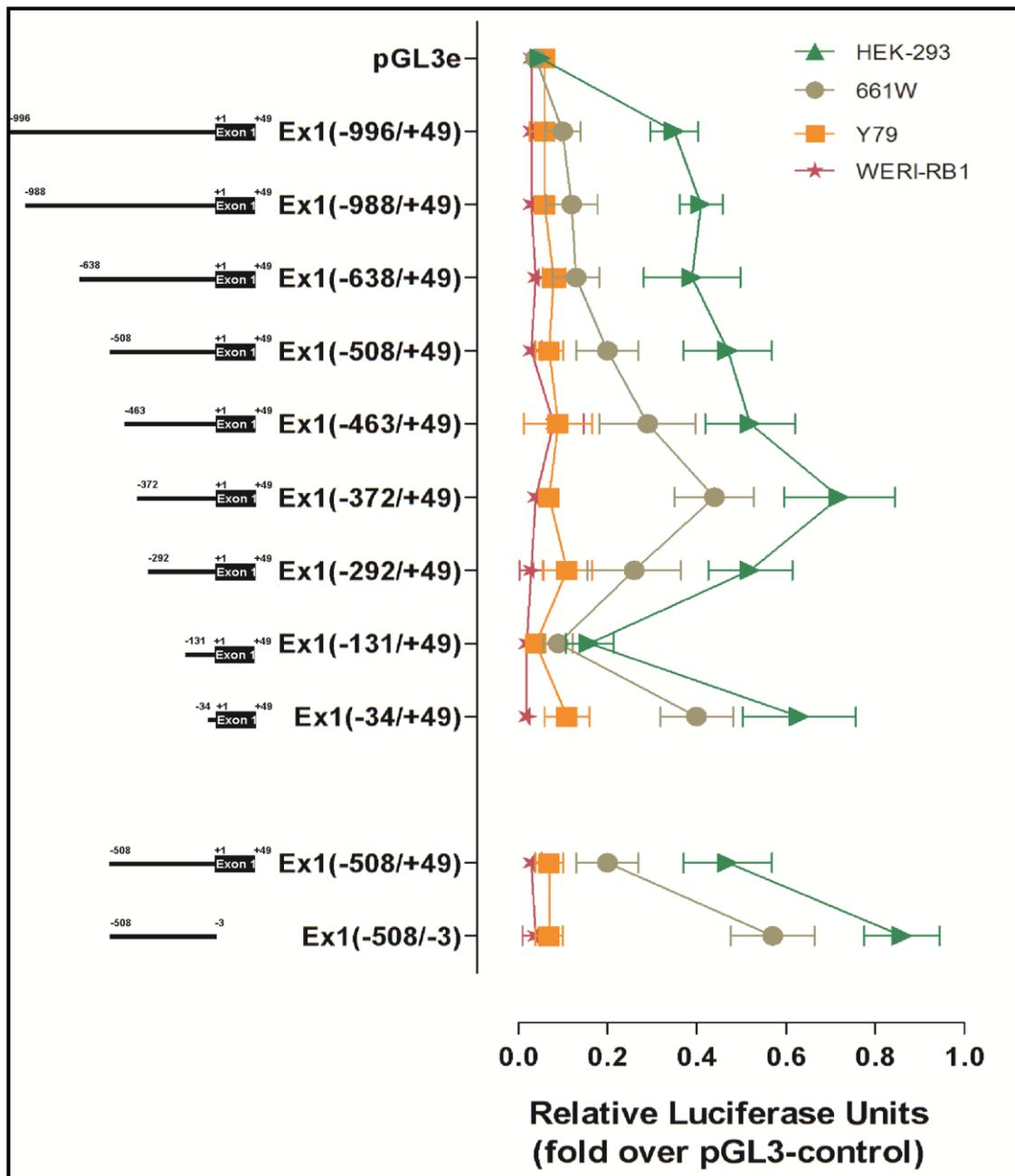
Figure 4: Deletion analysis of the CNGA3 promoter region in various cell lines. Fragments of different sizes of the upstream region of exon 1 of the human CNGA3 gene were cloned into pGL3-enhancer. The 5'-position of the promoter fragments relative to TSS.2 (+1) is indicated, whereas the 3'-position was always +49 excepting for the insert Ex1(-508/-3) whose 3'-position was -3. The Y79, WERI-Rb1, HEK-293 and 661W cell lines were cotransfected with the indicated reporter gene constructs and a pRL-TK reporter control plasmid and assayed 48 h after transfection. Luciferase activity was normalized to Renilla luciferase activity in relative luciferase units as a measure of transfection efficiency. The luciferase activities are plotted as fold increases over that of the pGL3-control construct which was set to 1. Values are presented as means ± the standard error of the mean (SEM) of 2 (WERI-RB1 and Y79) and 3 (HEK-239 and 661W) independent experiment in triplicates.

Mutational analysis of putative transcription factor binding sites in the *CNGA3* promoter region

Our prediction approach relied on profile scan for putative transcription factor binding sites using the MatInspector Professional program (http://www.genomatix.de; Quandt et al., 1995) considering core promoter elements and general as well as eye-specific transcription factors with a matrix-specific optimized threshold as cut-off. This profile-based scan allowed us to highlight 84 putative binding sites from vertebrates and more specifically 28 eye specific transcription factor binding sites in the 508 bp upstream of exon 1 (data not shown).

Prediction filtering was done using multiple sequence alignment of 500 to 1000 bp upstream of the different TSS from human, dog, mouse and rat using the MultiZ program (http://genome.ucsc.edu/; Blanchette et al., 2004). Prediction and filtering of the 508 bp upstream of TSS.2 (Figure 5A) revealed nine putative transcription factor binding sites for the following eye specific transcription factors: pituitary homeobox 1 (PTX1), Brn-2, POU-III protein class (BRN2), homeodomain protein Hoxb-9 (HOXB9), HMG box-containing protein 1 (HBP1), homeobox transcription factor Nanog (NANOG), sex determining region Y gene product (SOX/SRY), bipartite binding site of VDR/RXR heterodimers-DR3 sites (VDR/RXR) and orphan nuclear receptor rev-erb alpha/homodimer DR2 binding site (REV-ERBα). Also, the non tissue specific transcription factors: MYC-MAX binding sites (MYC/MAX), nuclear factor-like 2 (NFE2L2) and Yin Yang 1 activator site (YY1).

Mutational deletions of these putative transcription factor binding sites were generated by PCR in four of the reporter gene constructs: Ex1(-508/+49), Ex1(-508/-3), Ex1(-372/+49) and Ex1(-34/+49) (the mutant constructs were sequenced to ensure that undesirable mutations generated by PCR were absent). The mutant constructs were transiently transfected into the cell lines previously shown to elicit promoter activities (HEK-293 and 661W). Luciferase assays were performed in cells as described before to analyze potential effects of these mutational deletions.

The promoter activity profiles of the 5'-deletion mutants were found to be variable in these cell lines. Compared to the appropriate non-mutant construct deletion mutants for the transcription factors: BRN2/POU-III, VDR/RXR and YY1 elicit most significant and constant results among the different inserts showed increasing of luciferase activity (Figure 5 B-E). Decrease of luciferase activity in HBP1 and REV-ERBα mutants were also evident as consistent results (Figure 5 B-E).
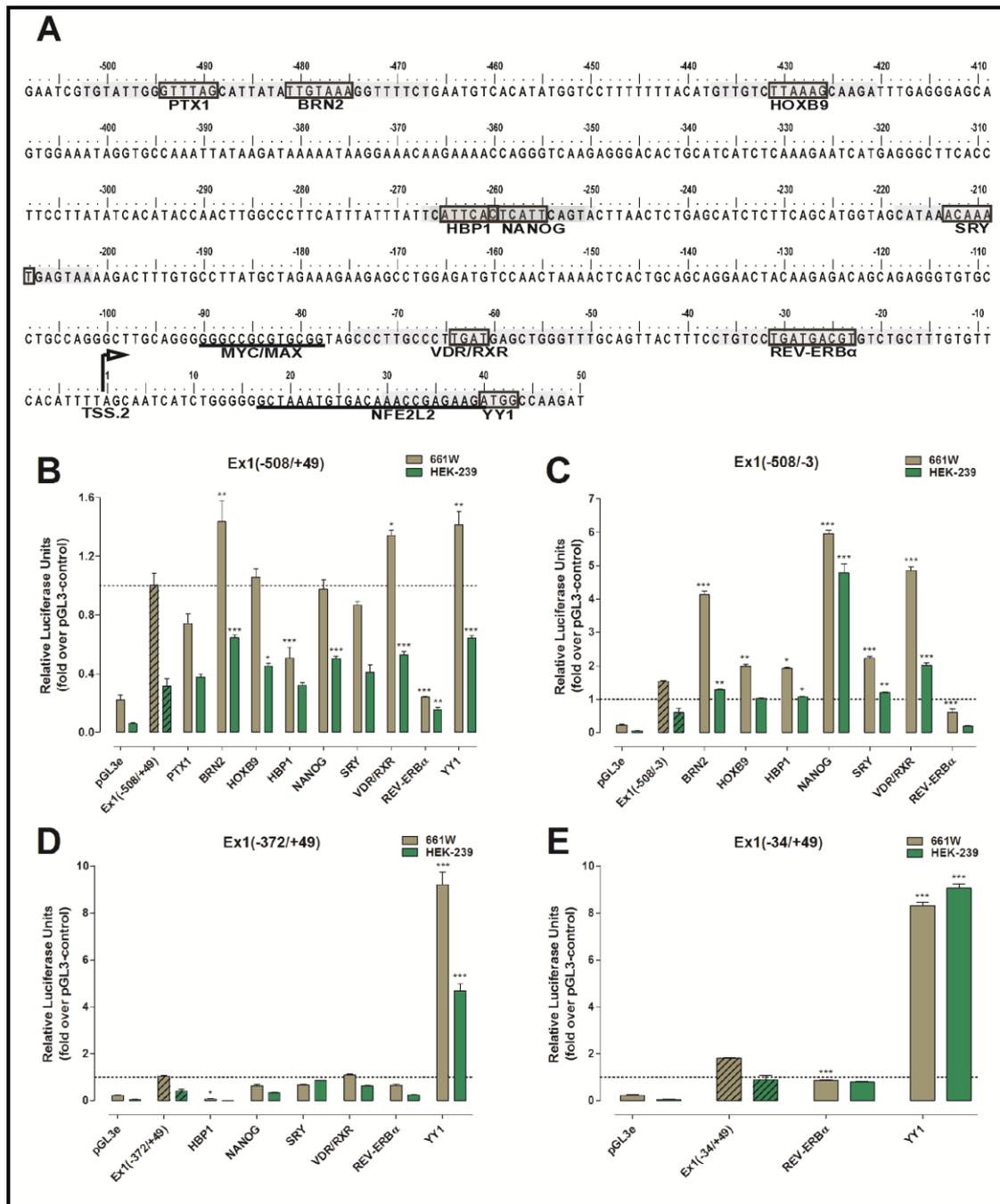
Figure 5: Site-directed deletion analysis of putative transcription factors binding sites in the CNGA3 promoter region. *(A) The CNGA3exon 1 5'-flanking region was searched for putative transcription factor binding sites using the MatInspector software from Genomatix BiblioSphere and compared by phylogenetic footprinting using MultiZ. The sequences for nine putative transcription factor binding sites are dashed in grey boxes and their respective site-directed deleted core region framed in black boxes. The deletions were done independently in four of the previous reporter constructs: Ex1(-508/+49), Ex1(-508/-3), Ex1(-372/+49) and Ex1(-34/+49). Transcription start site 2 (TSS.2) identified by 5'-RLM-RACE is marked by an arrow and is numbered as +1. (B-E) Two cell lines (HEK-293 and 661W) were cotransfected with the indicated site-directed deleted reporter-gene constructs and the pRL-TK*

reporter control plasmid and assayed 48 h after transfection. Luciferase activity was normalized to Renilla luciferase activity in relative luciferase units as a measure of transfection efficiency. Experiments were repeated two times, in triplicates. The luciferase activities are plotted as fold increases over that of the pGL3-control construct which was set to 1 (dotted line). Statistical analysis was performed using One-Way ANOVA with Dunnett's Multiple Comparison Test, comparing for each respective cell line all columns vs. the native-deletion free CNGA3 promoter construct (dashed columns). Significance was defined as p<0.05 (*), p<0.01 (**), p<0.001 (***) and p>0.05 was considered non-significant. Only significant results have been marked in the figure. Values are presented as means ± the standard error of the mean (SEM).

Effect of expression of transcription factors on the human *CNGA3* promoter region

We studied the effect of overexpression of some putative transcription factors on the Ex1(-372/+49) construct in the HEK-293 and 661W cells. The indicated construct was independently cotransfected with one of several expression plasmids for REV-ERBα, MAX, NFE2L2 and YY1 (cloned in pcDNA3.1) and the pRL-TK *Renilla* luciferase normalization vector. Luciferase assays were performed as described before to analyze effects of transcription factor overexpression. As shown in figure 6, only overexpression of YY1 caused a significant and constantly decreases of *CNGA3*-promoter luciferase reporter levels in both cell lines when comparing to the transfection without cotransfected with any pcDNA3.1- transcription factor plasmid.

Tissue expression pattern of the human *CNGA3* gene

In order to localize the different populations of *CNGA3* transcripts in the human retina, we performed *in situ* hybridization with two different sense and antisense probes. The probes where designed in a way to stain the human retinal tissue in a subtractive manner. Probe 1 comprised almost the whole length of exon U1 (probe-ExU1, 360 bp) while probe 2 contains exon 1 to 3 (probe-Ex1-3, 370 bp) (Figure 7A). Antisense probe-Ex1-3 detects both *CNGA3* transcripts which start transcription at TSS.1 and TSS.2, while the antisense probe-ExU1 labels only transcripts derived from both TSS.1 and TSS.2. Figures 7B to E show the results of the human *in situ* hybridization. Probe-ExU1 stained cells across the entire Outer Nuclear Layer (ONL) and Inner Nuclear Layer (INL, Figure 7B) while probe-Ex1-3 stained in addition cells of the Ganglion Cell Layer (GCL) and strongly the inner segments of the photoreceptor layer (Figure 7C). Sense control probes gave no signal when processed in the same way as the antisense reactions (Figure 7D and E). If we subtract the staining signal of probe-ExU1 from probe-Ex1-3, the data suggests a predominant expression of the TSS.2 transcript in the photoreceptor layer as well as in ganglion cells. Moreover, the expression of the transcript starting at TSS.1 is weak in ONL and INL layers.
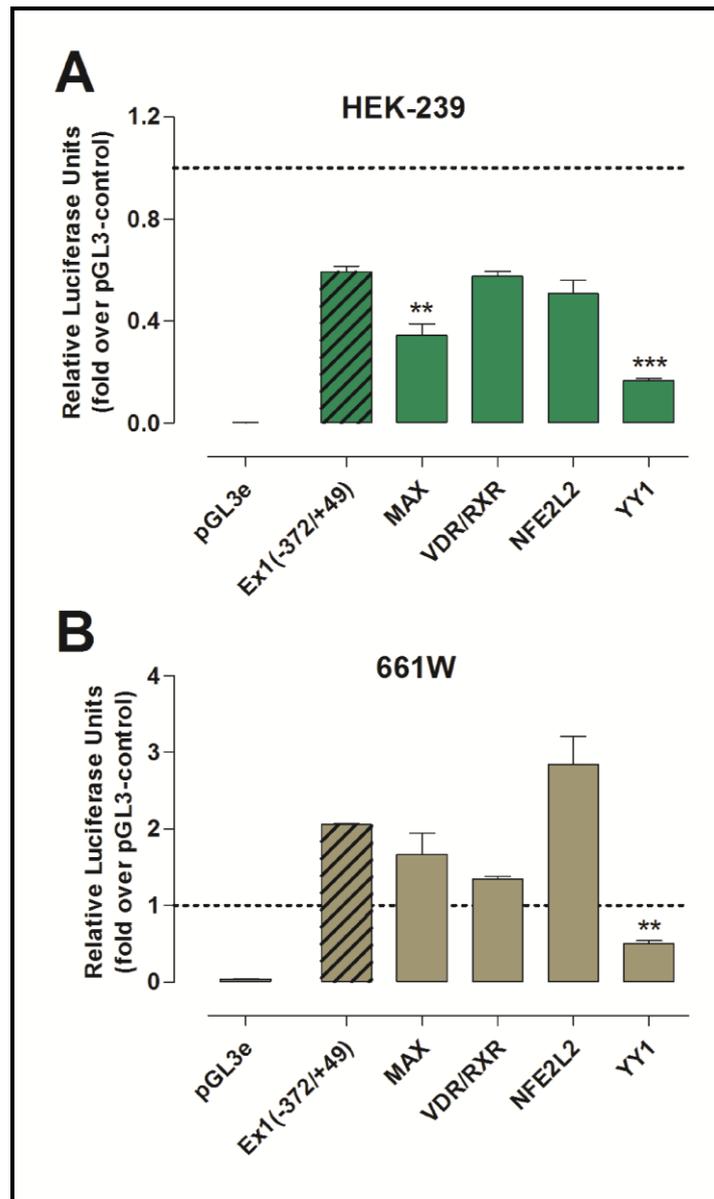
Figure 6: Transactivation assay on the CNGA3 on the promoter. *HEK-293 (A) and 661W cell lines (B) an expression plasmid were independently cotransfected with the indicated reporter gene construct Ex1(-372/+49), an expression plasmid containing the ORF for the transcription factors MAX, RXRγ, NFE2L2 and YY1 and a pRL-TK reporter control plasmid. Luciferase activity was normalized to Renilla luciferase activity in relative luciferase units as a measure of transfection efficiency. Two independent experiments were done in triplicates. The luciferase activities are plotted as fold increases over that of the pGL3-control construct which was set to 1 (dotted line). Statistical analysis was performed using One-Way ANOVA with Dunnett's Multiple Comparison Test, comparing for the respective cell line, all columns vs. native construct cotransfected with only the empty pcDNA3.1 vector. Significance was defined as p<0.01 (**), p<0.001 (***) and p>0.05 was considered non significant. Only significant results have been marked in the figure. Values are presented as means ± the standard error of the mean (SEM).*
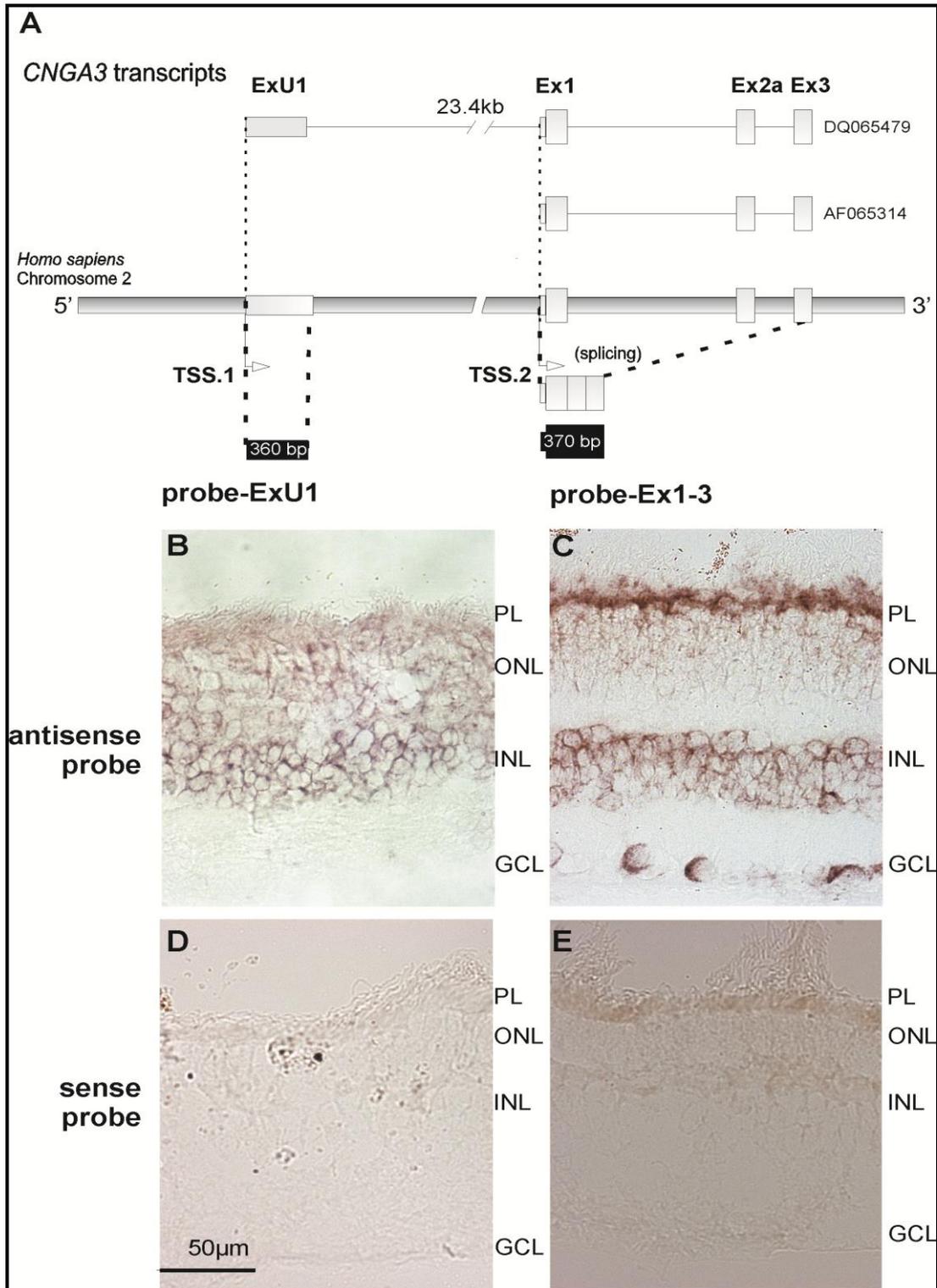
Figure 7: In situ hybridization of CNGA3 transcripts in human retina. *(A) Scaled diagram of the 5'-end of human CNGA3 gene containing the two main TSS (TSS.1 and TSS.2). The black boxes indicate the scaled sizes of the in situ hybridization probes. (B-C) In situ hybridization staining with antisense probes of the exon ExU1 (B) and Ex1 (C), respectively. (D-E) In situ hybridization staining with sense control-probes of the exon ExU1 (D) and Ex1 (E), respectively. PL, photoreceptor layer; ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer.*

Discussion

CNGA3 has two alternative TSS in retinal tissue

CNGA3 encodes the α-subunit of the cGMP gated channel that is expressed in cone photoreceptors and plays an important role in the phototransduction cascade (Biel et al., 1999). *CNGA3* has attracted considerable interest as a gene for achromatopsia (Kohl et al., 1998). Yet nothing is known about its regulation on the level of gene expression. In this study we have investigated the 5'complexity of CNGA3 transcription regulation. We were able to define the presence of at least two TSS for *CNGA3* as demonstrated by RLM-RACE in retinal tissue (Figure 1A). These identified transcripts confirmed previously published *CNGA3* cDNAs (DQ065479 and AF065314) (Kohl et al., 1998; Roni et al., 2007). According to the working definition used by Carninci et al., 2006, our results of the TSS regions of human *CNGA3* encompasses two non-overlapping clusters. While TSS.1 (with 3 different TSS) displays a broad distribution of start sites, TSS.2 displays a single start point. These two clusters initiate two alternative transcripts that differ in their untranslated sequences, but are identical within their coding regions. Sequence comparisons that showed a conserved splice domain sequence indicate that exon U1 is functional in different mammalian species (Roni et al., 2007). Nevertheless our findings do not rule out the existence of other TSSs in *CNGA3* in the human retina (as example TSS.3 indicated by the EST AK131300 from retinoblastoma samples).

Human *CNGA3* promoter is located upstream of TSS.2

Recent evidence suggests that 30–50% of the human and about one half of the mouse genes have multiple alternative promoters that may differ in location in the genome by thousands of kilobases (Baek et al., 2007; Kimura et al., 2006; Takeda et al., 2007). We investigated if the expression of both CNGA3 transcript variants might be controlled by different promoter sequences. Therefore we cloned different fragments located immediately upstream of the transcription start sites (TSS.1, TSS.2 and TSS.3) into a luciferase reporter vector and tested various retinal and non-retinal cell lines (ARPE-19, Y79, WERI-Rb1, HEK-293 and 661W) for reporter gene activity.

Our *in vitro* results localized an active promoter immediately upstream of TSS.2/exon 1. In contrast, putative promoter regions upstream of exon U1 or exon 2 did not stimulate significant luciferase activity in any of the five mentioned cell lines. Since we were not able to find substantial homology between human and mouse putative promoter regions (data not shown), this shift of the human promoter from TSS.1 to a more 3' directed TSS.2 of *CNGA3* gene could suggest a species-specific change in gene regulation. This TSS.2 is closer to translation initiation, leaving just a short 5'UTR that normally plays important roles in differential regulation of translation efficiency (Davuluri et al., 2000). Concordant with this idea, the length of the 5'UTR of transcripts containing TSS.1 could influence mRNA stability, which is another mechanism of *CNGA3* posttranscriptional regulation that under normal conditions may

impede efficient translation (Pickering and Willis, 2005). If such is the scenario, then this repressive regulation might be released by splicing out exon U1 from the endogenous *CNGA3* transcripts.

The presence of the alternative transcripts using different TSS was tested in a differential manner by *in situ* hybridization (ISH; Figure 7). ISH patterns suggest that TSS.2-transcript is strongly expressed in the photoreceptor inner segments and in ganglion cells, whereas TSS.1-transcripts are weakly expressed along the ONL and INL. These results support the idea that the promoter activating the TSS.2-transcript act more in a cone-photoreceptor and ganglion cells-specific way, while transcripts starting at TSS.1 don't have such a specificity of expression within the photoreceptor layer and are shown to be expressed in a weak and broad way along the ONL and INL or are simply strongly expressed in alternative tissues. Such independent regulatory events could give rise to different tissue expression patterns.

Since we could only detect a CpG island immediately downstream from exon U1, we speculate if this region might work as an enhancer element for the promoter of any of both *CNGA3* transcripts. Some examples of such CpG enhancer elements were described previously (Iguchi-Ariga and Schaffner, 1989; Unoki and Nakamura, 2003; Williams et al., 2005). Unfortunately we were not able to successfully clone this region due to the presence of a long GC rich tract on the sequence.

Cell-specific expression of the *CNGA3* promoter

The luciferase gene reporter assay showed the highest promoter activity for the gene fragment located upstream of Ex1 of CNGA3 expressed best in 661W and HEK-293 cells, followed by a weak luciferase activity in Y79 and WERI-Rb1 cells and no significant activity on ARPE-19 cells. These results suggest that the sequence upstream of exon1 of *CNGA3* gene is a robust functional promoter with restricted cell type specificity. It exhibits strong promoter activities in 661W and HEK-293, presumably because these responsive cell lines contain the appropriate trans-acting factors essential for its transcription.

The previously discussed use of alternative promoters could allow for a more complex regulation of gene expression, arguing for the possibility that the responsive promoter bring about cell-specific transcriptional regulation (Landry et al., 2003). In accordance with this observation, the CAGE tag viewer results for *CNGA3* revealed no tissue-specificity in the composition of the different tag clusters (data not shown).

Transcription factor repressive sequences are found in the *CNGA3* promoter

The lower activity of the longer constructs as compared to the higher activity of the shorter ones in the deletion analysis experiment is due to the fact that longer constructs could accumulate repressive signals and therefore have lower activities on the gene reporter assays. As shown in Figures 4 and 5, the only results that can be correlated from both progressive-deletion and mutational-deletion experiments suggest repressive properties for both BRN2/POU and YY1. Since the deleted YY1 core site

lays on the initiation codon sequence of *CNGA3* gene, we cloned this codon on the same reading frame as the firefly luciferase reporter gene avoiding misinterpretation of data caused by competition of the initiation codons of these two genes. Moreover, we could verify the opposite result for YY1 when overexpressing the putative repressor, decreasing the activity of the reporter Ex1(-372/+49). YY1 transcription factors are well known to work as an activator and repressor of transcription in many core promoter elements (Austen et al., 1997; Shi et al., 1981). If YY1 could be acting as a repressor of transcription, then this could raise the possibility of competition of alternative promoters from TSS.1 and TSS.2 for initiation of transcription from exon U1 when exon 1 promoter is in a repressed state by YY1.

Inconsistent results for other transcription factors cannot be explained by this approach since we cannot rule out the presence of other regulatory sequences working on the proximities which weren't predicted by the putative transcription factor binding site analysis from MatInspector.

In summary, we have verified two TSSs for the human retinal *CNGA3* and identified that the transcription of the human *CNGA3* gene is controlled by a promoter located immediately upstream of TSS.2-Ex1. Furthermore, we identified the minimal promoter needed for transcription. Finally, we have also defined a region in the *CNGA3* promoter potentially regulated by YY1, which most probably involves *CNGA3* transcription repression which is consistent with the presumed function of this transcription factor. These findings reveal a previously unknown level of complexity in the regulation of the expression of the *CNGA3* gene, and provide the basis for further studies of its transcriptional regulation. Moreover, this promoter could be useful as a tool on gene therapy trials, where cone photoreceptors are the final target for gene delivery.

SUMMARY OF RESULTS

The first chapter of the thesis describes the mapping of TSSs of genes expressed in human retina. Genes subject of study were selected on the basis of their physiological or developmental role in this tissue. *In silico* analysis of ESTs and known algorithm predictions of exons were employed together with experimental validation via Cap-finder RACE of their respective cDNA. This thesis reports the mapping of TSSs of 54 retina expressed genes, retrieving new sequences for 41 of them, some of which contain un-annotated exons. Some results were verified by primer extension experiments and by inter species analysis of splicing regions. Results can be grouped into five categories, compared to the RefSeq; (i) TSS located in new first exons, (ii) splicing variation of the second exon, (iii) extension of the annotated first exon, (iv) shortening of the annotated first exon, (v) confirmation of previously annotated TSS.

The second chapter of the thesis have verified two TSSs for the human retinal *CNGA3* and identified that the transcription of the human *CNGA3* gene is controlled by a promoter located immediately upstream of TSS.2-Ex1. Furthermore, it also identified the minimal promoter needed for transcription and also defined a region in the *CNGA3* promoter potentially regulated by YY1, which most probably involves *CNGA3* transcription repression which is consistent with the presumed function of this transcription factor.

GENERAL DISCUSSION

The results of this thesis give correct information about the complete 5' end of transcripts; this data will be useful to locate the respective core and proximal promoter elements. Cap-finder RACE was chosen to map the TSS of retinal expressed genes because this technique is selective for complete transcripts (Dike et al., 2004; Suzuki et al., 2004). With this technique it was found that previous gene annotation was missing part of the first exon or even complete exons at the 5' end. This experimental determination of TSSs shows that the current gene annotation was in most cases obtained from data sources that are not strictly selective for the complete transcribed form, and need to be updated. This procedure led the discover of several transcript isoforms that were un-annotated and locating retina specific TSSs. Therefore new isoforms give a more complete picture of alternative start sites use in retinal genes. New reported TSSs contribute to the identification of regulatory elements active in tissue specific gene regulation; and this computational approaches will only be effective after experimental validation.

Nevertheless, throughout the work of this thesis project, it was understood the importance of placing 5' RACE primers in a way that the experimental design allows multiple TSSs finding on contiguous exons. The *CNGA3* gene was the best example of this importance: after the results of the first chapter, further placement of *CNGA3* 5' RACE inner primers in human in experiments on the second chapter evidenced the presence of a second TSS, named TSS.2. This findings evidenced a major level of

complexity on the transcription of human *CNGA3* suggesting the existence of alternative promoters controlling alternative transcription initiation sites. Furthermore, cell specific expression could explain the existence of those alternative *CNGA3* transcripts where repressive transcription factors (as for YY1) could be limiting the expression of this gene to defined regions of the retina.

GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

In conclusion, the results of this thesis highlight the importance of applying a tissue specific approach with a systematic program of Cap-finder RACE using the known gene structures as a starting point and/or gene predictions to complete the existing gene annotation. The new TSSs and transcribed sequences provide crucial information for further exploration of the promoter and other *cis*-regulatory sequences at the 5'-end of the gene, and in particular for the study of those elements that are functionally active in human retina.

We have identified two TSSs for human retinal *CNGA3* and that transcription of the human *CNGA3* gene is controlled by a promoter located immediately upstream of TSS.2-Ex1. Furthermore, we identified the minimal promoter needed for transcription. Finally, we have also defined a region in the *CNGA3* promoter regulated by YY1. These findings reveal a previously unknown level of complexity in the regulation of the expression of the *CNGA3* gene, and provide the basis for further studies of its transcriptional regulation.

It will be a important task for the future to analyze the nature of transcription factors binding directly to the *CNGA3* promoter region *in vivo* to elucidate its regulation in detail, especially in view of the fact that *CNGA3* expression, which was assumed to be exclusive for cone photoreceptors, has also been described in other cell types. The characterization of transcriptional regulators influencing *CNGA3* expression might thus help elucidating the basis for tissue-specific expression of *CNGA3*. Moreover, this promoter could be useful as a tool on gene therapy trials, where cone photoreceptors are the final target for gene delivery.

REFERENCES

Adler, R., 1993. Plasticity and differentiation of retinal precursor cells. Int. Rev. Cytol. 149, 145-189.

Ahmad, I., Barnstable, C. J., 1993. Differential Laminar Expression of Particulate and Soluble Guanylate Cyclase Genes in Rat Retina. Experimental Eye Research. 56, 51-62.

Ahmad, I., et al., 1990. Developmental and tissue-specific expression of the rod photoreceptor cGMP-gated ion channel gene. Biochemical and Biophysical Research Communications. 173, 463-470.

Antequera, F., Bird, A., 1993. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A. 90, 11995-9.

Austen, M., et al., 1997. Characterization of the Transcriptional Regulator YY1. The Bipartite Transactivation Domain is Independent of Interaction with the TATA Box- Binding Protein, Transcription Factor IIB, TAFII55, or cAMP- Responsive Element- Binding Protein (CBP)-Binding Protein. J. Biol. Chem. 272, 1709-1717.

Babu, S., et al., 2006. Conserved cis-elements in the Xenopus red opsin promoter necessary for cone-specific expression. FEBS Lett. 580, 1479-84.

Baek, D., et al., 2007. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome Research. 17, 145-155.

Barragan, I., et al., 2005. Mutation screening of three candidate genes, ELOVL5, SMAP1 and GLULD1 in autosomal recessive retinitis pigmentosa. Int J Mol Med. 16, 1163-7.

Biel, M., et al., 1999. Selective loss of cone function in mice lacking the cyclic nucleotide-gated channel CNG3. Proceedings of the National Academy of Sciences of the United States of America. 96, 7553-7557.

Biel M, Z. X., Ludwig A, Sautter A, Hofmann F., 1999. Structure and function of cyclic nucleotide-gated channels. Rev Physiol Biochem Pharmacol. 135, 151-171.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6-21.

Blackshaw, S., et al., 2001. Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. Cell. 107, 579-89.

Blackshaw, S., et al., 2004. Genomic analysis of mouse retinal development. PLoS Biol. 2, E247.

Blake, M. C., et al., 1990. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. Mol Cell Biol. 10, 6632-41.

Blanchette, M., et al., 2004. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. Genome Research. 14, 708-715.

Bortoluzzi, S., et al., 2005. A multistep bioinformatic approach detects putative regulatory elements in gene promoters. BMC Bioinformatics. 6, 121.

Brandeis, M., et al., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature. 371, 435-8.

Burke, T. W., Kadonaga, J. T., 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev. 10**,** 711-24.

Burke, T. W., et al., 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb Symp Quant Biol. 63**,** 75-82.

Butler, J. E., Kadonaga, J. T., 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev. 16**,** 2583-92.

Carcamo, J., et al., 1991. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc Natl Acad Sci U S A. 88**,** 8052-6.

Carninci, P., 2006. Tagging mammalian transcription complexity. Trends Genet. 22**,** 501-10.

Carninci, P., et al., 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet. 38**,** 626-635.

Cassar, S. C., et al., 2004. Tissue specific expression of alternative splice forms of human cyclic nucleotide gated channel subunit CNGA3. Mol Vis. 10**,** 808-13.

Chen, F., et al., 1999. Retina-specific nuclear receptor: A potential regulator of cellular retinaldehyde-binding protein expressed in retinal pigment epithelium and Muller glial cells. Proc Natl Acad Sci U S A. 96**,** 15149-54.

Cheng, J., et al., 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science. 308**,** 1149-54.

Chowers, I., et al., 2003. Identification of novel genes preferentially expressed in the retina using a custom human retina cDNA microarray. Invest Ophthalmol Vis Sci. 44**,** 3732-41.

Cooper, S. J., et al., 2006. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res. 16**,** 1-10.

Davuluri, R. V., et al., 2001. Computational identification of promoters and first exons in the human genome. Nat Genet. 29**,** 412-7.

Davuluri, R. V., et al., 2000. CART Classification of Human 5â€² UTR Sequences. Genome Research. 10**,** 1807-1816.

Deng, W., Roberts, S. G., 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev. 19**,** 2418-23.

Deng, W., Roberts, S. G., 2007. TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma. 116**,** 417-29.

Dike, S., et al., 2004. The mouse genome: experimental examination of gene predictions and transcriptional start sites. Genome Res. 14**,** 2424-9.

Dunn, K. C., et al., 1996. ARPE-19, A Human Retinal Pigment Epithelial Cell Line with Differentiated Properties. Experimental Eye Research. 62**,** 155-170.

Eckey, M., et al., 2003. Gene silencing by the thyroid hormone receptor. Mol Cell Endocrinol. 213**,** 13-22.

Finn, J. T., et al., 1996. Cyclic Nucleotide-Gated Ion Channels: An Extended Family With Diverse Functions. Annual Review of Physiology. 58**,** 395-426.

FitzGerald, P. C., et al., 2006. Comparative genomics of Drosophila and human core promoters. Genome Biol. 7**,** R53.

Frith, M. C., et al., 2006. Evolutionary turnover of mammalian transcription start sites. Genome Res. 16, 713-22.

Frith, M. C., et al., 2008. A code for transcription initiation in mammalian genomes. Genome Res. 18, 1-12.

Gegenfurtner, K. R., Kiper, D. C., 2003. COLOR VISION. Annual Review of Neuroscience. 26, 181-206.

Gershenzon, N. I., Ioshikhes, I. P., 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics. 21, 1295-300.

Gershenzon, N. I., et al., 2006. The features of Drosophila core promoters revealed by statistical analysis. BMC Genomics. 7, 161.

Gerstner, A., et al., 2000. Molecular cloning and functional characterization of a new modulatory cyclic nucleotide-gated channel subunit from mouse retina. J Neurosci. 20, 1324-32.

Glushakova, L. G., et al., 2006. Human blue-opsin promoter preferentially targets reporter gene expression to rat s-cone photoreceptors. Invest Ophthalmol Vis Sci. 47, 3505-13.

Goto-Omoto, S., et al., 2006. Compound heterozygous CNGA3 mutations (R436W, L633P) in a Japanese patient with congenital achromatopsia. Visual Neuroscience. 23, 395-402.

Graham, F. L., et al., 1977. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. J Gen Virol. 36, 59-72.

Haeseleer, F., et al., 1998. Molecular characterization of a novel short-chain dehydrogenase/reductase that reduces all-trans-retinal. J Biol Chem. 273, 21790-9.

Haeseleer, F., et al., 2002. Dual-substrate specificity short chain retinol dehydrogenases from the vertebrate retina. J Biol Chem. 277, 45537-46.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol. 11, 394-403.

Halees, A. S., et al., 2003. PromoSer: A large-scale mammalian promoter and transcription start site identification service. Nucleic Acids Res. 31, 3554-9.

Hall, T. A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95-98.

Harpavat, S., Cepko, C. L., 2003. Thyroid hormone and retinal development: an emerging field. Thyroid. 13, 1013-9.

Hoang, E., et al., 1999. Characterization of a novel C-kinesin (KIFC3) abundantly expressed in vertebrate retina and RPE. Exp Eye Res. 69, 57-68.

Hochheimer, A., Tjian, R., 2003. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. 17, 1309-20.

Humphries, M. M., et al., 1997. Retinopathy induced in mice by targeted disruption of the rhodopsin gene. Nat Genet. 15, 216-9.

Hyatt, G. A., Dowling, J. E., 1997. Retinoic acid. A key molecule for eye and photoreceptor development. Invest Ophthalmol Vis Sci. 38, 1471-5.

Iguchi-Ariga, S. M., Schaffner, W., 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes & Development. 3**,** 612-619.

Janecke, A. R., et al., 2004. Mutations in RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy. Nat Genet. 36**,** 850-4.

Jin, V. X., et al., 2006. Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. BMC Bioinformatics. 7**,** 114.

Johan T. den Dunnen, S. E. A., 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. Human Mutation. 15**,** 7-12.

Juven-Gershon, T., et al., 2006. Perspectives on the RNA polymerase II core promoter. Biochem Soc Trans. 34**,** 1047-50.

Juven-Gershon, T., et al., 2008. The RNA polymerase II core promoter - the gateway to transcription. Curr Opin Cell Biol. 20**,** 253-9.

Kadonaga, J. T., 2002. The DPE, a core promoter element for transcription by RNA polymerase II. Exp Mol Med. 34**,** 259-64.

Kaupp, U. B., Koch, K. W., 1992. Role of cGMP and Ca2+ in Vertebrate Photoreceptor Excitation and Adaptation. Annual Review of Physiology. 54**,** 153-176.

Kawamura, S., Tachibanaki, S., 2008. Rod and cone photoreceptors: Molecular basis of the difference in their physiology. Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology. 150**,** 369-377.

Kent, W. J., 2002. BLAT--the BLAST-like alignment tool. Genome Res. 12**,** 656-64.

Kikuchi, T., et al., 1993. The proximal promoter of the mouse arrestin gene directs gene expression in photoreceptor cells and contains an evolutionarily conserved retinal factor-binding site. Mol Cell Biol. 13**,** 4400-8.

Kim, T. H., et al., 2005. A high-resolution map of active promoters in the human genome. Nature. 436**,** 876-80.

Kimura, K., et al., 2006. Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. Genome Research. 16**,** 55-65.

King, D. C., et al., 2005. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res. 15**,** 1051-60.

Kohl, S., et al., 2000. Mutations in the CNGB3 gene encoding the beta-subunit of the cone photoreceptor cGMP-gated channel are responsible for achromatopsia (ACHM3) linked to chromosome 8q21. Hum Mol Genet. 9**,** 2107-16.

Kohl, S., et al., 2002. Mutations in the Cone Photoreceptor G-Protein [alpha]-Subunit Gene GNAT2 in Patients with Achromatopsia. The American Journal of Human Genetics. 71**,** 422-425.

Kohl, S., et al., 1998. Total colourblindness is caused by mutations in the gene encoding the alpha-subunit of the cone photoreceptor cGMP-gated cation channel. Nat Genet. 19**,** 257-9.

Komaromy, A. M., et al., 2008. Targeting gene expression to cones with human cone opsin promoters in recombinant AAV. Gene Ther. 15**,** 1073.

Kutach, A. K., Kadonaga, J. T., 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol. 20, 4754-64.

Lagrange, T., et al., 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev. 12, 34-44.

Lander, E. S., et al., 2001. Initial sequencing and analysis of the human genome. Nature. 409, 860-921.

Landry, J.-R., et al., 2003. Complex controls: the role of alternative promoters in mammalian genomes. Trends in Genetics. 19, 640-648.

Li, Q., et al., 2008. Cone-specific expression using a human red opsin promoter in recombinant AAV. Vision Res. 48, 332-8.

Li, X. F., et al., 2002. Molecular cloning of a fourth member of the potassium-dependent sodium-calcium exchanger gene family, NCKX4. J Biol Chem. 277, 48410-7.

Lim, C. Y., et al., 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 18, 1606-17.

Lord-Grignon, J., et al., 2004. Characterization of new transcripts enriched in the mouse retina and identification of candidate retinal disease genes. Invest Ophthalmol Vis Sci. 45, 3313-9.

Macleod, D., et al., 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282-92.

Maeda, A., et al., 2000. Aberrant expression of photoreceptor-specific calcium-binding protein (recoverin) in cancer cell lines. Cancer Res. 60, 1914-20.

Mangelsdorf, D. J., Evans, R. M., 1995. The RXR heterodimers and orphan receptors. Cell. 83, 841-50.

Martinez-Morales, J. R., et al., 2003. OTX2 activates the molecular network underlying retina pigment epithelium differentiation. J Biol Chem. 278, 21721-31.

Matsui, T., et al., 1980. J. Biol. Chem. 255, 11992-11996.

Matys, V., et al., 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 31, 374 - 378.

McFall, R. C., et al., 1977. Characterization of a New Continuous Cell Line Derived from a Human Retinoblastoma. Cancer Res. 37, 1003-1010.

Morris, T. A., et al., 1997. Localization of upstream silencer elements involved in the expression of cone transducin alpha-subunit (GNAT2). Invest. Ophthalmol. Vis. Sci. 38, 196-206.

Muller, F., et al., 2007. New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. J Biol Chem. 282, 14685-9.

Ng, L., et al., 2001. A thyroid hormone receptor that is required for the development of green cone photoreceptors. Nat Genet. 27, 94-8.

Nishida, A., et al., 2003. Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. Nat Neurosci. 6, 1255-63.

Ohler, U., 2006. Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction. Nucleic Acids Res. 34, 5943-50.

Ohler, U., et al., 2002. Computational analysis of core promoters in the Drosophila genome. Genome Biol. 3.

Olsson, J. E., et al., 1992. Transgenic mice with a rhodopsin mutation (Pro23His): a mouse model of autosomal dominant retinitis pigmentosa. Neuron. 9, 815-30.

Otterson, G. A., et al., 1997. Incomplete penetrance of familial retinoblastoma linked to germ-line mutations that result in partial loss of RB function. Proceedings of the National Academy of Sciences of the United States of America. 94, 12036-12040.

Oyama, M., et al., 2004. Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. Genome Res. 14, 2048-52.

Peng, G. H., et al., 2005. The photoreceptor-specific nuclear receptor Nr2e3 interacts with Crx and exerts opposing effects on the transcription of rod versus cone genes. Hum Mol Genet. 14, 747-64.

Peng, G. H., Chen, S., 2005. Chromatin immunoprecipitation identifies photoreceptor transcription factor targets in mouse models of retinal degeneration: new findings and challenges. Vis Neurosci. 22, 575-86.

Perrault, I., et al., 2004. Retinal dehydrogenase 12 (RDH12) mutations in leber congenital amaurosis. Am J Hum Genet. 75, 639-46.

Pickering, B. M., Willis, A. E., 2005. The implications of structured 5' untranslated regions on translation and disease. Seminars in Cell & Developmental Biology. 16, 39-47.

Pickrell, S. W., et al., 2004. Deciphering the contribution of known cis-elements in the mouse cone arrestin gene to its cone-specific expression. Invest Ophthalmol Vis Sci. 45, 3877-84.

Ponjavic, J., et al., 2006. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome Biol. 7, R78.

Purnell, B. A., et al., 1994. TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. Genes Dev. 8, 830-42.

Quandt, K., et al., 1995. Matlnd and Matlnspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucl. Acids Res. 23, 4878-4884.

Reid, T. W., 1974. J. natn. Cancer Inst. 53, 347-360.

Roberts, M. R., et al., 2005. Retinoid X receptor (gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of the developing mouse retina. Invest Ophthalmol Vis Sci. 46, 2897-904.

Roberts, M. R., et al., 2006. Making the gradient: thyroid hormone regulates cone opsin expression in the developing mouse retina. Proc Natl Acad Sci U S A. 103, 6218-23.

Rogozin, I. B., et al., 2001. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. Bioinformatics. 17, 890-900.

Roni, V., et al., 2007. Mapping of transcription start sites of human retina expressed genes. Bmc Genomics. 8.

Roy, A. L., et al., 1993. An alternative pathway for transcription initiation involving TFII-I. Nature. 365**,** 355-9.

Roy, A. L., et al., 1991. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature. 354**,** 245-8.

Sachs, M. S., Geballe, A. P., 2006. Downstream control of upstream open reading frames. Genes Dev. 20**,** 915-21.

Sandelin, A., et al., 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet. 8**,** 424-36.

Schmid, C. D., et al., 2006. EPD in its twentieth year: towards complete promoter coverage of selected model organisms. Nucleic Acids Res. 34**,** D82-5.

Schulz, H. L., et al., 2004. The Retinome - defining a reference transcriptome of the adult mammalian retina/retinal pigment epithelium. BMC Genomics. 5**,** 50.

Seto, E., et al., 1991. YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. Nature. 354**,** 241-5.

Shaaban, S. A., Deeb, S. S., 1998. Functional analysis of the promoters of the human red and green visual pigment genes. Invest Ophthalmol Vis Sci. 39**,** 885-96.

Sharon, D., et al., 2002. Mutated alleles of the rod and cone Na-Ca+K-exchanger genes in patients with retinal diseases. Invest Ophthalmol Vis Sci. 43**,** 1971-9.

Sharpe, L. T. a. N., K., Total colour blindness: an introduction. In: R. F. Hess, Sharpe, L.T. and Nordby, K., (Ed.), Night Vision: Basic, Clinical and Applied Aspects. Cambridge University Press, UK Cambridge, 1990.

Shi, Y., et al., 1981. Transcriptional repression by YY1, a human GLI-Krüippel-related protein, and relief of repression by adenovirus E1A protein. Cell. 67**,** 377-388.

Shi, Y., et al., 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell. 67**,** 377-88.

Shiraki, T., et al., 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proceedings of the National Academy of Sciences of the United States of America. 100**,** 15776-15781.

Shuji Tachibanaki, Y. S.-M. D. A. S. K., 2007. Molecular Mechanisms Characterizing Cone Photoresponses. Photochemistry and Photobiology. 83**,** 19-26.

Smale, S. T., Baltimore, D., 1989. The "initiator" as a transcription control element. Cell. 57**,** 103-13.

Smale, S. T., Kadonaga, J. T., 2003. The RNA polymerase II core promoter. Annu Rev Biochem. 72**,** 449-79.

Sonnenburg, S., et al., 2006. ARTS: accurate recognition of transcription starts in human. Bioinformatics. 22**,** e472-80.

Srinivas, M., et al., 2006. Activation of the blue opsin gene in cone photoreceptor development by retinoid-related orphan receptor beta. Mol Endocrinol. 20**,** 1728-41.

Suzuki, Y., et al., 2002. DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. Nucl. Acids Res. 30**,** 328-331.

Suzuki, Y., et al., 2004. DBTSS, DataBase of Transcriptional Start Sites: progress report 2004. Nucleic Acids Res. 32**,** D78-81.

Tagle, D. A., et al., 1988. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol. 203**,** 439 - 455.

Takeda, J.-i., et al., 2007. H-DBAS: Alternative splicing database of completely sequenced and manually annotated full-length cDNAs based on H-Invitational. Nucl. Acids Res. 35**,** D104-109.

Thomas, M. C., Chiang, C. M., 2006. The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol. 41**,** 105-78.

Tokusumi, Y., et al., 2007. The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. Mol Cell Biol. 27**,** 1844-58.

Trifunovic, D., et al., 2008. A High-Resolution RNA Expression Atlas of Retinitis Pigmentosa Genes in Human and Mouse Retinas. Invest. Ophthalmol. Vis. Sci. 49**,** 2330-2336.

Trinklein, N. D., et al., 2003. Identification and functional analysis of human transcriptional promoters. Genome Res. 13**,** 308-12.

Turner, D. L., Cepko, C. L., 1987. A common progenitor for neurons and glia persists in rat retina late in development. Nature. 328**,** 131-136.

Unoki, M., Nakamura, Y., 2003. Methylation at CpG islands in intron 1 of EGR2 confers enhancer-like activity. FEBS Letters. 554**,** 67-72.

Weitz, D., et al., 2002. Subunit Stoichiometry of the CNG Channel of Rod Photoreceptors. Neuron. 36**,** 881-889.

Venter, J. C., et al., 2001. The sequence of the human genome. Science. 291**,** 1304-51.

Wetts, R., Fraser, S. E., 1988. Multipotent precursors can give rise to all major cell types of the frog retina. Science. 239**,** 1142-1145.

Wiechmann, A., Howard, E., 2003. Functional analysis of the human recoverin gene promoter. Curr Eye Res. 26**,** 25-32.

Williams, S., et al., 2005. CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. BMC Biotechnology. 5**,** 17.

Wissinger, B., 1998. Neuer Erkenntnisse zur genetischen Ursach der kompletten Achromatopsie (komplette Farbenblindheit). Z. prakt. Augenheilkd.**,** 467-472.

Wissinger, B., et al., 2001. CNGA3 mutations in hereditary cone photoreceptor disorders. Am J Hum Genet. 69**,** 722-37.

Wissinger, B., et al., 1997. Cloning, chromosomal localization and functional expression of the gene encoding the alpha-subunit of the cGMP-gated channel in human cone photoreceptors. Eur J Neurosci. 9**,** 2512-21.

Vlieghe, D., et al., 2006. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. Nucleic Acids Res. D95 - D97.

Wong, J. M., Bateman, E., 1994. TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. Nucleic Acids Res. 22**,** 1890-6.

Yanagi, Y., et al., 2002. Distinct functions of photoreceptor cell-specific nuclear receptor, thyroid hormone receptor beta2 and CRX in one photoreceptor development. Invest Ophthalmol Vis Sci. 43**,** 3489-94.

Yau, K. W., Baylor, D. A., 1989. A. Rev. Neurosci. 12**,** 289-327.

Ying, S., et al., 1998. A CAT reporter construct containing 277bp GNAT2 promoter and 214bp IRBP enhancer is specifically expressed by cone photoreceptor cells in transgenic mice. Curr Eye Res. 17**,** 777-82.

Yoshida, S., et al., 2004. Expression profiling of the developing and mature Nrl-/- mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl. Hum Mol Genet. 13**,** 1487-503.

Yu, J., et al., 2003. Annotation and analysis of 10,000 expressed sequence tags from developing mouse eye and adult retina. Genome Biol. 4**,** R65.

Zagotta, W. N., Siegelbaum, S. A., 1996. Structure and Function of Cyclic Nucleotide-Gated Channels. Annual Review of Neuroscience. 19**,** 235-263.

Zhang, T., et al., 2004. Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. Genome Res. 14**,** 79-89.

Zheng, J., et al., 2002. Rod Cyclic Nucleotide-Gated Channels Have a Stoichiometry of Three CNGA1 Subunits and One CNGB1 Subunit. 36**,** 891-896.

Zhong, H., et al., 2002. The heteromeric cyclic nucleotide-gated channel adopts a 3A:1B stoichiometry. Nature. 420**,** 193-198.

Zhu, X., et al., 2002. Mouse cone arrestin expression pattern: light induced translocation in cone photoreceptors. Mol Vis. 8**,** 462-71.