SUPPLEMENTARY INFORMATION

GENOMIC STRUCTURE OF IDENTIFIED NATs

1) *Pax6OS*

The analysis of the murine *Pax6* genomic locus revealed the presence of two overlapping mouse ESTs (one from E14.5 neural retina and the other from a neonate head), distributed over a 133 kb genomic interval upstream of the first coding exon and overlapping the first intron of a splice isoform of *Pax6* (Figure 1A). RT-PCR experiments on adult retina RNA confirmed the existence of these cDNAs and revealed the presence of further extent of alternative splicing (Figure 1A) with at least six isoforms and seven exons. Sequence analysis of this cDNA (*Pax6 opposite strand, Pax6OS*) did not reveal the presence of any long open reading frame (ORF). A similar analysis of the Human Genome Browser at UCSC (July 2003 assembly) revealed the presence of a spliced human retina cDNA sequence (Genbank accession BX648962), with a poly-adenylation site that spans a region of about 70 kb upstream the first exon of the *PAX6* gene and is transcribed from the opposite strand (*PAX6OS*, Fig. 1B).

2) *Six3OS*

The analysis of the *Six3* mouse genomic region revealed the presence of about twenty spliced ESTs, mostly derived from retina cDNA libraries, that cover a region of about 20 kb overlapping the first intron of the *Six3* gene. Sequence alignment of these ESTs allowed us to group them into 11 different splice isoforms (Figure 1A). In particular, isoform *Six3OS11* turned out to overlap 34 bp of the first exon of a *Six3* transcript (accession number X90871). We found a putative ORF of 597 nucleotides in form *Six3OS1* that was predicted to encode a protein of 199 amino acids with no significant sequence similarity to any known proteins. The *Six3OS1*, *5*, and *7* isoforms were characterized by the presence of a poly-adenylation site. The analysis of the corresponding human genomic sequence, through the UCSC server, allowed us to determine that in the human genome there were also several cDNAs localized upstream of the *SIX3* gene and transcribed in the opposite orientation (Figure 1B). The putative human *SIX3OS* cDNAs showed no or low sequence similarity to the corresponding

murine transcripts: the highest value of sequence similarity (86% over a region of 125 bp) was found between murine *Six3OS*5 and human *SIX3OS5*.

3) *Six6OS*

In silico analysis of the mouse *Six6* upstream genomic region revealed the presence of a full-length anonymous cDNA (Genbank accession NM_029444). This cDNA was also represented by several ESTs, some of which extended its sequence towards the 5' end of the *Six6* gene and overlapped its first intron (Figure 1A). Therefore we termed *Six6OS* the NM_029444 cDNA. By RT-PCR on RNA from mouse adult retina using both *Six6* and *Six6OS* specific primers, we were able to connect the 5' untranslated region (5'-UTR) of *Six6* with an exon of the *Six6OS6* isoform (Figure 1A). The overlap between the two transcripts extended for 190 bp of the *Six6* 5'-UTR. Sequence analysis of the *Six6OS* cDNA sequences allowed us to detect also in this case the presence of extensive alternative splicing, with at least seven alternative transcripts (Figure 1A). Furthermore, we detected in the *Six6OS1* isoform a long ORF of 1722 bp that is predicted to encode a protein of 574 amino acids with no significant sequence similarity to any known gene or protein product, except for the presence of a coiled coil region. A poly-adenylation signal was found in the *Six6OS1*, *2*, and *3* isoforms. Interestingly, *Six6OS1* was found to be significantly conserved in human (C14Orf39, accession number NM_174978) both in terms of sequence (with the ORF showing 81 and 70% identity at the nucleotide and at the protein level, respectively) and genomic position, as it is also localized in the region upstream the *SIX6* transcript with at least three splice isoforms (Figure 1B).

4) Vax2OS

The *Vax2* gene was found to be flanked at its 5' end by several spliced ESTs, mostly derived from retina cDNA libraries and belonging to the Unigene cluster Mm.185937. Sequence alignment of these ESTs as well as RT-PCR experiments allowed us to demonstrate that these sequences were part of the same transcript (termed *Vax2OS*). The *Vax2OS* cDNA encompasses at least 19 kb of genomic DNA, with its 5'-most exon extending into the first intron of *Vax2*, without overlapping *Vax2* exon sequences. Similar

to the previously described cDNAs, the *Vax2OS* also undergoes extensive alternative splicing with at least four different isoforms (Figure 1A), three of which (isoforms 1, 2, and 3) display a poly-adenylation signal. Sequence analysis of these isoforms failed to identify any long ORF, suggesting that the *Vax2OS* may be a non-coding RNA. By analyzing the Human Genome Browser and the human EST databases, we found no evidence of the presence of any spliced cDNA antisense to the human *VAX2* gene.

5) *CrxOS*

Several ESTs and an anonymous cDNA (AJ089116), all belonging to the UniGene cluster Mm.259041 flank the 5' end region of the mouse *Crx* gene (with a distance of 3 kb between their closest ends). Electronic assembly of these ESTs and RT-PCR experiments allowed us identify at least six alternative splice isoforms in the *CrxOS* transcript (Figure 1A): isoform 4 is characterized by the presence of a poly-adenylation signal at its 3' end. No evidence was found, either by RACE-PCR or by RT-PCR, of direct sequence overlap between the *Crx* and the *CrxOS* transcripts. Sequence analysis of the *CrxOS* isoforms revealed an ORF of 738 nucleotides that is predicted to encode a protein of 246 amino acids (accession XP_133196), characterized by the presence of two putative homeodomains (at position 13-75 and 123-185), as assessed by the SMART, Prosite and Pfam servers). Although the *CrxOS* ORF is not predicted to contain any canonical repeats, as determined through the Repeatmasker server (http://ftp.genome.washington.edu/RM/RepeatMasker.html), we determined, by BLASTN analysis, that the *CrxOS* ORF sequence seems to be present in several (at least 25) incomplete and unspliced copies (showing up to 98% nucleotide identity) in the mouse genome.

When we analyzed the corresponding human genomic region, we found one anonymous cDNA (AK097640) and two retina ESTs in the region upstream of the *CRX* gene. By RT-PCR experiments, we were able to demonstrate that both the AK097640 and the two ESTs were part of the same transcript (*CRXOS*). RACE-PCR experiments allowed us to extend the 5' end of the *CRXOS* transcript and to identify three further upstream exons (Figure 1B). Interestingly the human cDNA overlaps, over about 100 bp, an alternative 5'-UTR exon of the *CRX* transcript described by Hodges et al ((1)

(accession AF335593), which renders *CRXOS* a *bona fide* antisense to *CRX*. Similar to its murine counterpart, the human *CRXOS* cDNA also contains a long ORF, 1524 bp long, which is predicted to encode a protein of 508 amino acids containing a complete homeodomain sequence of the paired type. No sequence similarity could be found between the mouse *CrxOS* and the human *CRXOS*, either at the nucleotide or at the protein level, even in the homeodomains.

6) *Otx2OS*

The analysis of the *Otx2* locus through the UCSC Mouse Genome Browser allowed us to identify at least eight spliced ESTs and/or anonymous cDNA belonging to three Unigene clusters (Mm.134273, Mm.165199 and Mm.134184). These sequences span 200 kb of genomic DNA upstream the first exon of the *Otx2* gene and are transcribed in the opposite orientation. *In silico* assembly of these sequences revealed the presence of at least three splice isoforms of the *Otx2OS* transcript (Figure 1A) covering the *Otx2* upstream genomic region. RT-PCR and RACE-PCR experiments on adult mouse retina RNA using primer specific to these novel cDNAs and to the *Otx2* transcript, allowed us to determine that the 5' end of the *Otx2* transcript overlaps the *Otx2OS1* isoform by extending into its first exon (Figure 1A). Sequence analysis of the *Otx2OS* isoforms did not reveal the presence of any long ORF suggesting that these sequences may be part of a non-coding RNA. A similar *in silico* analysis of the corresponding human genomic locus allowed us to detect the presence of a putative *OTX2OS* cDNA represented by several ESTs, mostly derived from eye cDNA libraries. Extensive alternative splicing was found also in the human *OTX2OS*, with the presence of at least three different isoforms (Figure 1B). Significant sequence similarity between the murine and the human *Otx2OS* cDNAs (95% identity) was restricted to only one exon, corresponding to the second exon of form *Otx2OS*1 of the murine cDNA. The same exon was found to be conserved also in additional species, besides human and mouse, such as rat and pig, as assessed by BLASTN analysis vs. the EST databases, suggesting that this particular DNA fragment may harbor a sequence domain with an important functional role.

7) *Pax2OS*

Two mouse spliced ESTs, from a cDNA clone derived from an embryonic cDNA library, were found to be localized in the third intron of the *Pax2* gene (Figure 1A). This cDNA is transcribed in the opposite orientation with respect to the *Pax2* gene and was thus named *Pax2OS*. A poly-adenylation signal was found at the 3' end of this transcript. Sequence analysis of the *Pax2OS* cDNA revealed neither any significant sequence similarity to any known gene or protein nor the presence of any long ORF. Analysis of the corresponding human *PAX2* genomic locus, from the July 2003 release of the Human Genome Browser at UCSC, did not show the presence of any putative human *PAX2OS* cDNA.

8) *RaxOS*

By analysing the *Rax* genomic locus, we found two mouse eye overlapping cDNA sequences (AK033044 and BG292914), transcribed in the opposite orientation and complementary to the second and the third coding exons of the *Rax* cDNA (Figure 1A). Therefore, this cDNA, similar to the previously described *Six3OS*, *Six6OS Otx2OS* and the human *CRXOS*, behaves as a canonical antisense to the *Rax* cDNA (*RaxOS*). *RaxOS* does not show any long ORF, displays a poly-adenylation signal and does not seem to be conserved in the human genome, as assessed through the analysis of the July 2003 release of the Human Genome Browser at UCSC.

REFERENCE

1. Hodges, M.D., Vieira, H., Gregory-Evans, K. and Gregory-Evans, C.Y. (2002) Characterization of the genomic and transcriptional structure of the CRX gene: substantial differences between human and mouse. *Genomics,* **80,** 531-42.

SUPPLEMENTARY TABLE 1. Oligonucleotide primers used for expression analysis of murine OS transcripts by RT-PCR

GENE name	Oligonucleotide name	Oligonucleotide sequence	Product size (bp)	Annealing temperature (°C)
Otx2	Otx2-Es3-F	GGCAGGCCTCACTTTGTTCTGACC	311	65
	Otx2ex1R	GCTCCGCGTTCGCTGCGTCATTC		
Vax2	Vax2-F	GTAGCGCGAGTCCGAGGGAG	387	62
	Vax2-R	GCTTCTCCAGGTCTCTGCTC		
Crx	Crx-F	CATCCAGGAGAGTCCCCATTT	220	58
	Crx-R	GACAGCCTGGTGCATCAGG		
Rax	Rax-F	ACACCTTTCCTGCAGAGAGGAG	183	58
	Rax-R	GTGCTTCTCCTTGCTCCTTG		
Pax6	Pax6-F	CAGCTTGGTGGTGTCTTTGTC	344	58
	Pax6-R	CCCTCGGATAATAATCTGTCTCG		
Six3	Six3N-F	CACGCTCCCGGCTTCTCTTAC	290	58
	Six3N-R	CTGGAACATGGACAACTCTTCCG		
Six6	Six6EX1F	CAGCTGCCCATTTTGAATTTC	732	58
	Six6EX2R	GTGATGGAGATGGCCGAAG		
Otx2OS	Otx2-AS-F	GCAGGTCCTGGCTCTTGGTCCG	149	67
	Otx2-AS-R	CAACTTTTCCAGGCGACACAGGAG		
Vax2OS	Vax2AS-F	CTTCCCTCTGTTCCCTATATCTCC	158	58
	Vax2AS-R	CAAGGAGTTCATTCCAGATTCTAGG		
CrxOS	Crx-AS-F	TCCTGTACCCTGGGTCCC	700	58
	Crx-AS-R2	ACATGCCATCTCTGCTTTGG		
RaxOS	Rax-AS-F	TTTAACCGGCGGCTTACAAG	377	58
	Rax-AS-R	AGTGTCACTGACGGCTTGGA		
Pax6OS	Pax6-AS-F-B	GGAACTCTCGGAGGAGAGGAC	315	58
	Pax6-AS-R2-B	TACCAGCTTGCTCCAGTCCC		
Six3OS	Six3AS-F	AGCTCTGAGCGCGACCCTAC	200	58
	Six3AS-R	GAGCATACGAAGATGGGCTTC		
Six6OS	Six6AS-F2	GAATTTGGGTCCCTGGACAC	568	58
	Six6AS-R2	CTAAGCACACAAAATTGAGTGGG		

SUPPLEMENTARY TABLE 2. Oligonucleotide primers used for expression analysis of human OS transcripts by RT-PCR

GENE name	Oligonucleotide name	Oligonucleotide sequence	Product size (bp)	Annealing temperature (°C)
OTX2OS	OTX2AS-HsF	CTCTTTGTTGATTCCTGGCACC	211	60
	OTX2AS-HsR	AGGAATTCTGCCCAGACTCTGC		
CRXOS	CRXAS-HsF	AACTACGACCAGCGACTGAATCTG	396	58
	CRXAS-HsR	GGAATTGGGCCTGGAATTGAG		
SIX3OS	SIX3AS-HsF	CTCCAAAGTTTGAGTGCCCTC	404	58
	SIX3AS-HsR	CCTGGGCAGAGTTGTTGAATG		
SIX6OS	SIX6AS-HsF	AGCCAACATGTGATGTTTTTCG	238	58
	SIX6AS-HsR	GTTCAGTACATGCCAACACTCTGC		
OTX2	OTX2HsF	CAGTCAATGGGCTGAGTCTGAC	475	58
	OTX2HsR	GACAGTGGGGAGATGGAAGC		
CRX	CRXHsF	CTATTCTGTCAACGCCTTGGC	618	58
	CRXHsR	CTGAAATAGGAGCTCGGAGACC		
SIX3	SIX3HsF	CCACTTCTTGTTGCCAAACTTC	642	58
	SIX3HsR	CTGGAGGTTACCGAGAGGATG		
SIX6	SIX6HsF	AGCTGCCCATCTTGAATTTCAG	734	58
	SIX6HsR	TCAGATGTCGCACTCGCTG		

SUPPLEMENTARY TABLE 3. Oligonucleotide primers used to obtain, by RT-PCR, templates for RNA *in situ* hybridization experiments.

Gene name	Oligonucleotides	Oligonucleotide sequences	Size (bp)
Crx	Crx-F	CATCCAGGAGAGTCCCCATTT	1284
	Crx-R2	CAAGTTCTCAGCACCTTTACGG	
Rax	Rax-F	ACACCTTTCCTGCAGAGAGGAG	624
	Rax-R2	AAGAAAAGTTCGGCCACCAG	
Six6	Six6-EX1F	CAGCTGCCCATTTTGAATTTC	703
	Six6-EX2R	GTGATGGAGATGGCCGAAG	
Vax2OS	Vax2AS-F	CTTCCCTCTGTTCCCTATATCTCC	596
	Vax2AS-R2	TGCCAGGGACCTTGACCTAG	
CrxOS	Crx-AS-F	TCCTGTACCCTGGGTCCC	700
	Crx-AS-R2	ACATGCCATCTCTGCTTTGG	
RaxOS	Rax-AS-F	TTTAACCGGCGGCTTACAAG	570
	Rax-AS-R2	GTGCTCTGCAAGACCGGAG	
Pax6OS	Pax6-AS-F-B	GGAACTCTCGGAGGAGAGGAC	320
	Pax6-AS-R2-B	TACCAGCTTGCTCCAGTCCC	
Six3OS	Six3AS-F	AGCTCTGAGCGCGACCCTAC	610
	Six3AS-R2	ACTGGGCATCTTGCACTGTACTC	
Six6OS	Six6AS-F2	GAATTTGGGTCCCTGGACAC	565
	Six6AS-R2	CTAAGCACACAAAATTGAGTGGG	