 Human Molecular Genetics, 1999, Vol. 8, No. 1 **11–23**

EYA4, a novel vertebrate gene related to *Drosophila eyes absent*

Giuseppe Borsani[1,+], Alessandro DeGrandi[1,+], Andrea Ballabio[1,2], Alessandro Bulfone[1], Loris Bernard[1], Sandro Banfi[1], Claudio Gattuso[1], Margherita Mariani[3], Michael Dixon[4], Dian Donnai[5], Kay Metcalfe[5], Robin Winter[6], Marie Robertson[7], Richard Axton[7], Alison Brown[7], Veronica van Heyningen[7] and Isabel Hanson[7,*]

[1]Telethon Institute of Genetics and Medicine, [2]Universitá Vita e Salute San Raffaele and [3]Department of Biological Technological Research (DIBIT), San Raffaele Biomedical Science Park, Milan, Italy, [4]School of Biological Sciences, 3.239 Stopford Building, University of Manchester, Oxford Road, Manchester M14 9PT, UK, [5]Department of Medical Genetics, St Mary's Hospital, Hathersage Road, Manchester M13 0JH, UK, [6]Institute of Child Health, 30 Guildford Street, London WC1N 1EH, UK and [7]MRC Human Genetics Unit, Western General Hopital, Crewe Road, Edinburgh EH4 2XU, UK

Received July 31, 1998; Revised and Accepted 16 October, 1998 DDBJ/EMBL/GenBank accession nos Y17114, Y17115, Y17847, AJ007991-AJ008004

We have isolated a family of four vertebrate genes homologous to *eyes absent* (*eya*), a key regulator of ocular development in *Drosophila*. Here we present the detailed characterization of the *EYA4* gene in human and mouse. *EYA4* encodes a 640 amino acid protein containing a highly conserved C-terminal domain of 271 amino acids which in *Drosophila* eya is known to mediate developmentally important protein–protein interactions. Human *EYA4* maps to 6q23 and mouse *Eya4* maps to the predicted homology region near the centromere of chromosome 10. In the developing mouse embryo, *Eya4* is expressed primarily in the craniofacial mesenchyme, the dermamyotome and the limb. On the basis of map position and expression pattern, *EYA4* is a candidate for oculo-dento-digital (ODD) syndrome, but no *EYA4* mutations were found in a panel of ODD patients.

INTRODUCTION

The *eyes absent* (*eya*) gene is a key regulator of *Drosophila* development. The original allele at this locus, called *clift*, results in defective head development and is an embryonic lethal. Subsequently, a series of viable alleles has been identified, allowing genetic dissection of *eya* function (1). These alleles show a variety of phenotypic effects including reduced or absent eyes and ocelli, abnormal brain morphology and sterility (2). However, it is in the developing eye that *eya* has been studied most extensively and it is now clear that *eya* is a key regulator of ocular differentiation, playing an essential role in ensuring the ocular fate of the cells in the eye imaginal disc. The onset of ommatidial differentiation is marked by movement of the 'morphogenetic furrow' across the disc, which leaves in its wake clusters of maturing photoreceptors (3). *eya* normally is expressed in cells immediately ahead of the furrow but in *eya* mutants these cells die instead of being recruited into the normal pathway of ocular development (1). The discovery that ectopic *eya* expression can induce eye formation in non-eye imaginal discs—as was previously shown for *eyeless* (the *Drosophila* homologue of *PAX6*)—places *eya* near the top of the ocular regulatory hierarchy (4). A total of four genes, *eya*, *eyeless*, *sine oculis* (*so*) and *dachshund* (*dac*), are now known which are capable of acting synergistically to impose an ocular fate on non-eye discs: these loci form a genetic 'cassette' which specifies eye development (5–8). The molecular basis of the complex relationship between these genes lies at least in part in direct physical interactions between the protein products of eya, dac and so (5,6). These findings immediately raise the exciting possibility that the molecular interactions of the *Drosophila* eye cassette may also be conserved in vertebrates, establishing a fundamental genetic link between the way in which developmental programmes are initiated and executed in widely diverged species.

As a starting point for the isolation of vertebrate homologues of *eya*, we searched the expressed sequence tag database (dbEST) for entries with homology to *eya* (9) and identified two distinct human ESTs. Based on a comparison of the sequence of the corresponding cDNA clones with *Drosophila eya*, we designed degenerate primers for cross-species RT–PCR from human, mouse and chick, which resulted in identification of a family of four different vertebrate *eya*-related cDNA sequences. These cDNAs, designated *EYA1–4*, all contain a large C-terminal domain with extensive homology to *eya*. Three of these vertebrate genes (*EYA1–3*) have been reported by others (10,11). Functional conservation of the *EYA* genes has been strikingly illustrated by

*To whom correspondence should be addressed. Tel: +44 131 332 2471; Fax: +44 131 343 2620; Email: isabel.hanson@hgu.mrc.ac.uk
+These authors contributed equally to this work

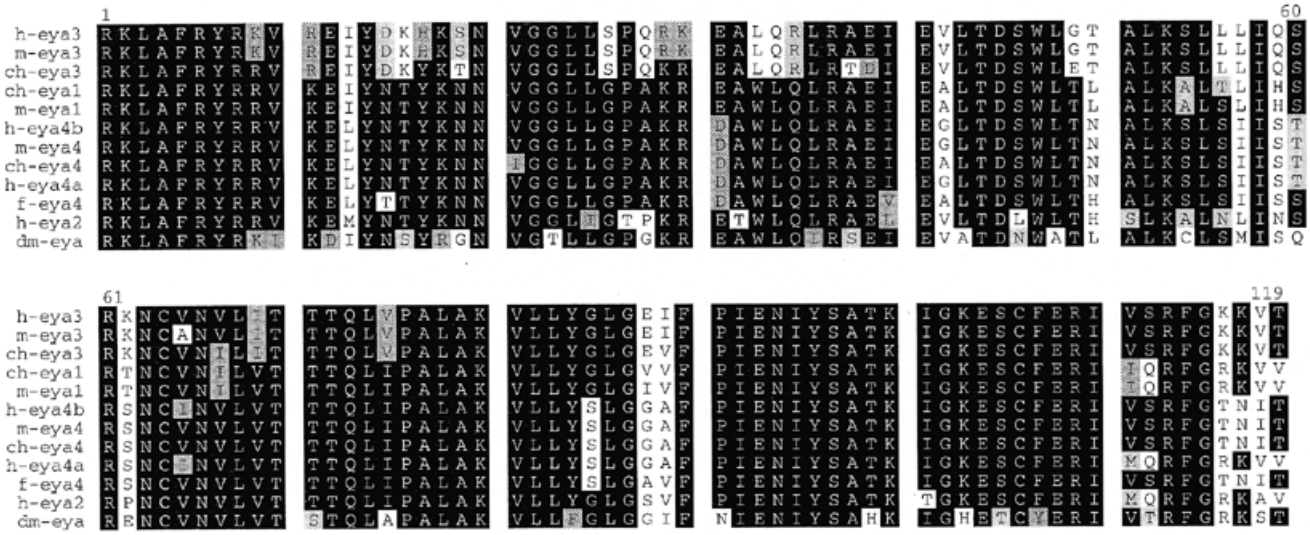


Figure 1. A family of four vertebrate cDNAs related to the *eya* gene of *Drosophila*. Alignment of conceptually translated cDNA sequences obtained by cross-species RT–PCR using degenerate primers against the highly conserved EYA-domain peptide motifs GGVDWM and YVVIGDG. h, human; m, mouse; ch, chick; f, *Fugu*; dm, *Drosophila* (1). Identical residues are shown on a black background, and conservatively substituted residues are shaded. Two RT–PCR products were identified for human *EYA4* which differ in nucleotide sequence at the 3′ end; these are derived from alternative splice forms using either exon 20 (h-eya4a) or exon 19 (h-eya4b; see text). The *Fugu Eya4* (f-eya4) sequence was obtained with the same primers from genomic clones (see text). Alignments were created with the programs Pileup and Prettybox from the GCG package. The nucleotide sequences of the RT–PCR products have the following accession nos: human EYA2, AJ007992; human EYA3, AJ007991; human EYA4'a' (exon 20 splice form), AJ007993; human EYA4'b' (exon 19 splice form), AJ007994; mouse Eya1, AJ007995; mouse Eya3, AJ007996; mouse Eya4, AJ007997; chicken Eya1, AJ008002; chicken Eya3, AJ008003; chicken Eya4, AJ008004; the *Fugu* Eya4-like sequence is comprised of four genomic fragments, AJ007998–AJ008001.

the ability of mouse *Eya3* to rescue the *Drosophila eya* phenotype (4), while in human development the importance of *EYA* genes is highlighted by the discovery that mutations in the *EYA1* gene underlie two dominantly inherited syndromes, branchio-oto-renal syndrome and branchio-oto syndrome (11–13). Here we describe the detailed characterization of mouse and human *EYA4*. Based on map location and expression pattern, *EYA4* was considered a strong candidate for human oculo-dento-digital syndrome (ODD) (14). However, we were unable to find any *EYA4* mutations in ODD patients despite a thorough search, and we conclude that ODD is not caused by general loss-of-function mutations in *EYA4*.

RESULTS

Isolation of vertebrate *EYA* genes

We began our search for vertebrate homologues of *eya* by exploiting the wealth of data generated by the EST sequencing project (9,15). By performing a simple text string search of dbEST, it was possible to identify sequences which had already been annotated as having significant homology to *eya* through routine screening of ESTs against known genes in the public databases. This initial stringsearch approach yielded two distinct human ESTs (accession nos H07988 and Z39529) which appeared to be derived from two separate human *eya*-related genes on the basis of nucleotide sequence. The corresponding cDNA clones (IMAGE IDs 45134 and 54132) were sequenced completely and found to have highly significant homology to the C-terminal region of *eya*. To identify any further members of the vertebrate gene family, we designed degenerate oligonucleotides against two peptide motifs GGVDWM and YVVIGDG which are separated by 119 amino acids and are conserved between the fly eya protein and the predicted product of IMAGE clone 45134. Degenerate RT–PCR products from human glioblastoma and lens cell lines, mouse embryos and chick embryos were cloned, and several dozen recombinants from each RNA source were sequenced. By combining the degenerate cross-species RT–PCR and EST searching approaches, and performing pairwise comparisons of the nucleotide and predicted amino acid sequences, we characterized a total of four distinct vertebrate *eya*-related cDNAs designated *EYA1–4* (Fig. 1). We also hybridized a pufferfish cosmid library with the insert of IMAGE clone 45134 (human *EYA2*) and identified a subset of cosmids which amplified at the genomic level with the degenerate primers. When sequenced, the genomic PCR product was found to contain four *eya*-related exons. Pairwise comparisons of the predicted translation product of these exons with those of the RT–PCR clones revealed that these *Fugu* cosmids contained the putative pufferfish homologue of *EYA4* (Fig. 1).

The mouse and human genes *EYA1–3* have been described in detail by others (10,11). We chose to focus on a new member of the family, *EYA4*.

Identification of mouse and human *EYA4* cDNAs

A mouse 11 day embryo cDNA library was screened with a mouse *Eya4* probe, MRT11, which was obtained by degenerate RT–PCR (Fig. 1, m-eya4). Analysis of eight positive clones and 5′ RACE PCR products led to the definition of a 2699 bp contig (accession no. Y17115) containing a 2551 bp open reading frame (ORF). The first ATG is located 319 bp from the beginning of the sequence, while a TAA translation termination codon was present at nucleotide 2171, resulting in a predicted 617 amino acid polypeptide. BLASTN analysis of dbEST revealed the presence of ESTs corresponding to both the mouse (accession no.

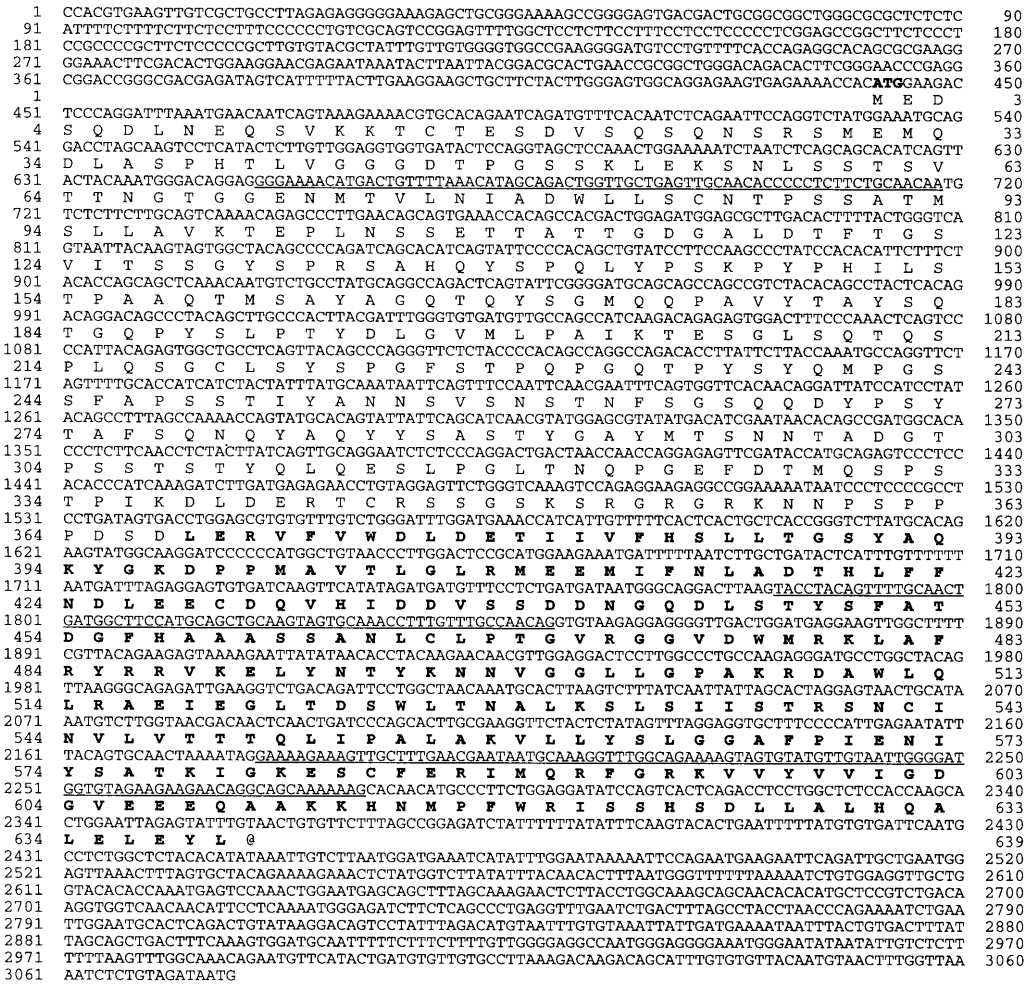


Figure 2. Nucleotide and predicted amino acid sequence of the human *EYA4* gene (accession no. Y17114). In the nucleotide sequence, the first in-frame ATG is shown in bold and the three regions affected by alternative splicing are underlined. These correspond to exon 5, the first 68 nucleotides of exon 16, and exon 20. The cryptic splice acceptor in exon 16 consists of the last 14 nucleotides of the second underlined region. In the amino acid sequence, the highly conserved C-terminal 271-amino acid 'EYA domain' is shown in bold. The position of the stop codon is indicated with @.

AA072879) and the human (accession no. AA176744) *EYA4* transcripts.

Screening of a human skeletal muscle cDNA library using the IMAGE cDNA clone 611269 (corresponding to EST AA176744) yielded >100 positive clones out of 1×10^6 recombinant phage plated. Nucleotide sequence analysis of 10 overlapping cDNA clones allowed us to build the 3077 nucleotide consensus sequence shown in Figure 2 (accession no. Y17114). The human transcript contains a single 1920 bp ORF with the first in-frame ATG codon at nucleotide 442. This ATG partially fulfils the Kozak criteria for an initiation codon (16) and is preceded, 12 bp upstream, by an in-frame stop codon. A TAA stop codon is present at nucleotide 2359, giving a predicted protein product of 639 amino acids with an estimated mol. wt of 69.5 kDa. The 5′- and 3′-untranslated regions are 441 and 716 bp, respectively.

Both mouse and human transcripts correspond to a novel mammalian gene that we named *EYA4*, according to the Human Genome Organization (HUGO) Nomenclature Committee's recommendation.

Sequence analysis and homology with *EYA* genes

Comparison of human and mouse *EYA4* nucleotide sequences revealed 79.7% identity for the entire cDNA sequence, a value that reaches 88.9% in the coding region, with an overall amino acid sequence identity of 90.9%. Interestingly, the human polypeptide has a stretch of 23 amino acids which is absent in the mouse protein (Fig. 3). Human cDNA clones have also been isolated which encode a protein lacking these amino acids. Genomic analysis revealed that the 23 amino acid sequence is encoded by exon 5 in the human gene, indicating the presence of alternative splicing.

A BLAST comparison of the EYA4 protein with a non-redundant protein database revealed significant homology with all the members of the mammalian EYA family, with the *Drosophila* eya protein and with a predicted *Caenorhabditis elegans* polypeptide from a 2.2 Mb genomic sequence from chromosome III. The homology is particularly high in the 271 amino acid C-terminal 'EYA-domain' region which begins at residue

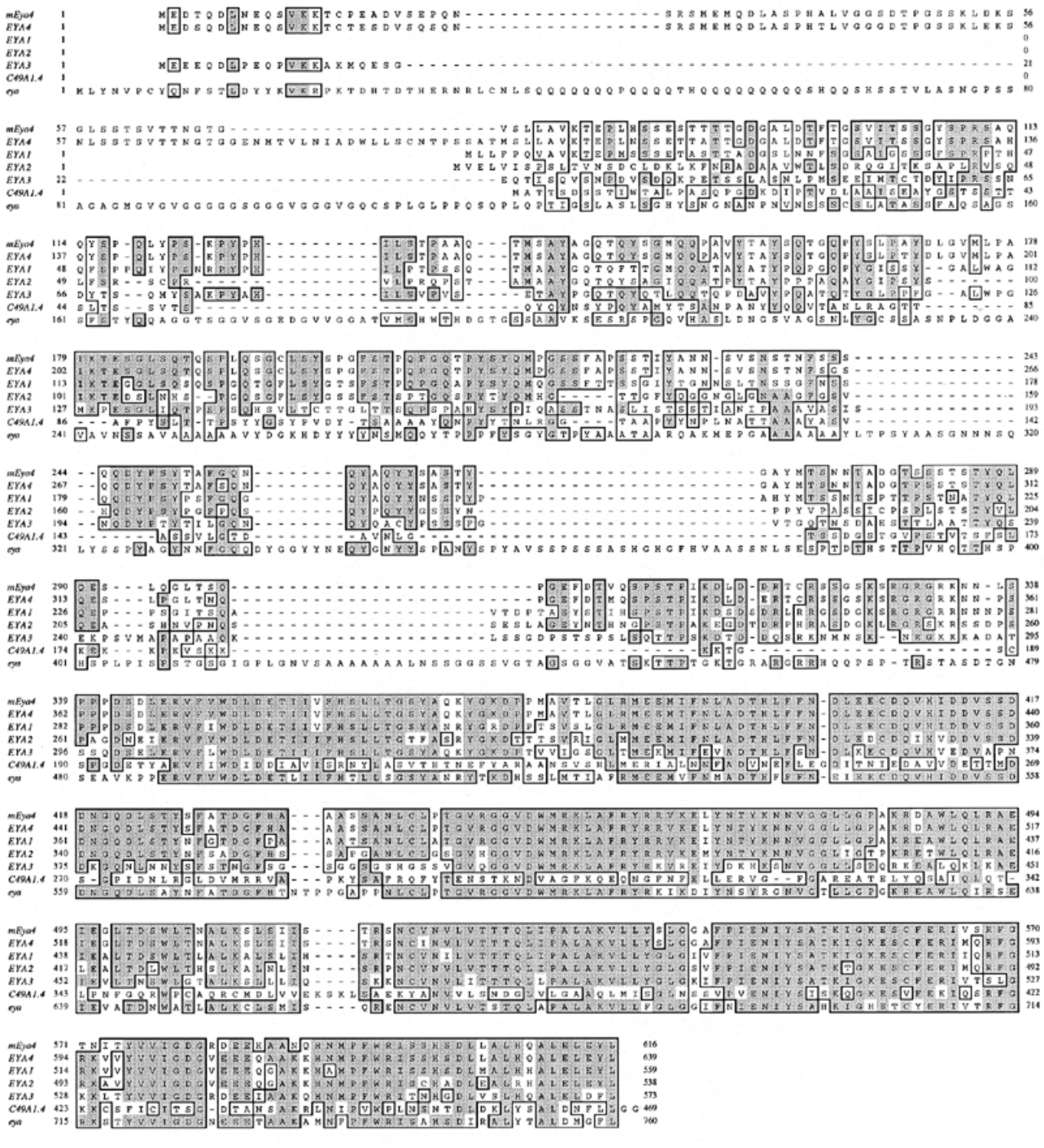


Figure 3. Multiple sequence alignment obtained using the Clustal W algorithm. mEya4, mouse Eya4 polypeptide (accession no. Y17115); EYA4, human EYA4 polypeptide (accession no. Y17114); EYA1, EYA2 and EYA3 polypeptides (accession nos Y10260, Y10261 and Y1026, respectively); eya, *D.melanogaster* eya protein (accession no. A45174); C49A1.4 *C.elegans* predicted protein product (accession no. Z83221). Identical residues (shaded) and conservatively substituted residues (unshaded) are boxed. The gaps inserted by the Clustal W program are represented with dashed lines.

369 in the human EYA4 protein sequence (Figs 2 and 3). The EYA4 EYA-domain shares 88.2% amino acid identity with EYA1, 79.7% with EYA2, 72.3% with EYA3, 70.1% with *Drosophila* eya and 24% with *C.elegans* predicted protein C49A1.4.

Genomic mapping of *EYA4* and *Eya4*

To determine the chromosomal assignment of the human *EYA4* gene, we performed radiation hybrid (RH) analysis using the Genebridge 4 panel as described in Materials and Methods.

Linkage was detected, with a lod score >3, to previously mapped reference markers, allowing us to localize *EYA4* between CHLC.GATA23B12 and AFMA074ZG9 on the MIT RH map of the long arm of chromosome 6 (data not shown). The chromosomal localization on 6q23 was confirmed independently by fluorescence *in situ* hybridization (FISH) analysis with a 2.1 kb *Xba*I genomic fragment containing exon 13 (data not shown).

Yeast artificial chromosome (YAC) clones (855_a_2, 958_g_10) corresponding to *EYA4* were identified in the CEPH human Mega-YAC library by PCR-based screening with the primers RH1 and RH2. These YACs are part of the Whitehead Institute/MIT Center for Genome Research chromosome 6 integrated map.

The mouse *Eya4* gene was genetically mapped on chromosome 10 between markers *D10Xrf5* and *Aco2* in a region that is homologous to human chromosome 6q22–q23 (Fig. 4).

To test a possible involvement of this gene in human inherited disorders, we used the Online Mendelian Inheritance in Man (OMIM) database to retrieve information on the disease loci mapped to the same chromosomal region. Interestingly, *EYA4* maps within the critical region for ODD syndrome (OMIM 164200) (14).

Genomic structure

Hybridization of cDNA probes to both cosmid and phage P1-derived artificial chromosome (PAC) arrayed libraries led to the identification of genomic clones corresponding to the *EYA4* gene. *Eco*RI cosmid fragments hybridizing to cDNA probes were subcloned in plasmid vectors and sequenced using oligonucleotide primers designed from the cDNA sequence. In addition, both cosmid and PAC clones were sequenced directly using *EYA4* exon primers. Twenty-one exons were identified and the sequence of all exon–intron boundaries was determined. Exon sizes and splice junction sequences are shown in Table 1.

Alternative splicing

The analysis of different *EYA4* cDNA clones and RT–PCR products led to the identification of three alternatively spliced forms of the major transcript presented in Figure 2. Form a is characterized by the absence of 70 nucleotides corresponding to exon 5 (Figs 2 and 5A) and is present in a number of cDNA clones identified from a human skeletal muscle cDNA library. Mouse cDNA clones derived from an 11 day embryo library also lack the sequence corresponding to human exon 5.

Form b was identified by RT–PCR in human lymphoblastoid and lens epithelial cell lines and is generated by the use of a cryptic splice acceptor site within exon 16 (bases 1836–1849 in Fig. 2; Table 1; Fig. 5). The removal of 68 bp from the start of exon 16 results in a truncated predicted protein product of 452 amino acids, terminating at a TGA stop codon (bases 1866–1868, Fig. 2). A cDNA clone corresponding to form b has also been identified in mouse (data not shown).

Table 1. Splice junction table of the human *EYA4* gene

Exon no.	Splice acceptor		Splice donor		Exon size (bp)
1			CGACGAG	**gt**gagtgacccc	–
2	ttcttgtttt**ag**	ATAGTCA	ACAATCA	**gt**aagtcttcat	98
3	cttttcttac**ag**	GTAAAGA	ATTCCAG	**gt**acagtaaaat	50
4	ctgctctacc**ag**	GTCTATG	ACAGGAG	**gt**aagtgtacta	125
5	tcctgttctc**ag**	GGGAAAA	GCAACAA	**gt**atgagagatg	70
6	tcttctacgt**ag**	TGTCTCT	GGGTCAG	**gt**aaagccttta	93
7	tgtgtcttac**ag**	TAATTAC	CTTCCAA	**gt**aagtggtcag	67
8	tctgatattt**ag**	GCCCTAT	ACTTACG	**gt**atttcacatc	143
9	ttttctgaac**ag**	ATTTGGG	ATGCCAG	**gt**aagtagctac	144
10	ttttacctct**ag**	GTTCTAG	ACAACAG	**gt**atagtagctt	80
11	ttattgtttc**ag**	GATTATC	CAACCAG	**gt**acagatcttc	166
12	tgttatctat**ag**	GAGAGTT	CCTGGAG	**gt**atgcctactc	135
13	ttttgtcttc**ag**	CGTGTGT	TGGCAAG	**gt**aagaaattca	84
14	ttttctttac**ag**	GATCCCC	TTTAGAG	**gt**aagaatttta	90
15	tgtttttaac**ag**	GAGTGTG	ACTTAAG	**gt**aagctatgcc	59
16	tgggtgttac**ag**	TACCTAC	GTTGGAG	**gt**atgtgtggct	161
16a	gtttgccaac**ag**	GTGTAAG	GTTGGAG	**gt**atgtgtggct	93
17	ttggtgttgc**ag**	GACTCCT	GCACTAG	**gt**aagtggaatt	115
18	tttctcccat**ag**	GAGTAAC	AAAATAG	**gt**aaggaaatta	122
19	attgtttttc**ag**	GCAAGGA	TAACCAG	**gt**aacttcactc	101
20	tttatcttcc**ag**	GAAAAGA	AAAAAAG	**gt**aacctgtctc	101
21	ctaaccacac**ag**	CACAACA			

Exon 16a corresponds to the last 93 bp of exon 16 and is generated by the presence of a cryptic splice acceptor site.

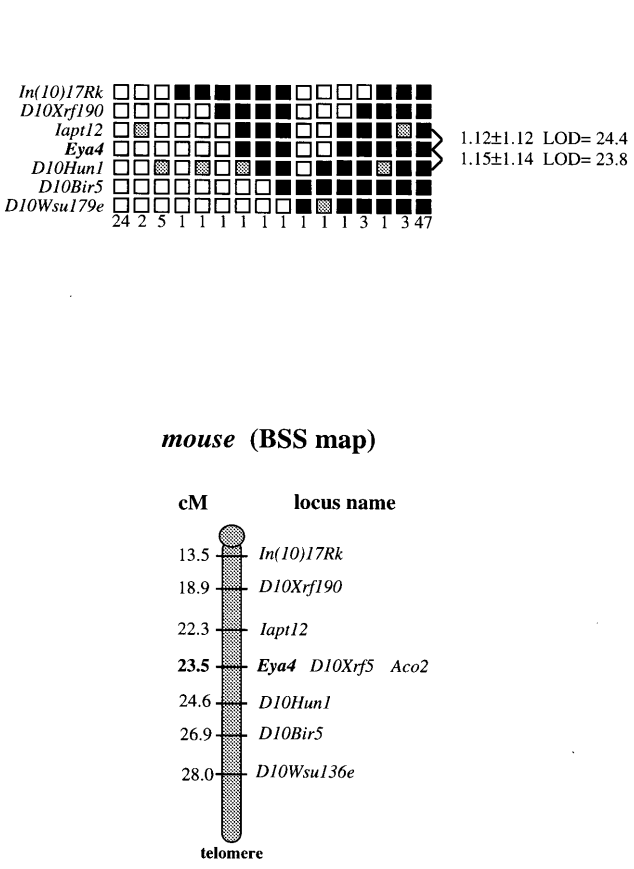


Figure 4. Mapping of *Eya4* in the mouse genome. (**A**) Haplotype and linkage analysis of *Eya4* and flanking loci on mouse chromosome 10 through the analysis of the BSS backcross (Jackson Laboratory). Empty squares, *Mus spretus* allele; solid squares, C57BL/6J allele; stippled squares, genotype not determined. The numbers to the right, between rows, indicate recombination fractions ± standard error, and LOD scores. The columns represent different haplotypes observed on chromosome 10. The numbers below the columns define the number of individuals sharing each haplotype. (**B**) Position of *Eya4* on chromosome 10 with respect to nearby markers independently mapped by others on the BSS backcross. The numbers on the left represent approximate genetic distances from the most centromeric chromosome 10 marker in this cross.

Form c is characterized by the substitution of exon 20 by a novel exon, exon 19 (Figs 2 and 5). Exons 19 and 20 (both 101 bp long) show 69.3% nucleotide sequence identity and 67.6% amino acid sequence identity and are separated by just 38 bp in genomic DNA. RT–PCR experiments with oligonucleotide primers located in the flanking exons demonstrated that form c is present in adult heart cDNA, while only the form containing exon 20 can be detected in skeletal muscle (data not shown). The mouse Eya4 sequence presented in Figure 3 appears to correspond to the form c splice variant. There is 95.2% identity between the amino acid sequences of the EYA-domains of Eya4 and the human EYA4 exon 20 isoform as shown in Figure 3, but the identity increases to 99.3% when Eya4 is compared with the human EYA4 exon 19 isoform.

Expression pattern of *Eya4*

The tissue distribution of *Eya4* expression was examined using northern blot analysis and RNA *in situ* hybridization to whole-mount embryos and sections.

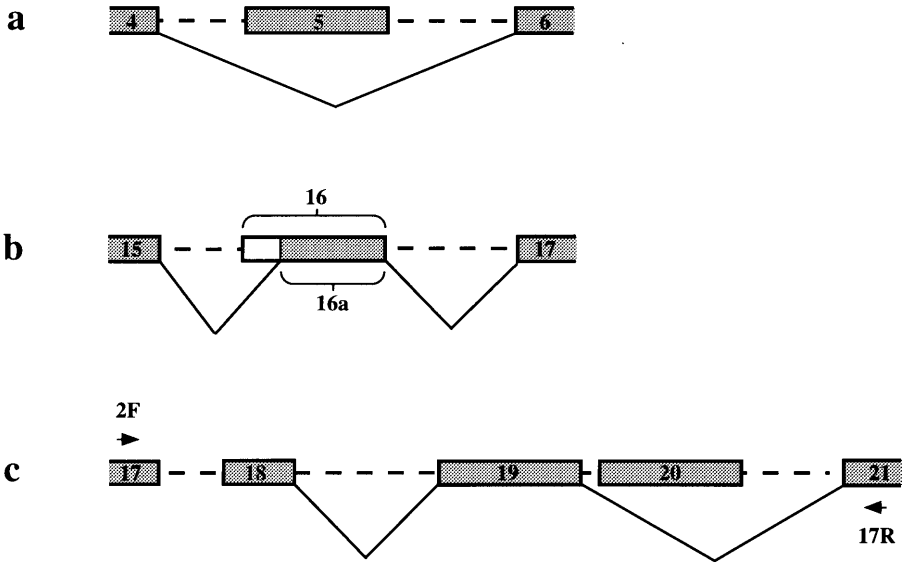
Hybridization of a cDNA probe from the 3′-untranslated region of mouse *Eya4* to a northern blot with poly(A)$^+$ RNA from various adult tissues revealed that *Eya4* is expressed in skeletal muscle (Fig. 6). These data are consistent with the fact that a large number of *EYA4* cDNAs were isolated from a human skeletal muscle cDNA library, suggesting a similar pattern of expression in man. The size of the mouse transcript, 5.5 kb, is larger than that of our *Eya4* cDNA contig, suggesting that some additional untranslated sequences are missing from our sequence.

To determine the expression pattern of *Eya4* in the developing mouse embryo, we performed RNA *in situ* hybridization to whole-mount embryos and sections. At E9.5, *Eya4* is expressed in the nasal placode, the otic vesicle and in two stripes flanking the dorsal midline above the developing forelimb bud (Fig. 7a and b). By E10.5, *Eya4* is clearly expressed in the branchial arch region and in the vicinity of the somites (Fig. 7c). At E11.5, *Eya4* expression is detected in a broad stripe of craniofacial mesenchyme above the nasal process and between the eyes (Fig. 7d and f). There is also very strong expression in the region of the somites, which on vibratome sectioning was found to correspond to the dermamyotome (Fig. 7e). At the level of the developing limbs, *Eya4*-positive cells appear to be migrating away from the dermamyotome into the limb structures in a pattern resembling that of migrating muscle precursor cells (17,18). At E12.5, the same craniofacial region is still positive (Fig. 8A and B) while in the limb, *Eya4*-expressing cells are now present in the condensing mesenchyme of the hand and foot plates, surrounding the pre-cartilage condensations of the digits in the region of the developing flexor tendons (Figs 7g and h and 8E and F). *Eya4* expression is also detected in the developing urogenital system (Fig. 8D) and in the tongue and jaw at E12.5 (Fig. 8B and C).

In summary, *Eya4*, like mouse *Eya1–3* and *Drosophila eya* itself, is widely expressed during development (2,10). However, in contrast to these other genes, we found no evidence for expression of *Eya4* in the developing eye at any of the stages examined. A major domain of *Eya4* expression includes the dermamyotome and cells apparently migrating from it to populate the developing limbs. Although the dermamyotomal localization was only confirmed at E11.5, the expression pattern at E10.5 (band of expressing cells above the forelimb, Fig. 7c) and E9.5 (parallel stripes of expression overlying the more rostral somites, Fig. 7b) is consistent with a role for *Eya4* in the maturation of limb muscle precursors. Like *Eya4*, *Eya1* and *Eya2* are expressed in the developing limb including the dermamyotome, migrating muscle precursors and tendons (19).

Mutation analysis of *EYA4* in ODD syndrome

The genomic map position of *EYA4*, together with the developmental expression pattern of *Eya4* in the limb and the face, made *EYA4* a promising candidate for ODD syndrome, a congenital disorder which shows autosomal dominant inheritance with high penetrance. Phenotypic expression is variable, but affected individuals often have a highly characteristic facial appearance with a thin nose, hypoplastic alae nasi (nasal wings) and anteverted nares (14). Typical involvement of the eyes, teeth and limbs includes microcornea or microphthalmia, hypoplasia of the dental enamel and syndactyly of the fourth and fifth fingers (type III syndactyly). Linkage analysis of six affected families placed the ODD locus at 6q22–q24 in an interval flanked by *D6S474* and *D6S292* (14); *EYA4* lies within this critical region.



Exon 19

```
G K E S C F E R I V S R F G T N I T Y V V I G D G R D E E H A A N Q
atagGCAAGGAAAGCTGTTTTGAGCGTATAGTGTCCAGATTTGGCACTAACATAACTTATGTTGTGATTGGAGATGGCCGAGATGAGGAGCATGCCGCTAACCAGcaca
: :: ::::: :: ::::: :: ::: ::   :: ::::::: :: ::  :::::::: ::::: :::::  ::: :: :: :: :: :: :: ::
atagGAAAAGAAAGTTGCTTTGAACGAATAATGCAAAGGTTTGGCAGAAAAGTAGTGTATGTTGTAATTGGGGATGGTGTAGAAGAAGAACAGGCAGCAAAAAAGcaca
G K E S C F E R I M Q R F G R K V V Y V V I G D G V E E E Q A A K K
```

Exon 20

Figure 5. *EYA4* undergoes alternative splicing. (Top) Schematic representation of three alternatively spliced forms (a, b and c) detected in the human *EYA4* transcript. Grey boxes represent either exons or portions of exons included in the alternative isoforms; arrows show the positions of oligonucleotide primers used in RT–PCR. Exon boundaries are reported in Table 1. (Bottom) Alignment of the alternative exons 19 and 20. Identical amino acids are underlined while identical nucleotides are indicated by **:**. Portions of the flanking exons 18 (left) and 21 (right) are indicated in lower case. The nucleotide sequence of exon 19 has been submitted to EMBL under the accession no. Y17847.

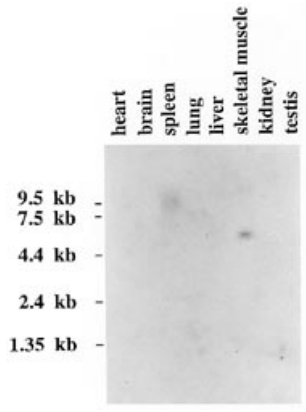


Figure 6. Northern blot analysis of the *Eya4* gene. A 400 bp cDNA probe corresponding to the 3′-untranslated region of *Eya4* was hybridized to a northern blot containing 2 μg of poly(A)$^+$ RNA from various mouse adult tissues. A band of ~5.5 kb is detected in skeletal muscle only.

Our patient panel consisted of four of the 6q-linked families previously reported, one individual from a two-generation family and three sporadic cases, thus giving a maximum possible total of eight different ODD alleles. We used a variety of approaches to search for mutations in the 21 exons of *EYA4*. Individual exons were amplified from genomic DNA and screened for sequence variants by single strand conformation polymorphism (SSCP), heteroduplex analysis and dideoxy fingerprinting. In addition, five of the exons (1, 2, 3, 7 and 19) were sequenced directly in seven of the eight ODD alleles. Southern blot analysis of patient DNA using three different *EYA4* probes covering both coding and non-coding regions failed to detect any altered hybridization pattern with respect to normal controls. Lymphoblastoid cell lines (LCL) were available from seven patients, three of which were from the 6q-linked family ODD 1 and one of which was from the 6q-linked family ODD 2 (14). LCL chromosomes were subjected to standard karyotypic analysis and in all cases were found to be normal. Chromosomal FISH analysis with a cosmid (Cos4) containing the 3′ end of *EYA4* was normal, with two strong signals detected on 6q in each case. In addition, LCL RNA was analysed by nested RT–PCR for splice mutations in the EYA-domain of *EYA4*. No difference was seen between patients and normal

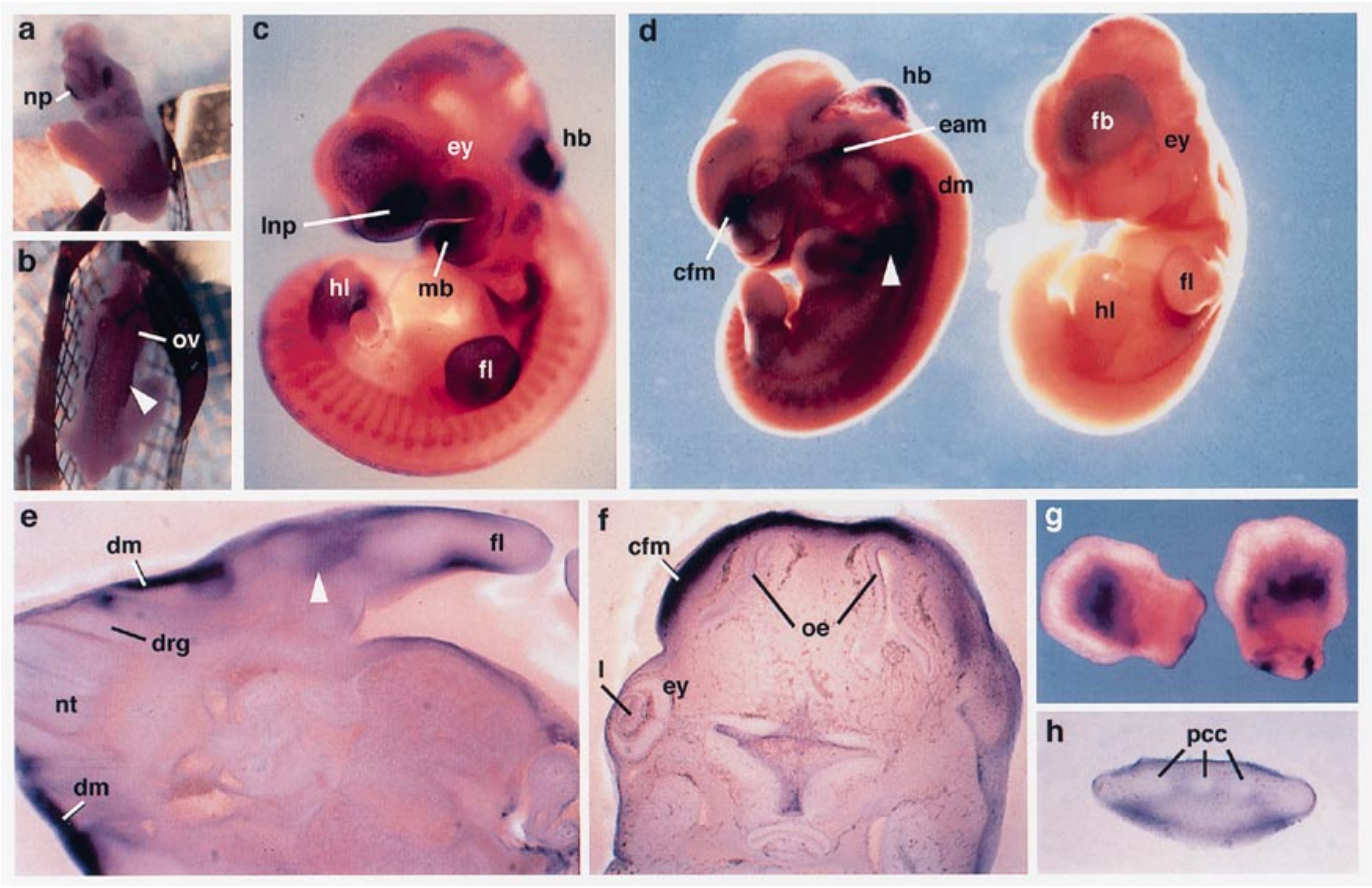


Figure 7. RNA *in situ* hybridization of whole-mount mouse embryos. (**a**) Ventral view and (**b**) dorsal view of E9.5 embryo. *Eya4* is expressed in the nasal placode (np), the otic vesicle (ov) and in two stripes flanking the dorsal midline above the level of the developing forelimb (arrowhead in b). (**c**) E10.5 embryo. *Eya4* is expressed in the lateral nasal process (lnp), mandibular process (mb), forelimb (fl) and hindlimb (hl). The eye (ey) is negative. The signal in the hindbrain (hb) is due to trapping. (**d**) E11.5 embryos hybridized with antisense (left) and sense (right) probes. *Eya4* is expressed in the craniofacial mesenchyme (cfm) above the nose and between the eyes, the future external auditory meatus (eam), the dermamyotome (dm), the forelimb and the hindlimb. Strong expression was also seen in a domain linking the dermamyotome with the limb (arrowhead). (**e**) Transverse vibratome section of an E11.5 whole-mount embryo at the level of the forelimb. *Eya4* is expressed strongly in the dermamyotomal compartment of the somites but not in the neural tube (nt) or dorsal root ganglia (drg). *Eya4* is also expressed in streams of cells which appear to be moving from the dermamyotome into the forelimb. (**f**) Transverse vibratome section through the head of an E11.5 whole-mount embryo. *Eya4* expression is clearly visible in the craniofacial mesenchyme overlying the olfactory epithelium (oe). The eye, including the lens (l), is negative. (**g**) Whole-mount hybridization of hindlimb plate (left) and forelimb plate (right) at E12.5. *Eya4* is expressed in the vicinity of the condensing digits. (**h**) A section through a whole-mount-hybridized forelimb showing that *Eya4* is expressed around but not within the pre-cartilage condensations (pcc) of the individual digits.

individuals. Given the size of our patient panel and the variety of analyses performed, we conclude that ODD is not caused by general loss-of-function mutations of *EYA4*, although we cannot completely rule out the possibility that ODD is caused by a specific *EYA4* mutation which we have been unable to detect.

DISCUSSION

We set out to isolate vertebrate homologues of the *Drosophila eya* gene. As a starting point, we screened dbEST, then complemented this with cross-species degenerate RT–PCR and cDNA library screening to define four vertebrate *EYA* genes. We have presented the detailed characterization of *EYA4*, a novel member of the family, in mouse and man. *EYA4* is most closely related to *EYA1*, but has a distinct map location and developmental expression pattern.

Like EYA1–3, the EYA4 protein contains a C-terminal 271 amino acid domain with extensive homology to *Drosophila* eya. Although the function of this EYA-domain is unknown at present, it recently has been shown that the eya protein plays a key role in *Drosophila* ocular development by tethering the DNA-binding function of so protein to the trans-activation function of dac protein under the control of *eyeless* (*PAX6*) (5,6,8). The protein–protein interactions eya–so and eya–dac form the molecular basis of a developmental 'cassette' in which a unique combination of gene products imposes a specific fate on a group of cells (7). These findings immediately raise the question of whether such regulatory cassettes are a general phenomenon, essential for the successful execution of specific developmental programmes, and whether they exist in vertebrates. Given that all the genes of the eye cassette have vertebrate homologues (20,21) it seems highly likely that the same molecular interactions will be conserved, although it is intriguing to note that the potential combinations are numerous given that there are four known homologues of *eya* (*EYA* genes) and five known homologues of *so* (*SIX* genes). In *Drosophila*, all the genes of the eye cassette are expressed separately elsewhere in non-ocular tissue; thus it is the specific combination of gene activity that is crucial, so that in theory a very small number of 'successful' proteins can be deployed in different

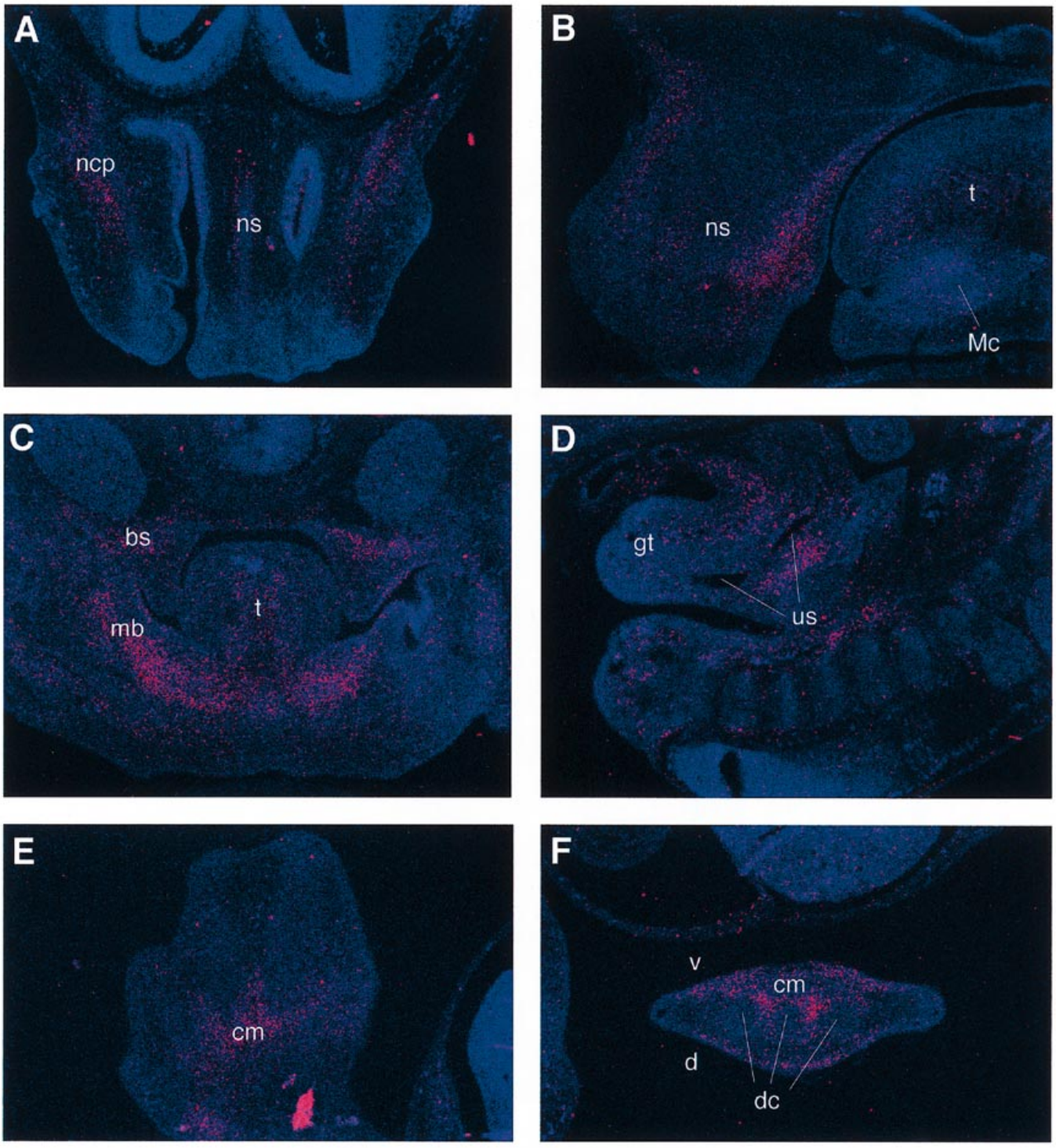


Figure 8. RNA *in situ* hybridization of E12.5 mouse tissue sections. Craniofacial expression of *Eya4* detected in coronal (**A** and **C**) and sagittal (**B**) sections. *Eya4* is expressed in the pre-cartilage primordium of the nasal septum (ns) and nasal capsule (ncp), intrinsic muscle of the tongue (t), condensing basisphenoid (bs) and mandible (mb) regions. Expression is detected in the pelvic region (**D**, sagittal section), in the genital tubercle and around the urogenital sinus. In the developing limbs, transcripts were detected in condensing mesenchymal cells (cm) in the ventral region of developing flexor tendons and around the digital cartilage condensations (dc), and in the ventral and dorsal ectoderm (**E**, hand plate longitudinal section; **F**, hand plate transverse section). Mc, Meckel cartilage; V, ventral; D, dorsal. The *Eya4* sense probe did not show any specific hybridization (data not shown).

combinations with different partners to induce specific tissue identities. In this respect, it is interesting to note that the expression pattern of *Eya4* in the dermamyotome and in cells emerging from the dermamyotome to populate the limb is remarkably similar to that of *Six1*, a mouse homologue of *Drosophila so* (22), and also overlaps extensively with *Pax3* (17)

and with *Eya1* and *Eya2* (19). The complex expression pattern of *Pax3, Six1, Six2, Eya1, Eya2* and *Eya4* in the precursors of the limb musculature and connective tissue may play a role in ensuring that these migrating cells remain committed to their fate until they reach their destination, and could clearly involve co-expression of PAX, EYA and SIX proteins at the cellular level (18,19,22). Given that *Six1* and *Six2* are expressed from E8.2 and E8.5, respectively, it will be of interest to examine the expression pattern of *Eya4* at these earlier stages (22).

Alternatively spliced transcripts have been observed in the *Drosophila eya* gene (1,2) and in human *EYA1* (12). In both cases, the splice variants are restricted to the 5′ end of the gene and create isoforms with alternative ATG start codons. Three alternatively spliced *EYA4* transcripts have been identified and it is notable that two of these affect the highly conserved C-terminal domain (Fig. 5). Form b results in a frameshift leading to the generation of a truncated protein, while form c results in the substitution of 11 amino acids. It will be of interest to determine whether the form c alternative splice, in which exon 20 is replaced by exon 19, involves the unusually short intron (just 38 bp) separating the two exons. The relative abundance and functional relevance of all the alternative splice forms remain to be investigated.

On the basis of the mouse embryonic expression pattern (in particular in the face and the limb) and human map location, *EYA4* was considered to be a good candidate for ODD syndrome. We used a variety of approaches to look for mutations and rearrangements of *EYA4* in a panel of ODD patients, including individuals from families which show linkage to 6q. No mutations were found, and we conclude that ODD is not caused by general loss-of-function mutations in *EYA4*. Perhaps significantly in view of the frequent eye involvement of ODD patients, we did not detect ocular expression of mouse *Eya4* at the developmental stages examined even though we identified several *EYA4* clones amongst the RT–PCR products from human lens cell lines. We could not rule out the possibility that human *EYA4* is expressed in the embryonic eye or that mouse *Eya4* is transiently expressed but not observed by us. In summary, we consider it very unlikely that *EYA4* is the ODD gene, although it remains a candidate for developmental syndromes which map to 6q23.

MATERIALS AND METHODS

EST database searching

The initial search was performed with a text screen of dbEST as previously described (9). Two distinct human ESTs were identified: GenBank accession nos H07988, from IMAGE clone 45134, and Z39529, from IMAGE clone 54132. The IMAGE cDNA clones were obtained from Research Genetics (USA) and the MRC HGMP Resource Centre (Hinxton, Cambridgeshire, UK). The sequence of each cDNA clone was determined by automated sequencing. IMAGE clone 45134 is derived from human *EYA2* and is present in the nucleotide sequence databases as DRES12 (accession no. U69178); IMAGE clone 54132 is derived from human *EYA3* and is contained entirely within the previously published *EYA3* sequence (11).

Degenerate RT–PCR

Based on the nucleotide sequence of human IMAGE clone 45134, two peptide motifs were identified (GGVDWM and YVVIGDG) which are conserved between *Drosophila* and human, and show relatively low codon redundancy, thus facilitating the design of degenerate primers. The forward primer (GGVDWM motif, nucleotides 1875–1858 in Fig. 2) was 5′-GGN GGN GTN GA(TC) TGG ATG-3′ and the two reverse primers (YVVIGDG motif, complement of nucleotides 2233–2252 in Fig. 2) were 5′-CCA TCN CC(TGA) ATN ACN ACA TA-3′ and 5′-CCA TCN CC(TGA) ATN ACN ACG TA-3′, where N is any nucleotide. For each RNA source, first strand cDNA was amplified in two separate PCRs containing the forward primer with each of the two reverse primers to avoid excessive degeneracy.

Total RNA was isolated from human glioblastoma cell lines, human lens epithelial cell lines, 12.5 day mouse embryos, 3.5 day chick embryos and 5.5 day chick embryo heads using total RNA isolation reagent (Advanced Biotechnologies). One microgram of total RNA was reverse transcribed in a volume of 20 µl with 25 U AMV reverse transcriptase (Boehringer Mannheim) using the supplied buffer in the presence of 50 pmol of random hexamers (Pharmacia), 1 mM dNTPs and 40 U RNase inhibitor (Boehringer Mannheim). Reverse transcription was carried out at 42°C for 1 h followed by 8 min at 75°C. For the amplification step, the reverse transcription reaction was diluted into 100 µl containing 1 µM of each degenerate oligo, 1× PCR buffer (Cetus) and 2 U Amplitaq (Cetus). PCR conditions were: step 1 (94°C for 30 s) ×1 cycle; step 2 (94°C for 30 s, 52°C for 45 s, 72°C for 30 s) ×30 cycles; step 3 (72°C for 10 min) ×1 cycle. At step 2, there was a ramp of 3°C/s between the annealing temperature (52°C) and the extension temperature (72°C). PCR products were blunt-ended with Klenow, kinased and ligated to *Eco*RI linkers (New England Biolabs) before cutting with *Eco*RI and ligating into *Eco*RI-cut, alkaline phosphatase-treated pBluescript SK+ (Stratagene). The recombinants from each RT–PCR reaction were screened by hybridization with the insert of IMAGE clone 45134, and at least 24 positive clones were sequenced in each case.

Isolation of *Fugu* genomic *EYA* clones

A gridded *Fugu rubripes* genomic cosmid library was obtained from the HGMP Resource Centre and screened by hybridization with the insert of IMAGE clone 45134. Positive clones were obtained from HGMP and used as templates for PCR amplification with the degenerate primers. Four cosmids (032C09, 096P07, 107H09 and 198L17) gave an identical 1.2 kb genomic PCR product which was cloned and sequenced to reveal the presence of exons related to the *EYA4* subfamily (f-eya4 in Fig. 1).

cDNA and genomic library screening

Phage library plating and screening conditions were as previously described (23). Human cDNA clones were isolated from a human adult skeletal muscle cDNA library (Clontech HL5002a) in Lambda GT10 vector. Mouse cDNA clones were isolated from a mouse 11 day embryo cDNA library (Clontech ML3003a) in Lambda GT10 vector. Recombinant phage recognized by the probes were isolated and the insert was recovered by *Eco*RI digestion and subcloned in pBluescript II SK– (Stratagene).

Both cosmid (Lawrence Livermore National Laboratory chromosome 6 arrayed cosmid library) and PAC (RPCI-5) human

genomic clones were isolated by hybridization with various *EYA4* cDNA probes.

DNA sequencing

Both manual (using a Sequenase Version 2.0 7-deaza-dGTP DNA sequencing kit from USB) and automated (using an Applied Biosystems ABI 377 fluorescent sequencer) sequencing were performed using vector- and gene-specific oligonucleotide primers. The nucleotide sequences of the human and the mouse *EYA4* cDNAs have been submitted to the databases under accession nos Y17114 (human) and Y17115 (mouse).

Computer sequence analysis

Sequence assembly and editing was performed using both the AutoAssembler version 1.4 (Applied Biosystems, Perkin Elmer) and DNA Strider 1.2 programs (24). Multiple sequence alignment was performed using the Clustal W algorithm (25) and Pileup [Wisconsin Package Version 9.1; Genetics Computer Group (GCG), Madison, WI]. Nucleotide and amino acid sequences were compared with the non-redundant sequence databases present at the National Center for Biotechnology Information (NCBI) using version 2.0 of BLAST (26). The global alignment program (GCG) was used to analyse identity and similarity among different nucleotide and amino acid sequences.

FISH mapping

FISH was performed on chromosome preparations obtained from lymphocyte metaphases from a 46,XY subject after 5-bromodeoxyuridine synchronization and thymidine release, with a 1 h colchicine block. A 2.1 kb *EYA4 Xba*I genomic probe, labelled with biotin-16-dUTP by nick translation (Boehringer Mannheim), was hybridized to chromosome preparations previously denatured at 70°C in 70% formamide/2× SSC. Hybridization mixture (30 µl/slide) contained the biotinylated probe (200 ng/slide) plus salmon sperm DNA (1 µg/slide) and Cot-1 DNA (10 µg/slide). After overnight hybridization (37°C), slides were washed at 37°C in 50% formamide/2× SSC (3×5 min) and in 2× SSC (3×5 min). Detection was performed by incubation of the slides with fluorescein isothiocyanate–avidin, according to the Oncor detection kit protocol with two amplification steps. Propidium iodide-counterstained slides were banded with 4′,6-diamidino-2-phenylindole (DAPI) and mounted in antifade solution (Vectashield mounting medium; Vector). Only those chromosomes with signals present on both chromatids at the same band position were taken into consideration.

Radiation hybrid mapping

For RH mapping, we used the Genebridge 4 panel (Research Genetics, Huntsville) which contains 93 human/hamster clones. The *EYA4* PCR primers were RH1 (5′-CAA GAT GAT CTC TAG GAA GGG AGA C-3′; located in the intron preceding exon 13) and RH2 (5′-ATA AGA CCC GGT GAG CAG TGA GTG-3′; located in exon 13, complement of nucleotides 1591–1614 in Fig. 2). These primers amplify a 163 bp product from human genomic DNA and give no specific amplification from hamster genomic DNA. DNA (25 ng) for each of the 93 hybrid clones, plus human and hamster genomic DNA controls, were used for PCR amplification in 96-well microtitre plates. Thirty-five cycles of amplification were performed (initial denaturation at 94°C for 2 min, followed by 94°C for 45 s, 60°C for 45 s and 72°C for 45 s). The result of PCR analysis was sent to the Radiation Hybrid Mapper server at the Whitehead Institute/MIT Center for Genome Research (http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl). This server returns an MIT framework marker linked to the subject STS with a LOD score >3.0.

Linkage mapping of *Eya4* in mouse

Genetic mapping was achieved utilizing a (C57BL/6j×SPRET/Ei)F_1 × SPRET/Ei (BSS) backcross generated and distributed by the Jackson Laboratory (Bar Harbor, ME) (27). A *Taq*I restriction fragment length polymorphism (RFLP) was identified by hybridization of C57BL/6j and SPRET/Ei parental DNAs cut with each of the six restriction enzymes (*Eco*RI, *Eco*RV, *Kpn*I, *Msp*I, *Taq*I and *Xba*I). Four Southern panels containing *Msp*I-cut parental DNAs and N_2 progeny DNAs (n = 94) were hybridized with an *Eya4* cDNA probe (clone 10.2). The resulting strain distribution pattern (SDP) was analysed with the Map Manager 2.6 program (28).

Expression studies

A northern blot of poly(A)$^+$ RNA from various mouse adult tissues was purchased from Clontech and hybridized using standard protocols (23).

RT–PCR experiments were carried out according to the manufacturer's recommended conditions (Gibco BRL and Boehringer Mannheim) (29). To investigate alternative splicing of *EYA4*, primers M4 (5′-CTC CAC ACC CAT CAA AGA TC-3′; nucleotides 1437–1456 in Fig. 2) and L868 (5′-TAA ATA TTC TCA ATG GGG AAA G-3′; complement of nucleotides 2141–2162 in Fig. 2) were used to amplify first strand cDNA from human lymphoblastoid and lens cell lines; a nested PCR was then performed using oligonucleotides M145 (5′-ACC TGT AGG AGT TCT GGG TCA-3′; nucleotides 1468–1488 in Fig. 2) and L722 (5′-CTA TAG AGT AGA ACC TTC GC-3′; complement of nucleotides 2110–2129 in Fig. 2). Agarose gel electrophoresis revealed the presence of a product of the expected size (662 bp) and a smaller product (~600 bp), which were sequenced directly as previously described (29).

Adult heart and skeletal muscle human cDNAs were amplified by RT–PCR using oligonucleotide primers 2F (5′-CTG ACA GAT TCC TGG CTA AC-3′; nucleotides 2002–2021 in Fig. 2) and 17R (5′-GTT TAA CTC CAT TCA GCA ATC-3′; complement of nucleotides 2508–2528 in Fig. 2). The resulting amplification products were analysed with restriction enzymes, subcloned into pBluescript vector and sequenced.

Digoxigenin-labelled sense and antisense riboprobes for use in whole-mount *in situ* hybridization were prepared from the mouse *Eya4* RT–PCR clone MRT11 (Fig. 1, m-eya4).

Hybridization to E9.5, E10.5, E11.5 and E12.5 mouse embryos was carried out as described previously (21).

Mouse embryo sectioning and radioactive *in situ* hybridization were performed as described (30). Embryos were fixed in 4% paraformaldehyde in phosphate-buffered saline and embedded in paraffin. Sets of serial sections were hybridized with [^{35}S]UTP-labelled antisense or sense riboprobes. A 900 bp mouse *Eya4* cDNA clone in pBluescript KS– was linearized with an appropriate restriction enzyme in order to transcribe either sense or antisense ^{35}S-labelled riboprobes, using the Stratagene RNA

transcription kit. Slides were exposed for 2–3 weeks. Sections were counterstained in Hoechst 33258 dye to stain the cell nuclei. The red colour represents the hybridization signal (Fig. 8).

ODD patient cohort

The ODD patient panel included individuals from the 6q-linked families ODD 1, ODD 2, ODD 4 and ODD 5 described by Gladwin *et al*. (14). Two affected members of each of these families were analysed, except for pedigree ODD 4 where all members were examined. In addition, we analysed one affected individual from a two-generation family, and three separate sporadic cases. LCL were available from three members of family ODD 1, one member of family ODD 2, one member of the two-generation family and all three sporadic cases.

Mutation analysis

Individual exons of the *EYA4* gene were amplified from genomic DNA under standard PCR conditions using primers located within the flanking introns (sequences available on request). SSCP analysis (31) was performed on 6% polyacrylamide gels (29:1 acrylamide:bis-acrylamide) with PCR fragments internally labelled by incorporating 0.1 µl of [α-^{32}P]dCTP (3000 Ci/mmol). Combined SSCP and heteroduplex analysis was carried out as previously described (32).

Southern blot analysis

Southern blot analysis was performed on 7 µg of genomic DNA using standard protocols (23).

ACKNOWLEDGEMENTS

We gratefully acknowledge J. Fletcher and P. Rashbass for technical guidance on *in situ* hybridization, R. Quinlan for the human lens cell lines, O. Zuffardi for FISH mapping assignment, A.M. Barbieri for radiation hybrid mapping, G. Consalez for assistance in mapping mouse *Eya4*, and the YAC Screening Centre at San Raffaele Biomedical Science Park, Milan. We also thank the TIGEM sequencing facilities for technical support. This work was supported by the Italian Telethon, the BIOMED2 CT97-BMH4-2284 EC grant (EURO-IMAGE) and the Merck Genome Research Institute (MGRI) grant no. 37. I.H. is supported by the Caledonian Research Foundation and the Iris Fund for Prevention of Blindness. V.v.H. is an International Scholar of the Howard Hughes Medical Institute.

REFERENCES

1. Bonini, N.M., Leiserson, W.M. and Benzer, S. (1993) The *eyes absent* gene: genetic control of cell survival and differentiation in the developing *Drosophila* eye. *Cell*, **72**, 379–395.
2. Bonini, N.M., Leiserson, W.M. and Benzer, S. (1998) Multiple roles of the *eyes absent* gene in *Drosophila*. *Dev. Biol*., **196**, 42–57.
3. Bonini, N.M. and Choi, K.W. (1995) Early decisions in *Drosophila* eye morphogenesis. *Curr. Opin. Genet. Dev.*, **5**, 507–515.
4. Bonini, N.M., Bui, Q.T., Gray-Board, G.L. and Warrick, J.M. (1997) The *Drosophila eyes absent* gene directs ectopic eye formation in a pathway conserved between flies and vertebrates. *Development*, **124**, 4819–4826.
5. Pignoni, F., Hu, B., Zavitz, K.H., Xiao, J., Garrity, P.A. and Zipursky, S.L. (1997) The eye-specification proteins So and Eya form a complex and regulate multiple steps in *Drosophila* eye development. *Cell*, **91**, 881–891.
6. Chen, R., Amoui, M., Zhang, Z. and Mardon, G. (1997) Dachshund and eyes absent proteins form a complex and function synergistically to induce ectopic eye development in *Drosophila*. *Cell*, **91**, 893–903.
7. Desplan, C. (1997) Eye development: governed by a dictator or a junta? *Cell*, **91**, 861–864.
8. Halder, G., Callaerts, P., Flister, S., Walldorf, U., Kloter, U. and Gehring, W.J. (1998) Eyeless initiates the expression of both sine oculis and eyes absent during *Drosophila* compound eye development. *Development*, **125**, 2181–2191.
9. Banfi, S., Borsani, G., Rossi, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitiello, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O. and Ballabio, A. (1996) Identification and mapping of human cDNAs homologous to *Drosophila* mutant genes through EST database searching. *Nature Genet*., **13**, 167–174.
10. Xu, P.X., Woo, I., Her, H., Beier, D.R. and Maas, R.L. (1997) Mouse *Eya* homologues of the *Drosophila eyes absent* gene require *Pax6* for expression in lens and nasal placode. *Development*, **124**, 219–231.
11. Abdelhak, S., Kalatzis, V., Heilig, R., Compain, S., Samson, D., Vincent, C., Weil, D., Cruaud, C., Sahly, I., Leibovici, M., Bitner-Glindzicz, M., Francis, M., Lacombe, D., Vigneron, J., Charachon, R., Boven, K., Bedbeder, P., Van Regemorter, N., Weissenbach, J. and Petit, C. (1997) A human homologue of the *Drosophila eyes absent* gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nature Genet*., **15**, 157–164.
12. Abdelhak, S., Kalatzis, V., Heilig, R., Compain, S., Samson, D., Vincent, C., Levi-Acobas, F., Cruaud, C., Le Merrer M., Mathieu, M., Oumlnig, R., Vigneron, J., Weissenbach, J., Petit, C. and Weil, D. (1997) Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of EYA1. *Hum. Mol. Genet*., **6**, 2247–2255.
13. Vincent, C., Kalatzis, V., Abdelhak, S., Chaib, H., Compain, S., Helias, J., Vaneecloo, F.M. and Petit, C. (1997) BOR and BO syndromes are allelic defects of *EYA1*. *Eur. J. Hum. Genet*., **5**, 242–246.
14. Gladwin, A., Donnai, D., Metcalfe, K., Schrander-Stumpel, C., Brueton, L., Verloes, A., Aylsworth, A., Toriello, H., Winter, R. and Dixon, M. (1997) Localization of a gene for oculodentodigital syndrome to human chromosome 6q22–q24. *Hum. Mol. Genet*., **6**, 123–127.
15. Banfi, S., Borsani, G., Bulfone, A. and Ballabio, A. (1997) *Drosophila*-related expressed sequences. *Hum. Mol. Genet*., **6**, 1745–1753.
16. Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res*., **12**, 857–872.
17. Bober, E., Franz, T., Arnold, H.H., Gruss, P. and Tremblay, P. (1994) Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development*, **120**, 603–612.
18. Christ, B. and Ordahl, C.P. (1995) Early stages of chick somite development. *Anat. Embryol. (Berl.)*, **191**, 381–396.
19. Xu, P.X., Cheng, J., Epstein, J.A. and Maas, R.L. (1997) Mouse *Eya* genes are expressed during limb tendon development and encode a transcriptional activation function. *Proc. Natl Acad. Sci. USA*, **94**, 11974–11979.
20. Oliver, G., Mailhos, A., Wehr, R., Copeland, N.G., Jenkins, N.A. and Gruss, P. (1995) *Six3*, a murine homologue of the *sine oculis* gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development*, **121**, 4045–4055.
21. Hammond, K.L., Hanson, I.M., Brown, A.G., Lettice, L.A. and Hill, R.E. (1998) Mammalian and *Drosophila dachshund* genes are related to the *Ski* proto-oncogene and are expressed in eye and limb. *Mech. Dev*., **74**, 121–131.
22. Oliver, G., Wehr, R., Jenkins, N.A., Copeland, N.G., Cheyette, B.N., Hartenstein, V., Zipursky, S.L. and Gruss, P. (1995) Homeobox genes and connective tissue patterning. *Development*, **121**, 693–705.
23. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1987) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24. Marck, C. (1988) 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res*., **16**, 1829–1836.
25. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.
26. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.

27. Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Ko, M.S., Thurston, S.J. and Birkenmeier, E.H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome*, **5**, 253–274.
28. Manly, K.F. and Elliott, R.W. (1991) RI Manager, a microcomputer program for analysis of data from recombinant inbred strains. *Mamm. Genome*, **1**, 123–126.
29. Hanson, I.M., Seawright, A., Hardman, K., Hodgson, S., Zaletayev, D., Fekete, G. and van Heyningen, V. (1993) *PAX6* mutations in aniridia. *Hum. Mol. Genet*., **2**, 915–920.
30. Rugarli, E.I., Lutz, B., Kuratani, S.C., Wawersik, S., Borsani, G., Ballabio, A. and Eichele, G. (1993) Expression pattern of the Kallmann syndrome gene in the olfactory system suggests a role in neuronal targeting. *Nature Genet*., **4**, 19–26.
31. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl Acad. Sci. USA*, **86**, 2766–2770.
32. Axton, R.A., Hanson, I.M., Love, J., Seawright, A., Prosser, J. and van Heyningen, V. (1997) Combined SSCP/heteroduplex analysis in the screening for *PAX6* mutations. *Mol. Cell Probes*, **11**, 287–292.