Journal of Neurochemistry, 2003, **84**, 840–853 doi:10.1046/j.0022-3042.2003.01583.x

New *Otx2* mRNA isoforms expressed in the mouse brain

Virginie Courtois,* Gilles Chatelain,* Zhi-Yan Han,† Nicolas Le Novère,† Gilbert Brun* and Thomas Lamonerie*

**LBMC, ENS-Lyon, Lyon, France*
†*Institut Pasteur, Paris, France*

Abstract

The mouse *Otx2* gene is essential throughout head and brain development, from anterior–posterior polarity determination and neuroectoderm induction to post-natal sensory organ maturation. These numerous activities must rely on a very finely tuned regulation of expression. In order to understand the molecular control of the *Otx2* gene, we set out to isolate its promoter. During this quest, we identified three remote transcription start sites, two defining two new upstream exons and one mapping within the previously reported first exon. The three transcripts differed in their 5′ non-coding region but encoded the same protein. The transcription start nucleotides of each mRNA species have been mapped by RNase protection assays and by an RNA circularization technique. We have demonstrated that they are all used and linked to functional promoters. In addition to leader versatility, we also detected alternative splicing within the coding sequence that gives rise to a new protein endowed with an 8 amino-acid insertion upstream of the homeodomain. Combined analysis of the relative abundance of *Otx2* mRNA isoforms in representative tissues and *in situ* hybridization studies revealed distinct spatial and temporal, although partially overlapping, expression patterns of the mRNA isoforms. These findings provide new clues to a better understanding of the relationships between *Otx2* gene architecture and its complex regulatory requirements.

Keywords: alternative promoters, brain, development, *Otx2*, RNA circularization, transcription.
J. Neurochem. (2003) **84,** 840–853.

The *Otx2* gene has been identified as a key player in the molecular events that control early developmental programmes. The murine *Otx2* gene, a member of a family including *Otx1* and *Crx* genes, is an ortholog of the *Drosophila orthodenticle* (*otd*) gene. *Otd* mutation prevents formation of the head segments where it is expressed (Finkelstein *et al*. 1990; Simeone *et al*. 1993; Hirth *et al*. 1995; Royet and Finkelstein 1995). In the mouse embryo, *Otx2* shows an expression pattern strikingly reminiscent of its *Drosophila* counterpart and knockout mice totally lack anterior head and brain structures. Knockout phenotypes and chimera studies have demonstrated that *Otx2* is necessary for gastrulation and is also required for induction and maintenance of the anterior neural plate (Acampora *et al*. 1995; Matsuo *et al*. 1995; Ang *et al*. 1996; Rhinn *et al*. 1998, 1999). Detailed analysis of the *Otx2* early expression pattern has contributed to the establishment of the role of the visceral endoderm in the induction of the neuroectoderm (Rhinn *et al*. 1998). Elegant knock-in experiments and transgenic approaches have demonstrated that, in early gastrula, *Otx2* is required for cell movements that convert the early embryo proximal–distal axis into the definitive anterior–posterior axis (Kimura *et al*. 2000). At later development stages a role in sensory organ formation has been ascribed to *Otx2*, especially for the eye and retina (Matsuo *et al*. 1995; Baas *et al*. 2000). Haploinsufficiency leads to numerous cranial, brain and eye defects, confirming the importance of *Otx2* gene dosage for proper anterior head development (Acampora *et al*. 1997; Matsuo *et al*. 1995; Makiyama *et al*. 1997; Martinez-Morales *et al*. 2001). Finally, early developmental defects of *Otx2* null mouse embryos can be rescued by the knock-in of the *Drosophila otd* gene in the *Otx2* locus and, conversely, overexpression of the *Otx2* gene allows the recovery of brain and ventral nerve cord in the *Drosophila otd* mutant (Acampora *et al*. 2001; Leuzinger *et al*. 1998;

Received August 30, 2002; revised manuscript received November 8, 2002; accepted November 12, 2002.

Address correspondence and reprint requests to Thomas Lamonerie, LBMC, ENS-Lyon, 46 allée d'Italie, 69364 Lyon cedex 07, France.
E-mail: tlamoner@ens-lyon.fr

Abbreviations used: CMV, cytomegalovirus; ES cells, embryonic stem cells; EST, expressed sequence tag; nt, nucleotide; RACE, rapid amplification of cDNA ends; SDS, sodium dodecyl sulphate; SEAP, secreted alkaline phosphatase; UTR, untranslated region.

Nagao *et al*. 1998). These observations clearly establish that Otx2-dependent basic genetic programmes controlling head development are conserved between mammals and insects.

The *Otx2* gene encodes a homeodomain transcription factor of the paired class (Finkelstein and Perrimon 1990). It can bind to and transactivate a synthetic promoter containing a bico target site (Simeone *et al*. 1993). Logically, Otx2 is thought to exert its functions through transcriptional regulation of genes, although few direct targets have been identified. Among them are *tenascin-C*, *N-CAM*, *calponin*, *IRBP*, *Gn-RH*, *Clock* and *XWnt5a* (for a review, see Boncinelli and Morgan 2001). The biochemical and functional–structural relationships of the gene have been much less documented. Besides the initial exon–intron structure described by Simeone *et al*. (1993), gene determinants for cranial mesenchyme and visceral endoderm expression have been ascribed to only ill-defined regions (Kimura *et al*. 1997, 2000).

The fact that the *Otx2* gene can operate in such different contexts as early gastrula or differentiating neuroretina suggests that numerous developmental and/or morphogenesis signals participate in its regulation. In order to identify the *cis*-regulatory elements that govern spatio-temporal expression of this gene, we have started to characterize its proximal promoter. Since the transcription start site of the *Otx2* gene had not been mapped precisely, we first set out to determine the 5′ end of this gene. This led us to identify two new *Otx2* transcripts, diverging by their 5′-untranslated region (UTR). We have mapped their start sites and verified the functional promoter activity of the respective upstream regions. Finally, this paper describes a more complex structural organization than originally thought and suggests that the spatial and temporal expression of the *Otx2* gene also relies on subtle choices among different *Otx2* messages.

Experimental procedures

Cell culture

P19, 3T3 and 293T cell lines were cultured at 37°C, 5% CO_2 and 95% humidity in Dulbecco Modified Eagle's Medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal calf serum (Perbio Science, Helsingborg, Sweden), 50 U/mL penicillin, 50 μg/mL streptomycin and 2 mM L-glutamine (Invitrogen). ES 129/svPa cells (Gauthier *et al*. 1999) were cultured on 0.1% gelatin dishes and 1% non-essential amino acids (Invitrogen), 1 mM pyruvate, 0.2 mM β-mercaptoethanol and 1000 U/mL of recombinant leukaemia inhibitory factor were added to the culture medium. Differentiation was induced by removing the leukaemia inhibitory factor and plating the cells on non-treated bacterial dishes.

Primers

(1) 5′-GGCACTGAAAATCAACTTGC
(2) 5′-TCCAAGCAATCAGTGGTTGA
(3) 5′-GACAGGATGCAGAAGGAGAT
(4) 5′-TTGCTGATCCACATCTGCTG
(5) 5′-ACGTGGATCCAGCATGATGTCTTATCTA
(6) 5′-ACGTTCTAGATCACTTGTCATCGTCGTCCTTGTAG-TCTACTAGTCTTCCCAAAACCTGGAATTTCCA
(7) 5′-ACGTGCTAGCAGACCTGTAGAAGCTA
(8) 5′-ACGTACTAGTTTTGGGCTAGAACCATT
(9) 5′-ACGTGGATCCGCAAATCTCCCTGAGA
(10) 5′-ACGTGGATCCGTGGAGACGGGGCTT
(11) 5′-ACGTGGATCCCTTTTTAGTTAGTGCTGGA
(12) 5′-GATCACTAGTCTTCACAAAACCTGGAAT
(13) 5′-ACGTACTAGTGTGCTTGAATTAGACAGGA
(14) 5′-ACGTTGCTAGCTGGCCCTGCTCTGTGCAGCT
(15) 5′-AGAGCTTCCAGAACGTCGAG
(16) 5′-AAGATCAGACGTCCTCATGG
(17) 5′-TGTGCCCTAGTAAAAGTCGT
(18) 5′-ACTGGATCCTGGCGACTGAGCTCCTCTAGC
(19) 5′-GACCATGATTACGCCAAGCTCG
(20) 5′-ACTGGATCCTGCTCTCTCCTCCAGGAGCAA
(21) 5′-ACGTGAATTCCCGGGAGACGCTGGGCA
(22) 5′-ACGTGAATTCCTTTCCCTGGCAGAGGT

RNA isolation and analysis

Mice were killed by cervical dislocation and total RNA was prepared from fresh tissues with TRIzol® (Invitrogen) according to the supplier's conditions. Cytoplasmic RNA was prepared from tissue culture as follows: cells were collected with a rubber policeman, pelleted at 500 *g*, rinsed with phosphate-buffered saline and lysed in 50 mM Tris-HCl, pH 7, 150 mM NaCl, 12 mM magnesium acetate, 150 mM sucrose and 0.5% NP40. Nuclei were pelleted at 10 000 *g* and the supernatant fluid was adjusted to 25 mM EDTA and 1% sarkosyl. Proteins were removed by two rounds of phenol–chloroform extraction and RNA was recovered by ethanol precipitation.

For northern blotting analysis, 20 μg of mRNA were loaded per lane in a 1.5% agarose gel containing 5% formaldehyde and separated by electrophoresis in a 1× 3-(*N*-morpholino)propanesulfonic acid (MOPS; 20 mM) buffer. RNA was transferred onto a nylon membrane (Hybond N) in 150 mM NH_4CH_3COO. Hybridization was performed overnight at 65°C in a buffer containing 1% bovine serum albumin, 200 mM sodium phosphate, pH 7, 15% formamide, 1 mM EDTA and 7% sodium dodecyl sulphate (SDS). A PCR product corresponding to the coding sequence of *Otx2* (see below) was labelled using [^{32}P]δXTΠ (3000 Ci/mmol) and a Ready-to-go DNA labelling kit from AP-Biotech (Freiburg, Germany). Ethidium bromide staining was used to verify gel loading.

For RT-PCR, 2 μg of total RNA were used. The first cDNA strand was synthesized with avian myeloblastosis virus (AMV) reverse transcriptase (Roche Molecular Biochemicals, Indianapolis, IN, USA) primed with random hexanucleotide primers. For amplification, 1/40 of the total cDNA was processed with the Expand High Fidelity PCR system (Roche Molecular Biochemicals) in the presence of primers no. 1 (exon 2) and no. 2 (exon 3), annealing at 62°C, 32 cycles. Control PCRs were realized with β-actin primer nos 3 and 4 (annealing at 62°C, 28 cycles). *Otx2* and *β-actin* PCR fragments are 561 and 146 bp long, respectively.

For RACE-PCR, we used the 5′/3′RACE Kit (Roche Molecular Biochemicals) according to the manufacturer's instructions. The PCR products were cloned into pGEM-T vector (Promega, Madison, WI, USA) and sequenced using the ABI PRISM™ Dye Terminator Sequencing Kit (Perkin Elmer, Wellesley, MA, USA)

on an automated DNA sequencer (ABI model 373; Applied Biosystems, Foster City, CA, USA).

For RNase protection assays, genomic Otx2 fragments were amplified, subcloned into pGEM-T and sequenced. The positions of probes relative to A of the ATG initiation codon numbered + 1 were: A, − 5047 to − 4918; B, − 2638 to − 2516; C, − 325 to − 119; D, 780–1399 and E, 2460–3620. The genomic sequence of the Otx2 locus is available at the website http://www.ensembl.org. After linearization, [^{32}P]UTP-labelled RNA probes (1.5×10^5 dpm/ng) were generated by *in vitro* transcription using the SP6 or T7 riboprobe system (Promega) depending on the orientation of the fragment. Hybridization (carried out for 16 h with 1 ng of labelled probe and 10 μg of total RNA at 65°C) and RNase treatment were performed with the RPAII kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions. Products were resolved on 5% denaturing polyacrylamide gels that were dried and autoradiographed.

Plasmids

All sequences were obtained by RT-PCR as described above. For complete Otx2 coding sequences containing or not the 8 amino-acid insertion, primer nos 5 and 6 were used (annealing at 60°C, 30 cycles). The PCR products containing an inframe C-terminal M2 tag were cloned into BamH I and Xba I sites of pCDNA3 (Invitrogen).

For 3′-UTR sequences, primer nos 7 and 8 were used (annealing at 55°C, 30 cycles). The PCR products were added at the end of the tag using a unique Xba I site.

For 5′-UTR sequences, forward primer nos 9, 10 and 11, respectively, for A, B and C transcripts, and the unique reverse primer no. 12 were used (annealing at 60°C, 30 cycles). These PCR products were digested by BamH I (present in the 5′ primer overhang) and Nco I (present in the Otx2 coding sequence) and fused to the same sites of the pCDNA3 recombinant vector containing the coding sequence and 3′-UTR.

The transcriptional activity of the Otx2 proteins was assayed using the region − 66 to + 68 of the *IRBP* promoter (Fong and Fong 1999) obtained by genomic DNA PCR with primer nos 13 and 14 (annealing at 60°C, 30 cycles). This 134-bp fragment was cloned into the Spe I site of pSEAP2-Basic (Clontech, Palo Alto, CA, USA). The IRBP-secreted alkaline phosphatase (SEAP) construct was transfected in cells with an equimolar amount of the Otx2 expression vector.

RNA circularization

The protocol of RNA circularization was adapted from Mandl *et al.* (1991). Cytoplasmic RNA (5 μg) was treated for 1 h with 10 U of calf intestine phosphatase (New England Biolabs, Beverly, MA, USA), then incubated with 1 U of tobacco acid pyrophosphatase (Fermentas, Vilnius, Lithuania) and finally with 80 U of RNA ligase (Promega). Each enzyme was removed by phenol–chloroform extraction. All circularized RNA was submitted to RT with coding sequence specific primer no. 15 and to PCR with primer nos 16 and 17 (annealing at 55°C, 30 cycles). The PCR products were subcloned into the pGEM-T vector and sequenced.

Reporter gene assay

Promoter activity was tested after transient transfection using SEAP (Clontech) as a reporter gene. The vectors pSEAP2-Basic (promoter-less) and pSEAP2-Control (SV40 promoter) were used as negative and positive controls, respectively. The different fragments of each promoter were obtained by PCR on 5.6-kb (constructs A and B) or 2.5-kb (construct C) EcoR I genomic fragments subcloned from the λ*Otx2* phage (a gift of D. Acampora, London, UK) into the pSK+ plasmid (Stratagene, La Jolla, CA, USA). Construct A, corresponding to nucleotide (nt) positions − 7800 to − 4925, was obtained with *Otx2* primer no. 18 and pSK+-specific primer no. 19. Construct B, corresponding to − 7800 to − 2503, was obtained with *Otx2* primer nos 20 and 19. Constructs C-207 and C-639, corresponding to nt positions − 1905 to − 207 and − 1905 to − 639, were obtained with primer nos 21 and 22, respectively, together with primer no. 19 (annealing at 55°C, 30 cycles). These fragments were cloned in both orientations into Bgl II or EcoR I sites of the pSEAP2-basic vector.

For assays, cells were seeded in six-well plates and transfected with 1 μg of the SEAP constructs and 0.1 μg of cytomegalovirus (CMV)-β-gal using the calcium-phosphate coprecipitation procedure. Transfections were carried out in duplicate and each experiment was reproduced three times. After 60 h, supernatant fluids were removed and SEAP activity was determined as relative light units with the Great EscAPe™ SEAP Chemiluminescence Detection Kit (Clontech) in a luminoscan (Labsystem, Helsinki, Finland). For normalization, β-galactosidase activity was measured as described in Herbomel *et al.* (1984). Relative activity was calculated as the ratio of the normalized SEAP value : normalized pSEAP2-basic value.

In situ hybridization

The procedure used was described in Han *et al.* (2000). Frozen tissues were cut with a cryostat (14-μm thick sections), thaw mounted on superfrost slides and stored at − 80°C. Sections were fixed with 4% paraformaldehyde, acetylated and stored in 80% ethanol until hybridization. Sections were delipidated in ethanol and chloroform, prehybridized for 2–4 h at 37°C in a moist chamber and hybridized for 20 h at 37°C under parafilm coverslips. Probes were added in the hybridization mixture at a concentration of 0.55 nM (i.e. 15 fmol/section). Non-specific staining was assessed by hybridizing one in every four consecutive sections in the presence of 100-fold excess of unlabelled probe. Sections were rinsed twice for 5 min in 2× SSC (1× SSC is 0.15 M NaCl and 0.03 M sodium citrate) and three times for 15 min in 1× SSC at room temperature, once for 15 min in 1× SSC at melting temperature minus 20°C and once for 15 min in 1× SSC at room temperature. After a quick rinse in ice-cold water, the slides were dehydrated with an ethanol gradient, dried at room temperature and exposed to autoradiography films (Biomax; Kodak, Rochester, NY, USA) for 15 days. Oligonucleotides used as probes were:

(i) A-specific oligonucleotide, 5′-TGAGAGCGGAACCTTCCTCAGCTCCAACTCAGCCTCCACTGTTAC;

(ii) B-specific oligonucleotide, 5′-GGAGGAGAGAGCAGTCGGGGAGTCGCCGCGCTATCGCTATTT;

(iii) C-specific oligonucleotide, 5′-TAGTTAGTGCTGGAACGTGGAGGAAGCTGCTCCCTCCGAAGCAGT and

(iv) common oligonucleotide, 5′-AATTTGGGCCGACTTTGCGCCTCCAAACAACCTTAGCATGATGTC.

Protein extraction and western blot analysis

Cells were pelleted, homogenized and lysed by a 30-min incubation in ice-cold lysis buffer (30 mM Tris HCl, pH 7.5, 40 mM NaCl and 0.5% NP40) in the presence of protease inhibitors (Complete; Roche

Molecular Biochemicals). Lysates were sonicated on ice and the protein concentration was determined using the protein assay kit (Bio-Rad Laboratories, Hercules, CA, USA). Samples were boiled in 50 mM Tris-HCl, pH 6.8, 5% 2-mercaptoethanol, 10% glycerol, 1% SDS and 0.001% bromophenol blue. Equal amounts (50 μg) of total proteins were separated by 10% SDS-polyacrylamide gel electrophoresis and transferred onto nitrocellulose membranes (Schleicher and Schuell, Dassel, Germany) in Tris-glycine buffer containing 20% methanol. Membranes were blocked with Tris-buffered saline (100 mM Tris HCl, pH 7.5, 0.9% NaCl and 0.1% Tween-20) containing 5% skimmed milk and incubated overnight at 4°C with the primary mouse antibody against M2 flag (1/1000; Sigma, St Louis, MO, USA) or for 1 h at room temperature with the primary antibody against β-galactosidase (1/800; Biogenesis, Poole, Dorset, UK). Membranes were washed and incubated for 1 h with peroxidase-conjugated secondary antibody (AP-Biotech). After several washes in Tris-buffered saline, membranes were incubated with enhanced chemoluminescence reagent (AP-Biotech).

Results

Cell and tissue expression of the *Otx2* gene

Features of the *Otx2* gene transcript and of its cDNA are poorly documented and, except for embryonic stages, the expression pattern of the gene remains to be established (Simeone *et al*. 1993; Frantz *et al*. 1994). Moreover, structural features of the gene outside the coding sequence lack detailed description. For instance, the precise location of the transcription start site remains elusive. To gain better insight into the functional organization of the gene, we first screened, by RT-PCR, a series of candidate mouse tissues, primary fibroblasts and cell lines (Fig. 1a). Remarkably, only a few tissues derived from the nervous system revealed the presence of the *Otx2* transcript. The highest levels were found in the neural retina, cerebellum and pons and a weak expression was consistently found in the hypothalamus and hippocampus. Among the cell lines tested, only the P19 embryonic carcinoma cells and normal embryonic stem cells (ES) sustained *Otx2* expression. Northern blot experiments confirmed this observation: a transcript of 2.3 kb was detected in the P19 embryonic carcinoma cell line as well as in mouse retina (not shown) and in both non-differentiated and differentiated ES cells (Fig. 1b). In contrast, primary and immortalized 3T3 fibroblasts yielded no signal as previously reported (Nguyen Ba-Charvet *et al*. 1998). As differentiated ES cells contain much higher *Otx2* mRNA levels than their cycling precursors, we examined, by northern blot, the time-course of *Otx2* transcript accumulation during the differentiation process (Fig. 1b). *Otx2* mRNA was barely detectable at 0 time. Its level is increased 5–10 times within the first 6 h of induction of differentiation and RNA accumulation peaks between 36 and 72 h when embryoid bodies are formed. The large increase in Otx2 mRNA level following induction of differentiation strongly suggests that *Otx2* expression is

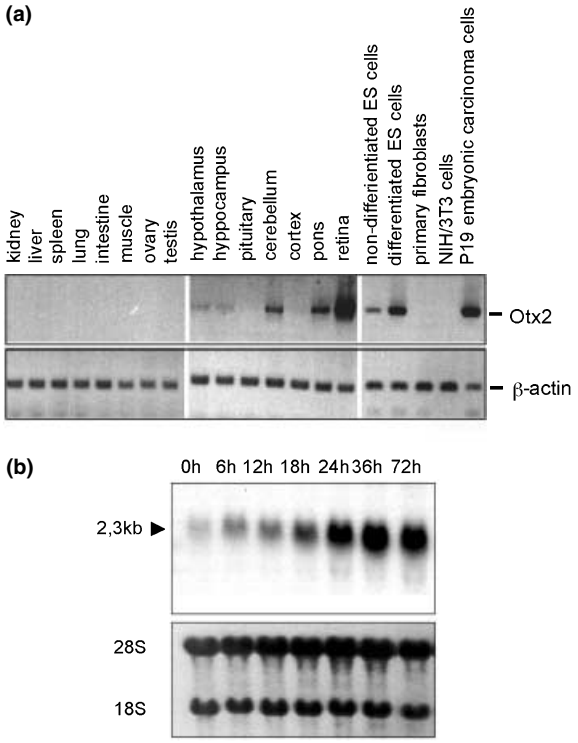

Fig. 1 *Otx2* cell- and tissue-specific expression. (a) RT-PCR analysis of the tissue distribution of *Otx2* mRNA. RT-PCR using *Otx2* (upper panel) and *β-actin* (lower panel) primers was performed as described in Experimental procedures on RNA from the indicated mouse tissues and cell lines. The lower panel shows that approximately equal amounts of RNA were used. (b) Northern blot analysis of *Otx2* regulation during cell differentiation. *Otx2* mRNA accumulation was monitored from RNA lots extracted from non-differentiated embryonic stem cells (0 h) or from cells induced to differentiate for the indicated times. The probe corresponds to the *Otx2* coding sequence. The size of *Otx2* mRNA is indicated. Ethidium bromide staining is shown below as a control of RNA loading. The position of 28S and 18S ribosomal RNAs is indicated. ES, embryonic stem cells.

required for this process. Therefore, both ES cells and retinal cells accumulate sufficient transcripts to allow the determination of the transcription start site of the *Otx2* gene.

Otx2 gene encodes multiple transcripts

In order to map precisely the 5′ end of the *Otx2* transcript, series of 5′ RACE-PCR experiments were carried out using total retina mRNA and oligonucleotide primers binding to the *Otx2* coding region. Unexpectedly, three types of *Otx2* cDNA products with different 5′ sequences were obtained (Fig. 2). They were named types A–C, respectively. In addition to a common coding sequence (Fig. 2a, light grey) the three cDNAs display unique 5′ sequences corresponding to specific UTRs starting 119 nt upstream of the translation ATG codon. Only the C type molecules have been reported previously (Kimura *et al*. 1997) (Fig. 2a, dark grey), while

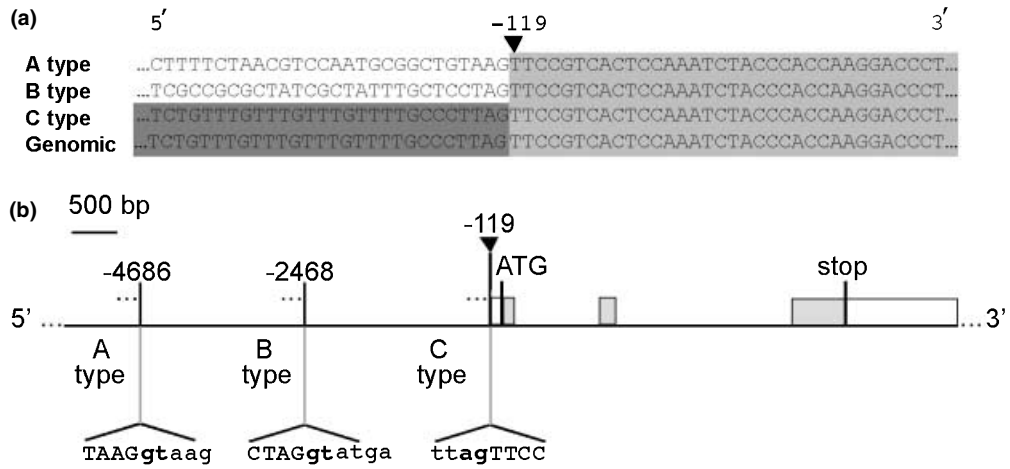

Fig. 2 Identification of three *Otx2* mRNA that differ by their 5′-untranslated region. (a) Partial nucleotide (nt) sequence of the three A, B and C *Otx2* cDNA species found by RACE-PCR. The shared sequence downstream of nt position 119 (arrowhead) is shown in light grey. The dark grey box emphasizes that C type cDNA is colinear to the genomic sequence (bottom line). (b) Genomic location of the new exons. The Otx2 coding sequence is represented by light grey boxes with start (ATG) and stop translation codons indicated. The 3′ end nt position, relative to the ATG start codon, of exons 1A, 1B and 1C is indicated together with the sequence overlapping the exon–intron splice sites. The consensus motifs are shown in bold. The nt sequence upstream of each A, B or C splice site is reported in (a). The arrowhead at − 119 shows the beginning of the common sequence for all A, B and C species.

the A and B type UTR corresponded to new species never described before. When aligned with a mouse genomic fragment, the A and B transcript extremities map to remote regions ending, respectively, at nt positions − 4686 and − 2468, with A of the ATG translation start codon taken as +1. The genomic sequence around these positions and position − 119 displays typical features of 5′ and 3′ intronic splice junctions, including a pyrimidine-rich tract and the consensus GT/AG sequence (Fig. 2b). This suggests that A and B transcripts arise from transcription and splicing of different exons bearing a common 3′ splice junction at position − 119. The 5′ splice site of the alternative 5′ exons is at position − 4686 for transcript A and at position − 2468 for transcript B. The 5′ sequence of transcript C, upstream of position − 119, is contiguous to the exon containing the translation start site. In contrast to the variability found at the 5′ ends, RACE-PCR and mRNA circularization (see below) experiments identified a unique 3′-UTR in all of the cell lines or tissues tested, corresponding to that already reported in the literature (Boyl *et al*. 2001).

Mapping of the transcription start sites

The above findings made it clear that transcription of the *Otx2* gene could be initiated at three alternative first exons. This prompted us to determine the limits of new exons of the A, B and C transcripts. The RACE products displayed 5′ end heterogeneity with maximal extensions reaching nt positions − 5010, − 2580 and − 211 for A, B and C types, respectively. To better map the different initiation sites, we performed RNase protection assays using genomic fragments covering the above positions (Fig. 3). Probe A, which contained 130 nt upstream position − 4918, yielded multiple protections ranging from 87 to 103 nt in P19 cells and adult retina. A weak signal could be detected in differentiated ES cells (Fig. 3a). The two strong 89 and 99 nt signals (arrowheads) correspond to initiation at positions − 5007 and − 5017. Probe B, which contained 123 nt upstream position − 2516, yielded multiple protections ranging from 76 to 98 nt only in P19 cells and adult retina (Fig. 3b). The major protections of 82, 84 and 95 nt (arrowheads) indicate initiation at positions − 2598, − 2600 and − 2611. Differentiated ES cells again showed a faint signal. Probe C, which contained 207 nt upstream position − 119, protected intense 88–93 nt fragments and weaker bands of 103–110 nt in P19 cells and differentiated ES cells. Non-differentiated ES cells showed modest signals of the same sizes, whereas only a weak band of 110 nt was detected in the retina (Fig. 3c). The strong 88 and 93 nt signals (arrowheads) indicated initiation at positions − 207 and − 212. These multiple protections with probes A, B and C must reflect real 5′ heterogeneity since probes encompassing exon 2 (nt 780–1399) and exon 3 (nt 2460–3620) displayed sharp protections at the expected size (152 and 377 nt, respectively, Figs 3d and e). Noticeably, a second protected fragment of 176 nt was reproducibly observed with the exon 2 probe, raising the possibility of a minor alternative splicing 24 nt upstream of the previously reported position (see below). From these experiments, we conclude that transcription of A and B type mRNA occurs preferentially in P19 cells and adult retina and starts at a cluster of positions located, respectively, between nt − 5017 and − 5003 and between nt − 2611 and − 2598, whereas C transcript is mainly restricted to embryonic P19 and ES cells and initiates between nt positions − 212 and − 207.

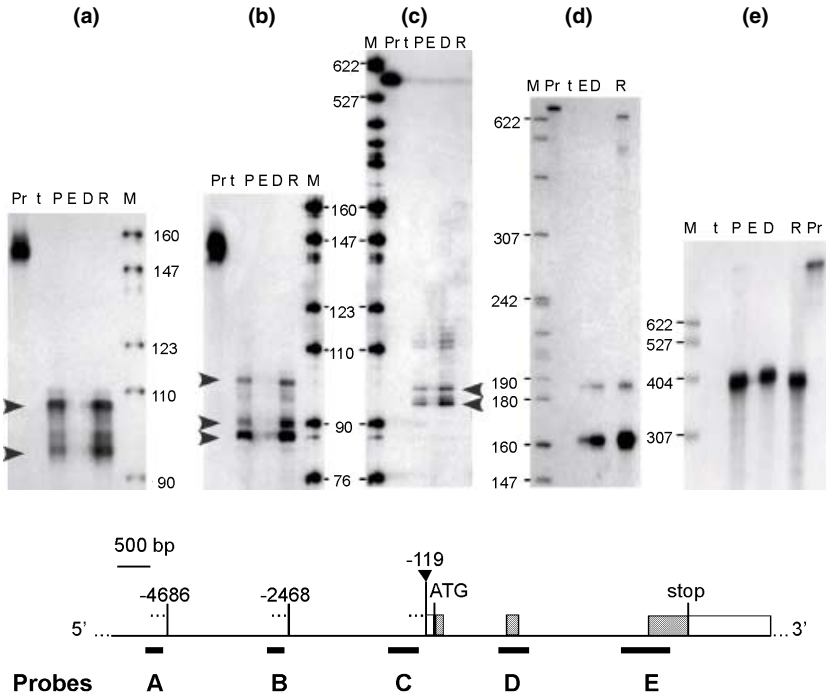

Fig. 3 Structural analysis of the different *Otx2* transcripts by RNase protection assay. (a–e) RNA from the indicated sources was hybridized with the corresponding probes. The probe localization is shown below the *Otx2* gene represented as in Fig. 2. Pr, untreated probe. Samples: t, tRNA used as a control for total digestion of the probe; P, P19 cells; E, non-differentiated embryonic stem cells (ES); D, differentiated ES cells and R, adult retina. M, ^{32}P end-labelled Msp I digest of pBR322. The size of the DNA fragments is indicated in nucleotides. In these conditions (5% denaturing polyacrylamide gel), single-stranded DNA migrates about 1.06 times faster than RNA. The apparent size of protected fragments was corrected by this factor.

In order to confirm these results we used another approach, based on RNA circularization assays. This method relies on ligation, with T4 RNA ligase, of the 5′ and 3′ ends of a single mRNA molecule. RT-PCR junction amplification allows the cloning and the determination of nts present at the very ends of the molecule. Selection of full length 5′-UTR is ensured by a phosphatase treatment which eliminates the 5′-phosphate group of incomplete molecules. After removal of the cap by tobacco acid pyrophosphatase, the resulting 5′-phosphoryl end can be joined to the 3′OH end by RNA ligase to give circular covalently joined RNA molecules. The PCR primers, mapping to a common region on all transcripts, allow single step and simultaneous amplification of the different 5′–3′ junctions. This reduces possible bias resulting from various PCR efficiencies. Thus, this assay provides a comparative estimation of each transcript level. The nt sequence of the 5′ end of a large number of PCR products, using circularized lots of RNAs from ES cells or mouse retina, was determined (Fig. 4). Figure 4(a, b and c) shows the relative abundance of the A, B and C transcripts initiated at the indicated nt positions. The major transcription start site of type A transcript is a G at position −5017 while B and C transcripts start, respectively, at a G at position −2598 and a C at position −207. These results are in good agreement with those obtained by RNase protection assays (Fig. 3). *In silico* examination, with the TESS database resource (http://www.cbil.upenn.edu/tess/), of each genomic DNA sequence upstream of these transcription start sites showed that A and B promoter regions contained numerous SP1 sites but were devoid of TATA and CAAT boxes. These two promoters thus fall in the category of GC-rich/TATA-less promoters. In contrast, characteristic TATA and CAAT boxes were found in the C promoter (data not shown).

The three different 5′ exons present in the A, B and C type transcripts were subsequently named exons 1A, 1B and 1C (Fig. 4d). For convenience, the part of the formerly so-called exon 1, which is common to the three A, B and C types of mRNAs (nt −119 to +97), was still referred to as exon 1. During the course of this work we also discovered that the 5′-UTR region of type A mRNA is transcribed from two small exons separated by a 155-bp long intronic sequence as shown in Fig. 4(d). The 5′ and 3′ splice sites of these two exons follow the canonical GT/AG rule (data not shown). They have been named exons 1A1 and 1A2. In addition to the complexity and heterogeneity of the 5′-UTR, sequencing of PCR products confirmed the presence of an alternative splice acceptor site in the coding sequence, 24 bp upstream of that already reported of the coding exon 2 (Fig. 8a). The added sequence encodes the octapeptide GPWASCPA, inserted 5 amino acids upstream

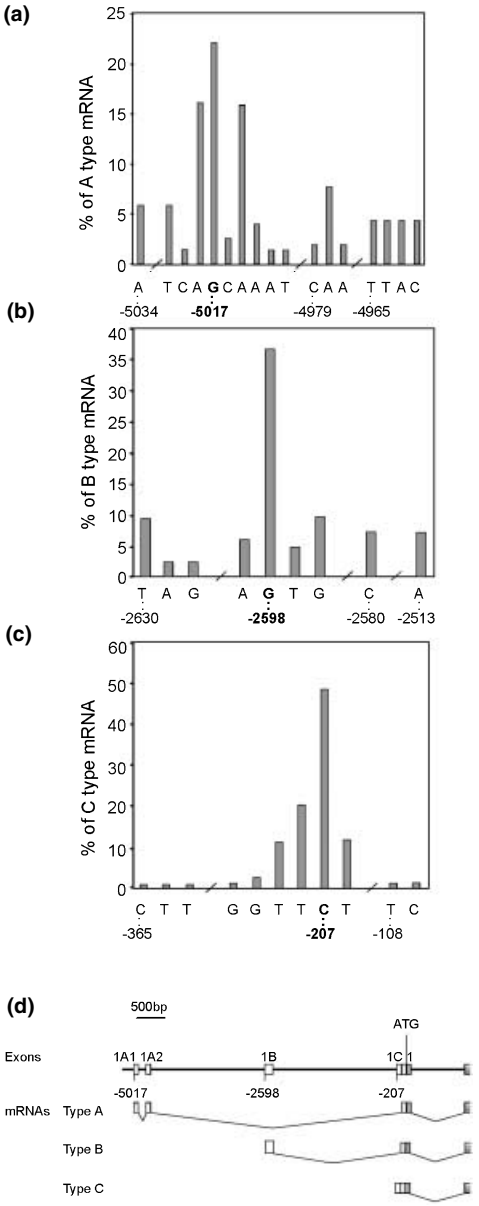

Fig. 4 Statistical mapping of the three types of *Otx2* mRNA transcription start sites. (a–c) Frequency diagrams of the first nucleotide (nt) positions transcribed, respectively, in type A, B and C mRNA. The abscissa shows the nt sequence and nt position numbers relative to the ATG start codon. The *Y* axis corresponds to the relative percentage of each type of mRNA. The nt position of the predominant transcription initiation sites is indicated in bold. (d) Deduced new structure of the *Otx2* gene. Genomic organization of the region of the *Otx2* gene transcription initiation is shown at the top. Boxes are exons. Hatched and grey-filled boxes are exon 1 non-coding and coding regions, respectively. Numbers indicate the nt position of the major transcription start sites. The splice scheme of the 5′ region of the three A, B and C type transcripts is presented underneath. The scale bar appears above.

of the homeodomain. From hereon, the new protein bearing this insertion will be referred to as $Otx2_L$.

The *Otx2* locus has three independent promoters

To further demonstrate that each of the above *Otx2* transcripts is driven by specific regulatory promoter sequences, the DNA regions upstream of each transcription start site A, B and C were fused to a SEAP reporter gene. Both orientations of each DNA fragment were tested using transient transfection assays (Fig. 5). The transcriptional activity of the A promoter (− 7800 to − 4925) alone or in combination with the B promoter (− 2503 to − 7800) and of the C promoter (−207 to −1905) region was analysed in the *Otx2*-expressing cell line P19 and in the *Otx2* transcription-deficient 3T3 fibroblast cell line. Sense orientation of the A, B and C constructs stimulated *SEAP* expression, in P19 cells, on average 38, 11 and 111 times as compared with the activity of a promoter-less *SEAP* vector (Fig. 5). No activity was detected when each promoter was inserted in the reverse orientation. This confirms that each fragment does contain a functional orientation-dependent transcription start site. The presence of exon 1A in the B construct did not significantly change B promoter activity. Constructs without exon 1A displayed the same SEAP activity as the full length construct (data not shown).

Interestingly, the activity of the different promoters differs between P19 and 3T3 cell lines. Only the TATA box containing C promoter has the specificity expected for cells that normally express *Otx2*. In contrast, promoters A and B display significant activity in cells that do not express *Otx2* (Fig. 5), suggesting that tissue-specific repressor elements are absent from the DNA fragments used in this study. In addition, 3′ deletion of the proximal C promoter region from nt − 207 to – 639 completely abolished its activity (Fig. 5), indicating that the transcriptional start site of the C transcript has been removed. Published reports located this site at nt positions − 1008 (Simeone *et al*. 1995) or − 688 (Guazzi *et al*. 1998). Both our structural (Fig. 3c) and functional data show that it definitely resides closer to the ATG translation start site than originally thought.

Expression domains of *Otx2* transcripts

The predominance of C transcript in ES cells (Fig. 3), which can be assimilated to early embryonic cells, as opposed to the abundant A and B mRNA in adult retina, a structure composed of well-differentiated cells, suggested that the *Otx2* promoter choice could be developmentally regulated. To test this, we first scored all of the independent clones obtained by circularization of mRNAs from non-differentiated and differentiated ES cells as well as newborn or adult retina. In non-differentiated ES cells, the C transcript represented 100% of *Otx2* mRNA. In differentiated ES cells, its proportion decreased to 75% as A (20%) and B (5%) transcripts appeared. Similarly, the percentages of A, B and C mRNAs, respectively, moved from 19, 50 and 31% in the

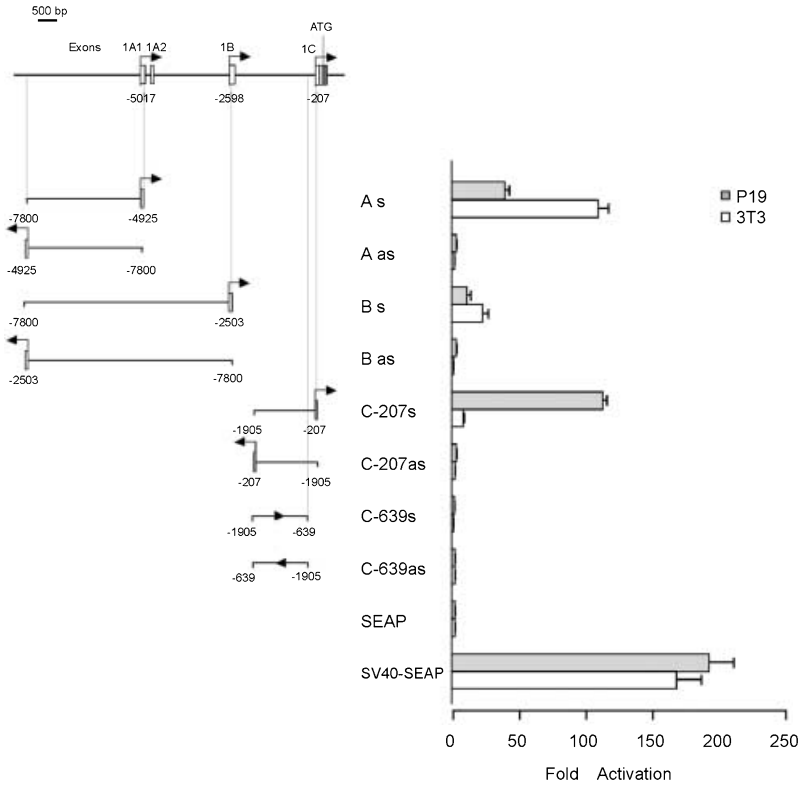

Fig. 5 Activity of the three *Otx2* gene promoters. Mouse P19 embryonic carcinoma cells and fibroblast 3T3 cells were transiently transfected by plasmids containing the indicated (A, B or C) promoter regions fused either in sense (s) or antisense (as) orientation to the secreted alkaline phosphatase (SEAP) reporter gene. The position of the promoter fragments is aligned to the gene map (top) and the nucleotide position of their ends is indicated. The arrowhead indicates the transcription direction. Boxes are exons and arrows with corresponding numbers indicate the position of the main transcription start sites. ATG is the translation start codon. Each transfection was normalized using a cotransfected *LacZ* gene. Phosphatase activity is presented relative to pSEAP2-basic (SEAP). The SV40-SEAP is used as a positive control. Error bars are SD calculated from three independent experiments performed in duplicate.

newborn retina to 46, 31 and 13% in the adult retina. Both upon ES differentiation and during maturation of the retina, we observed a similar increase in A transcript at the expense of C transcript. These temporal changes suggest that C type could be the early embryonic form of *Otx2* mRNA and that A and perhaps B type could take over later in development.

To further substantiate these findings, we carried out *in situ* hybridization experiments with radioactive oligonucleotides as probes in order to ensure specificity and sensitivity. Hybridizations were performed on adult mouse brain coronal sections and on newborn and adult retina sections. Probes mapping in the A, B and C 5′-UTRs and in the common exon 1 sequence were used. As shown in Fig. 6(a–c), all three mRNA types are expressed in ependymal cells of the brain ventricles (arrowheads). Although the signal intensity of the different probes cannot be compared with each other, expression in ependymal cells provides an internal control for expression in the medial habenula (arrows): in this structure, A appears high, B is moderate and C is low. Therefore, the ratio of each *Otx2* mRNA between the habenula and ependymal cells varies. In the newborn retina (Figs 6e–h), the A and C mRNA are detected throughout the as yet undifferentiated single neuronal layer but the B mRNA is mostly located in the outer part of the organ. In the adult retina (Figs 6i–l), the amount of transcripts A and B is increased and these mRNAs are located in both the inner and outer nuclear layers. On the contrary, the C mRNA is less abundant and mostly restricted to the inner nuclear layer. All of these variations correlate with the RNA circularization semiquantitative analysis and argue for strict spatial and temporal control of each *Otx2* promoter.

The three *Otx2* mRNAs are translated *in vivo*

The existence of different 5′ non-coding regions in the *Otx2* mRNA isoforms opens up the possibility of a differential translational regulation. In order to explore this hypothesis, the various cDNAs encoding flagged Otx2 or $Otx2_L$ proteins were inserted in a CMV expression vector. These constructs were cotransfected into 293T cells (which do not express *Otx2*) with a construct containing the Otx2-responsive *IRBP* promoter (Fong and Fong 1999) fused to the SEAP reporter gene and a CMV-β-galactosidase expression vector. Otx2

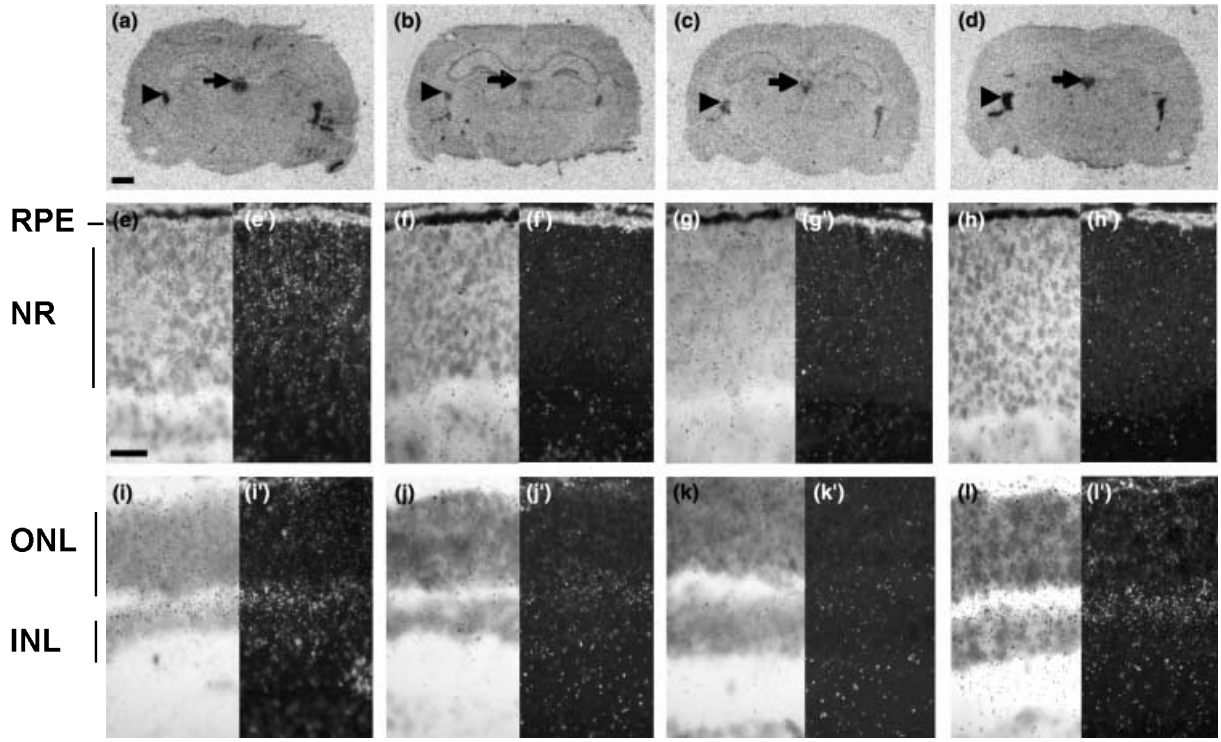

Fig. 6 Expression of the different *Otx2* mRNAs in the mouse brain and retina. In the top panels, bright field photographs of autoradiograms show the distribution of types A (a), B (b), C (c) and global (d) *Otx2* mRNAs in adjacent coronal adult mouse brain sections. Arrows point to the medial habenula and arrowheads point to the plexus choroid of the lateral ventricles. In the two lower panels, bright field (e–l) and corresponding dark field (e′–l′) photographs of emulsions show the distribution of type A (e and i), B (f and j), C (g and k) and global (h and l) *Otx2* mRNAs in adjacent newborn (e–h) and adult (j–l) mouse retina sections. INL, inner nuclear layer; NR, neural retina; ONL, outer nuclear layer; RPE, retinal pigmented epithelium.

protein accumulation was monitored both by western blot analysis and by measurement of its transcriptional activity (Fig. 7). In the absence of the 5′-UTR sequence, both proteins Otx2 and $Otx2_L$ are expressed to similar levels (Fig. 7, lanes 3 and 4). In the presence of A and C 5′-UTRs, after calibration for a constant amount of β-galactosidase, similar amounts of both proteins are detected (lanes 5, 6, 9 and 10), resulting in comparable transcriptional activation of the *IRBP* promoter (lower panel). However, the B 5′-UTR leads to increased amounts of the Otx2 proteins and simultaneously to stronger stimulation of the IRBP promoter (Fig. 7 lanes 7, 8 and lower panel). In all cases, the presence of the octapeptide in $Otx2_L$ does not dramatically influence the transcriptional properties of the protein. From this experiment, we concluded that translation can be achieved whatever the 5′-UTR is, although with variations, the B 5′-UTR conferring a better translation efficiency than the A and C 5′-UTRs. In addition, both Otx2 and $Otx2_L$ proteins activate transcription to similar levels.

Otx2 genomic organization

In addition to the six mRNAs that encode full length Otx2 and $Otx2_L$ proteins, an additional transcript expressed in P19 cells was recently deposited in the GenBank expressed sequence tag (EST) database. It fuses 174 nt from the intron that separates exons 1 and 2 (nt positions 607–780) to the second and third regular exons of the *Otx2* gene. Nucleotides around position 780 perfectly match a canonical 5′ splice sequence. RT-PCR analysis, with a specific forward primer, indicates that this alternative transcript is present in ES cells, although in low quantities, and is up-regulated simultaneously with ES cell differentiation (data not shown). This new transcript was named D. Figure 8(a) recapitulates the structural organization of the whole diversity of mRNA species discussed in this paper.

The discovery of these new *Otx2* gene features prompted us to see whether they are also conserved in the human genome. Nucleotide homology between 11 kb of mouse and human *Otx2* genomic sequence was scored (Fig. 8b) using a Blast analysis (Tatusova and Madden 1999). Sequences similar to the mouse transcription start sites for exons 1A, 1B and 1C are found, respectively, 5021, 2384 and 208 nt upstream of the translation start codon in the human *Otx2* locus. Mouse and human *Otx2* 5′-UTRs share 84% identity for type A, 58.5% for type B and 91% for type C. Despite the lower homology between human and mouse B 5′-UTR, a region of 1000 bp upstream of the

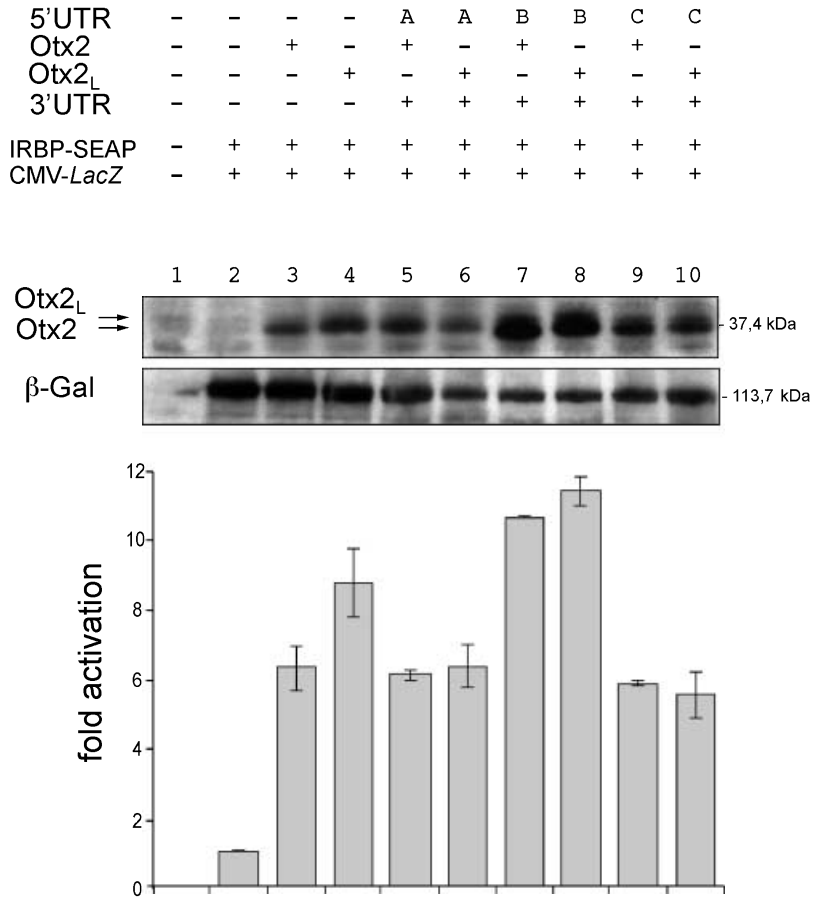

Fig. 7 Functional translation of the different *Otx2* mRNA in human 293T cells. Cytomegalovirus (CMV) expression vectors containing the indicated *Otx2* mRNA portions were transfected along with an *IRBP*-secreted alkaline phosphatase (SEAP) reporter and CMV-*LacZ* plasmids in the 293T cell line which does not express *Otx2*. Protein extracts were analysed by western blot for the presence of Otx2, $Otx2_L$ and β-galactosidase (β-gal). Lanes 1–10 correspond to the experimental conditions depicted above. The corresponding SEAP activity was measured and expressed as fold activation relative to *IRBP*-SEAP (lower graph). The position of the detected proteins and of molecular weight markers is indicated. UTR, untranslated region.

transcription start site retains 88.8% of nt identity. Thus, in both human and mouse, the main transcription start sites and the surrounding sequences are conserved. Similarly, the common sequence upstream of the ATG codon (84.8% nt identity), the exon 1D region and the 24-nt alternative exon 2 are highly conserved. Strong nt identity is also observed outside exons, especially in the 5′ flanking regions of each first exon, pointing to a conservation of the regulatory sequences in proximal promoters for each transcript isoform. Thus, the new mouse *Otx2* gene organization that we describe is conserved in human.

Discussion

Structural features of the *Otx2* gene

The present work provides a new understanding of the genomic organization of the *Otx2* gene (Fig. 8a). Instead of the single previously reported *Otx2* transcription start site, we have unveiled three short alternative non-coding first exons. The 1A (split into two close exons 1A1 and 1A2) and 1B new exons are located, respectively, 5 and 2.5 kb upstream of exon 1C and are controlled by GC-rich TATA-less promoters. Both splice to nt position − 119 which represents the first functional 3′ splice. Transcription of exon 1C begins 500–800 bp downstream of the previously described start positions and is likely to be driven by a regular TATA and CAAT box promoter. The 5′-UTR of the A, B and C mRNAs is, respectively, 176, 130 and 89 nt long, precluding their discrimination on electrophoresis gels (see Fig. 1b). All transcripts display multiple sites of initiation, a feature found in more and more genes. Similar situations have recently been described for *Pax6*, *c-myc* or *Lst1* genes (Okladnova *et al*. 1998; Futami *et al*. 2000; Yu and Weissman 2000). The complexity of the *Otx2* locus is increased by an alternative splice site in the coding region, which gives rise to a new

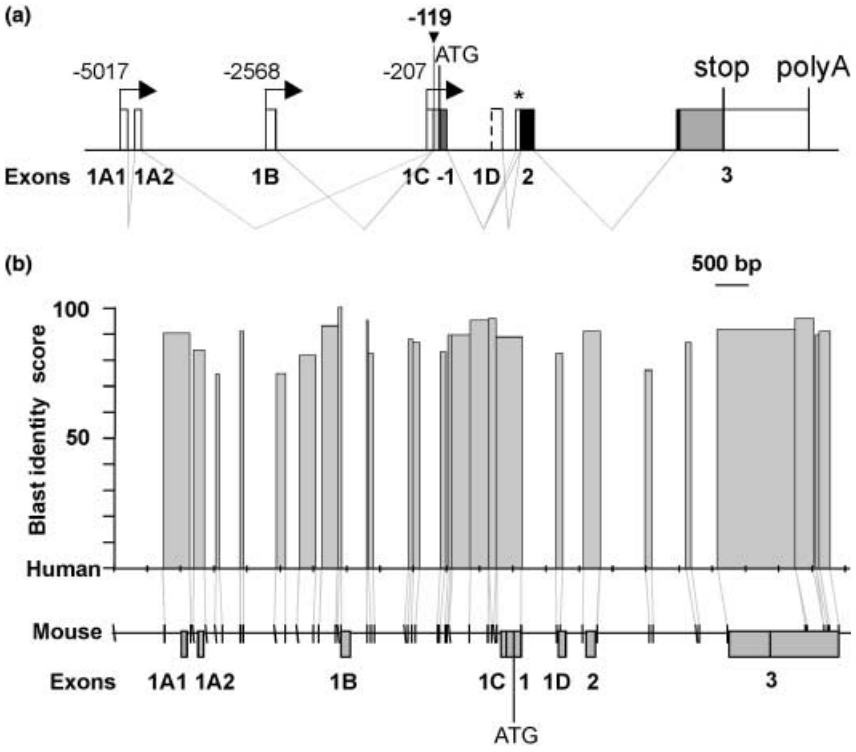

Fig. 8 A new conserved structure for the *Otx2* gene. (a) Schematic summarizing the exon–intron structure of the mouse *Otx2* gene. Exons are labelled and shown as boxes. Arrows symbolize the main transcription start sites. The 3′ splice site common to A and B transcript is shown by an arrowhead. Numbers indicate nucleotide positions relative to translation start codon (ATG). Translation termination (stop) and cleavage-polyadenylation (polyA) sites are shown. Otx2 coding regions are in grey and the homeobox is in black. The new exon 2 splice acceptor site present 24 nucleotides upstream of exon 2 is marked by an asterisk. (b) Nucleotide identity scoring of mouse and human *Otx2* genes. The Blast program (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bL2.html) was used to compare 11 kb of mouse and human *Otx2* sequences with the following parameters: word size = 11; match = 1; mismatch = – 2; gap open = 5; x-dropoff = 50. Only positive scores were retained. Position of the mouse exons is indicated. Lines align the position of each border of positive regions.

$Otx2_L$ protein species. In all tissues tested, the 24 nt longer isoform accounts for about 10% of the mRNAs, irrespective of their 5′-UTR. This suggests that constitutive alternative splicing, rather than UTR-dependent processing, is the main mechanism selecting for $Otx2_L$ synthesis. This yields a total of six *Otx2* mRNA isoforms, encoding two proteins. Despite the fact that transcripts encoding $Otx2_L$ have never been published, several cDNAs including the 24-nt insertion are referenced in the GenBank database (accession nos XM_127707, BC027104 and BC029667). Finally, a shorter *Otx2*-related mRNA, reported as an EST from P19 cells (database accession no. AA199520), splices a sequence derived from a putative additional exon 1D to exons 2 and 3. We found that it is coregulated with the other mRNAs in ES cells. Overall, the versatile gene organization presented here could provide flexible regulatory solutions to adapt *Otx2* expression to its diverse biological functions.

Functional significance of the *Otx2* mRNA diversity

The occurrence of multiple independent gene promoters, 5′-UTRs and proteins raises questions about their respective biological function. Numerous studies have correlated the use of alternative promoters to tissue-specific expression. For instance, Pendleton *et al*. (2002) have reported the usage of alternative 5′-UTRs in the *argininosuccinate synthase* gene to direct liver- or endothelial-specific expression. In the case of *Otx2*, we can expect a complex genetic control, given the very different and specific tasks which it deals with. During early development, it behaves successively as a determinant for cell movement and A/P axis polarity (Kimura *et al*. 2000), neuroectoderm induction and prevention of forebrain cell apoptosis (Rhinn *et al*. 1998, 1999). Its expression then becomes restricted to very specific areas, such as the tectum, ependymal cells or retina. The function that it exerts in these organs remains to be elucidated, but it is likely to regulate the transcription of specific sets of genes. Its multiple initiation sites provide a way of controlling protein production in different situations. Our data showing that the C type mRNA is the unique *Otx2* transcript present in ES cells and that the A type becomes predominant in adults support the idea of tissue- and/or time-course-specific usage of the different promoters. The fact that the relative abundance of the three

mRNA types changes during retina maturation adds weight to this hypothesis. Following this view, it will be very interesting to study the precise expression pattern of each type of mRNA throughout mouse development.

Potentially, differential 5′-UTR usage could confer specific subcellular locations (Thio *et al*. 2000) or peculiar translation properties. For instance, the leader region is involved in mRNA transport and ribosome assembly and, as such, has been recognized as the main level of translational regulation (van der Velden *et al*. 2002 and references therein). We have checked the translation hypothesis using transient expression of the six *Otx2* cDNA isoforms. Protein synthesis is effective with the three 5′-UTRs but the B type message appears more prone to translation. From this observation, one may suggest that the different translational capacity of transcripts helps to respond to tissue-specific properties of the translation machinery.

Alternative splicing within the coding region often gives rise to related polypeptides that exhibit different properties. In the present case, both Otx2 and $Otx2_L$ proteins stimulate the *IRBP* promoter with comparable magnitude. Although we cannot rule out different partnerships due to the presence of the octapeptide in $Otx2_L$, the two proteins appear, so far, to be functionally equivalent. However, if no other protein can be predicted (first exons are short and devoid of any open reading frame phased with that of the main Otx2) nor detected by expression of the A, B and C transcripts, we cannot exclude the occurrence of a new D mRNA. As exon D lies downstream of the Otx2 translation initiation codon, only N-terminal truncated Otx2 proteins would be expected from translation of the D mRNA, that might act as dominant negative forms of normal Otx2 versions. This point will be addressed in the future, using transient transfection of plasmids expressing various combinations of the cloned *Otx2* cDNAs. It is of note that, similar to the Otx2 situation, a shorter C-terminal form of the closely related protein Otx1 has been detected in mice bearing an *Otx1* allele disrupted at the exon 1 (Weimann *et al*. 1999).

Prospect on the consequence of a new gene organization

The general understanding of the structure and function of the *Otx2* gene, as deduced from previous studies, was re-examined in view of the present results. For instance, *Otx2* expression studies by mRNA *in situ* hybridization (Simeone *et al*. 1992), or by immunocytochemistry (Mallamaci *et al*. 1996), which are based on the global detection of all mRNA and proteins species are in full agreement with our conclusions. However, the present work adds new information and precision to the specific expression patterns of each isoform. It also sheds insight into the experiments that document the *in vivo* function of *Otx2*. *Otx2* loss of function was obtained by the knock-in of an *LacZ* gene in the so-called first exon (Acampora *et al*. 1995). For this purpose, the Sma I site, located 208 bp upstream of the translation start site, was used. This precisely fuses the reporter gene to the transcription start site of type C mRNA and simultaneously removes the possibility of splicing exons 1A and 1B. Therefore, the LacZ transcription is only driven by the promoter C. Similarly, in rescue experiments (Acampora *et al*. 2001), *Otd* or *Otx2* cDNAs only correct early developmental defects when fused to the same position. In contrast, the fusion of exogenous sequences to the ATG codon leads to spectacular improvement of the late development, provided that they are followed by the complete *Otx2* 3′-UTR sequence (Acampora *et al*. 2001; Boyl *et al*. 2001). Our findings suggest that transcription of the grafted sequences can be driven by the three A, B and C promoters (Fig. 8a). However, rescue of the *Otx2* mutation remains partial in a mouse line where the Otx1 coding sequence replaces *Otx2* and is driven by A, B and C promoters but lacks a full-length 3′-UTR (Suda *et al*. 1999). These differential temporal rescues might be explained by a model where promoter C operates during early embryogenesis while promoters A and B are required and dominant for late development. Altogether, these facts point to the importance of both 5′- and 3′-UTR to achieve optimal expression of the *Otx2* gene.

A paradigm of genomic organization and function

The unexpected complex genomic structure of the mouse *Otx2* locus raises the question of its general occurrence throughout evolution. The nt sequence identity in regions corresponding to mouse 5′-UTRs supports the possibility of a similar alternative splicing scheme in human. Alternative splicing is a strategy used to increase genetic diversity from a unique sequence and some authors estimate that 40–60% of genes are subjected to this process (Lander *et al*. 2001; Modrek and Lee 2002). The vast majority of known examples of alternative splicing events concern the coding region of mRNAs. For instance, immunoglobulin heavy chain alternative splicing provides a stop codon upstream of regions encoding a transmembrane domain, leading to secreted IgG antibodies rather than membrane-anchored IgM antigen receptors (Early *et al*. 1980). *Otx2* also displays protein diversification potential, but we have not been able to ascribe specific properties to the two protein isoforms, Otx2 and $Otx2_L$, using *in vitro* assays. However, reports of human transcripts without (Fong and Fong 1999) and with the 24-nt exon 2 insertion in the GenBank database (accession no. NM_021728) indicate that both protein isoforms are produced in humans. Moreover, evolutionary conservation of this process is also found in *Drosophila*, where about 10% of *otd* mRNAs follow the same alternative splicing scheme and result in the insertion of a hexapeptide GDLCYP upstream of the homeodomain (Vandendries *et al*. 1996).

Selectivity at the *Otx2* locus is first generated by the use of alternative promoters. We suggest that different sets of

cell-type regulatory factors exert independent control of their activities. The resulting mRNAs display additional differential translation properties, showing that their 5′-UTRs can integrate specific physiological signals. The combination of both mechanisms allows the extremely fine-tuned cell-specific production of Otx2 proteins which is required in the brain. Recent studies have provided a growing number of examples ascribing specific properties to the 5′-UTRs of mRNAs. For instance, the *neuronal nitric oxide synthase* gene has nine different transcription initiation exons, the use of which gives a unique protein (Wang *et al*. 1999). Alternative promoter usage allows tissue-specific expression and confers specific translation capacities in response to environmental changes. In the *bcl-X* gene, promoter choice plays a key role in determining alternative splicing which, in turn, influences the balance of pro- and antiapoptotic Bcl-X isoforms (Pecci *et al*. 2001). In this view, the comparison of translation of cloned *Otx2* cDNAs *in vivo* in areas where *Otx2* is normally expressed at various stages of development will be of particular interest.

In addition to the similarities between mouse and humans, data on the invertebrate *Otx2* gene also point to conservation during evolution. In the sea urchin *Strongylocentrotus purpuratus*, two *SpOtx* forms (α and β), that are entirely different in their N-terminal domain, are generated from a single gene (Li *et al*. 1997). *SpOtxα* mRNA is transcribed from a single start site. In contrast, *SpOtxβ* mRNA is transcribed from two separate initiation sites and undergoes alternative splicing of untranslated sequences that are associated with the same coding sequence. As in mouse, alternative promoter usage in sea urchin could regulate the expression of *Otx* from larval to adult stages. The existence of such conservation in the *Otx2* gene structure fits perfectly with the conservation of the Otx2 protein function observed from fly to mammals. It suggests that the complex architecture of coding and non-coding regions of this gene is instrumental for proper execution of developmental programmes that involve *Otx2* gene expression. This architecture would have been established before the divergence between invertebrates and vertebrates. Increasing genomic sequence information from new species shows special promise to identify the structural and regulatory motifs that support *Otx2* expression and function.

Acknowledgements

We are grateful to Dr D. Acampora for the gift of lambda *Otx2* phage. We thank Dominique Baas and Eloïse Hudry for constructs, Théophile Ohlmann and Evelyne Goillot for discussions and Evelyne Goillot and Guy Fuller for critical reading of the manuscript. This work was supported by grants from the CNRS, Retina France and The Association de la Recherche sur le Cancer. VC is a recipient of a fellowship of the French Ministry of Research and Education.

References

Acampora D., Mazan S., Lallemand Y., Avantaggiato V., Maury M., Simeone A. and Brulet P. (1995) Forebrain and midbrain regions are deleted in Otx2–/– mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development* **121**, 3279–3290.

Acampora D., Avantaggiato V., Tuorto F. and Simeone A. (1997) Genetic control of brain morphogenesis through Otx gene dosage requirement. *Development* **124**, 3639–3650.

Acampora D., Boyl P. P., Signore M., Martinez-Barbera J. P., Ilengo C., Puelles E., Annino A., Reichert H., Corte G. and Simeone A. (2001) OTD/OTX2 functional equivalence depends on 5′ and 3′ UTR-mediated control of Otx2 mRNA for nucleo-cytoplasmic export and epiblast-restricted translation. *Development* **128**, 4801–4813.

Ang S. L., Jin O., Rhinn M., Daigle N., Stevenson L. and Rossant J. (1996) A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. *Development* **122**, 243–252.

Baas D., Bumsted K. M., Martinez J. A., Vaccarino F. M., Wikler K. C. and Barnstable C. J. (2000) The subcellular localization of Otx2 is cell-type specific and developmentally regulated in the mouse retina. *Brain Res. Mol. Brain Res.* **78**, 26–37.

Boncinelli E. and Morgan R. (2001) Downstream of Otx2, or how to get a head. *Trends Genet.* **17**, 633–636.

Boyl P. P., Signore M., Acampora D., Martinez-Barbera J. P., Ilengo C., Annino A., Corte G. and Simeone A. (2001) Forebrain and midbrain development requires epiblast-restricted Otx2 translational control mediated by its 3′ UTR. *Development* **128**, 2989–3000.

Early P., Rogers J., Davis M., Calame K., Bond M., Wall R. and Hood L. (1980) Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* **20**, 313–319.

Finkelstein R. and Perrimon N. (1990) The orthodenticle gene is regulated by bicoid and torso and specifies *Drosophila* head development. *Nature* **346**, 485–488.

Finkelstein R., Smouse D., Capaci T. M., Spradling A. C. and Perrimon N. (1990) The orthodenticle gene encodes a novel homeo domain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures. *Genes Dev.* **4**, 1516–1527.

Fong S. L. and Fong W. B. (1999) Elements regulating the transcription of human interstitial retinoid-binding protein (IRBP) gene in cultured retinoblastoma cells. *Curr. Eye Res.* **18**, 283–291.

Frantz G. D., Weimann J. M., Levin M. E. and McConnell S. K. (1994) Otx1 and Otx2 define layers and regions in developing cerebral cortex and cerebellum. *J. Neurosci.* **14**, 5725–5740.

Futami K., Komiya T., Zhang H. and Okamoto N. (2000) Determination of heterogeneous transcription start points of two c-myc genes from the common carp (*Cyprinus carpio*). *Gene* **245**, 43–47.

Gauthier K., Chassande O., Plateroti M., Roux J. P., Legrand C., Pain B., Rousset B., Weiss R., Trouillas J. and Samarut J. (1999) Different functions for the thyroid hormone receptors TRalpha and TRbeta in the control of thyroid hormone production and post-natal development. *EMBO J.* **18**, 623–631.

Guazzi S., Pintonello M. L., Vigano A. and Boncinelli E. (1998) Regulatory interactions between the human HOXB1, HOXB2, and HOXB3 proteins and the upstream sequence of the Otx2 gene in embryonal carcinoma cells. *J. Biol. Chem.* **273**, 11092–11099.

Han Z. Y., Le Novere N., Zoli M., Hill J. A. Jr, Champtiaux N. and Changeux J. P. (2000) Localization of nAChR subunit mRNAs in the brain of *Macaca mulatta*. *Eur. J. Neurosci.* **12**, 3664–3674.

Herbomel P., Bourachot B. and Yaniv M. (1984) Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. *Cell* **39**, 653–662.

Hirth F., Therianos S., Loop T., Gehring W. J., Reichert H. and Furukubo-Tokunaga K. (1995) Developmental defects in brain segmentation caused by mutations of the homeobox genes orthodenticle and empty spiracles in *Drosophila*. *Neuron* **15**, 769–778.

Kimura C., Takeda N., Suzuki M., Oshimura M., Aizawa S. and Matsuo I. (1997) Cis-acting elements conserved between mouse and pufferfish Otx2 genes govern the expression in mesencephalic neural crest cells. *Development* **124**, 3929–3941.

Kimura C., Yoshinaga K., Tian E., Suzuki M., Aizawa S. and Matsuo I. (2000) Visceral endoderm mediates forebrain development by suppressing posteriorizing signals. *Dev. Biol.* **225**, 304–321.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Leuzinger S., Hirth F., Gerlich D., Acampora D., Simeone A., Gehring W. J., Finkelstein R., Furukubo-Tokunaga K. and Reichert H. (1998) Equivalence of the fly orthodenticle gene and the human OTX genes in embryonic brain development of *Drosophila*. *Development* **125**, 1703–1710.

Li X., Chuang C. K., Mao C. A., Angerer L. M. and Klein W. H. (1997) Two Otx proteins generated from multiple transcripts of a single gene in *Strongylocentrotus purpuratus*. *Dev. Biol.* **187**, 253–266.

Makiyama Y., Shoji S. and Mizusawa H. (1997) Hydrocephalus in the Otx2+/– mutant mouse. *Exp. Neurol.* **148**, 215–221.

Mallamaci A., Di Blas E., Briata P., Boncinelli E. and Corte G. (1996) OTX2 homeoprotein in the developing central nervous system and migratory cells of the olfactory area. *Mech. Dev.* **58**, 165–178.

Mandl C. W., Heinz F. X., Puchhammer-Stockl E. and Kunz C. (1991) Sequencing the termini of capped viral RNA by 5′-3′ ligation and PCR. *Biotechniques* **10**, 486.

Martinez-Morales J. R., Signore M., Acampora D., Simeone A. and Bovolenta P. (2001) Otx genes are required for tissue specification in the developing eye. *Development* **128**, 2019–2030.

Matsuo I., Kuratani S., Kimura C., Takeda N. and Aizawa S. (1995) Mouse Otx2 functions in the formation and patterning of rostral head. *Genes Dev.* **9**, 2646–2658.

Modrek B. and Lee C. (2002) A genomic view of alternative splicing. *Nat. Genet.* **30**, 13–19.

Nagao T., Leuzinger S., Acampora D., Simeone A., Finkelstein R., Reichert H. and Furukubo-Tokunaga K. (1998) Developmental rescue of *Drosophila* cephalic defects by the human Otx genes. *Proc. Natl Acad. Sci. USA* **95**, 3737–3742.

Nguyen Ba-Charvet K. T., von Boxberg Y., Guazzi S., Boncinelli E. and Godement P. (1998) A potential role for the OTX2 homeoprotein in creating early 'highways' for axon extension in the rostral brain. *Development* **125**, 4273–4282.

Okladnova O., Syagailo Y. V., Mossner R., Riederer P. and Lesch K. P. (1998) Regulation of PAX-6 gene transcription: alternate promoter usage in human brain. *Brain Res. Mol. Brain Res.* **60**, 177–192.

Pecci A., Viegas L. R., Baranao J. L. and Beato M. (2001) Promoter choice influences alternative splicing and determines the balance of isoforms expressed from the mouse bcl-X gene. *J. Biol. Chem.* **276**, 21062–21069.

Pendleton L. C., Goodwin B. L., Flam B. R., Solomonson L. P. and Eichler D. C. (2002) Endothelial argininosuccinate synthase mRNA 5′-untranslated region diversity. *J. Biol. Chem.* **277**, 25363–25369.

Rhinn M., Dierich A., Shawlot W., Behringer R. R., Le Meur M. and Ang S. L. (1998) Sequential roles for Otx2 in visceral endoderm and neuroectoderm for forebrain and midbrain induction and specification. *Development* **125**, 845–856.

Rhinn M., Dierich A., Le Meur M. and Ang S. (1999) Cell autonomous and non-cell autonomous functions of Otx2 in patterning the rostral brain. *Development* **126**, 4295–4304.

Royet J. and Finkelstein R. (1995) Pattern formation in *Drosophila* head development: the role of the orthodenticle homeobox gene. *Development* **121**, 3561–3572.

Simeone A., Acampora D., Gulisano M., Stornaiuolo A. and Boncinelli E. (1992) Nested expression domains of four homeobox genes in developing rostral brain. *Nature* **358**, 687–690.

Simeone A., Acampora D., Mallamaci A., Stornaiuolo A., D'Apice M. R., Nigro V. and Boncinelli E. (1993) A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo. *EMBO J.* **12**, 2735–2747.

Simeone A., Avantaggiato V., Moroni M. C., Mavilio F., Arra C., Cotelli F., Nigro V. and Acampora D. (1995) Retinoic acid induces stage-specific antero-posterior transformation of rostral central nervous system. *Mech. Dev.* **51**, 83–98.

Suda Y., Nakabayashi J., Matsuo I. and Aizawa S. (1999) Functional equivalency between Otx2 and Otx1 in development of the rostral head. *Development* **126**, 743–757.

Tatusova T. A. and Madden T. L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**, 247–250.

Thio G. L., Ray R. P., Barcelo G. and Schupbach T. (2000) Localization of gurken RNA in *Drosophila* oogenesis requires elements in the 5′-and 3′ regions of the transcript. *Dev. Biol.* **221**, 435–446.

Vandendries E. R., Johnson D. and Reinke R. (1996) orthodenticle is required for photoreceptor cell development in the *Drosophila* eye. *Dev. Biol.* **173**, 243–255.

van der Velden A. W., van Nierop K., Voorma H. O. and Thomas A. A. (2002) Ribosomal scanning on the highly structured insulin-like growth factor II-leader 1. *Int. J. Biochem. Cell Biol.* **34**, 286–297.

Wang Y., Newton D. C. and Marsden P. A. (1999) Neuronal NOS: gene structure, mRNA diversity, and functional relevance. *Crit. Rev. Neurobiol.* **13**, 21–43.

Weimann J. M., Zhang Y. A., Levin M. E., Devine W. P., Brulet P. and McConnell S. K. (1999) Cortical neurons require Otx1 for the refinement of exuberant axonal projections to subcortical targets. *Neuron* **24**, 819–831.

Yu X. and Weissman S. M. (2000) Characterization of the promoter of human leukocyte-specific transcript 1. A small gene with a complex pattern alternative transcripts. *J. Biol. Chem.* **275**, 34 597–34 608.