

Graded phenotypic response to partial and complete deficiency of a brain-specific transcript variant of the winged helix transcription factor RFX4

Perry J. Blackshear^{1,2,4,5}, Joan P. Graves³, Deborah J. Stumpo², Inma Cobos⁶, John L. R. Rubenstein⁶ and Darryl C. Zeldin^{1,3,4,*}

¹Office of Clinical Research, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

²Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

³Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

⁴Department of Medicine, Duke University Medical Center, Durham, NC 27710, USA

⁵Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

⁶Nina Ireland Laboratory of Developmental Neurobiology, Department of Psychiatry, University of California, San Francisco, San Francisco, CA 94143, USA

*Author for correspondence (e-mail: zeldin@niehs.nih.gov)

Development 130, 4539-4552

© 2003 The Company of Biologists Ltd

doi:10.1242/dev.00661

Accepted 10 June 2003

Summary

One line of mice harboring a cardiac-specific epoxygenase transgene developed head swelling and rapid neurological decline in young adulthood, and had marked hydrocephalus of the lateral and third ventricles. The transgene was found to be inserted into an intron in the mouse *Rfx4* locus. This insertion apparently prevented expression of a novel variant transcript of RFX4 (RFX4_v3), a member of the regulatory factor X family of winged helix transcription factors. Interruption of two alleles resulted in profound failure of dorsal midline brain structure formation and perinatal death, presumably by interfering with expression of downstream genes.

Interruption of a single allele prevented formation of the subcommissural organ, a structure important for cerebrospinal fluid flow through the aqueduct of Sylvius, and resulted in congenital hydrocephalus. These data implicate the RFX4_v3 variant transcript as being crucial for early brain development, as well as for the genesis of the subcommissural organ. These findings may be relevant to human congenital hydrocephalus, a birth defect that affects ~0.6 per 1000 newborns.

Key words: Hydrocephalus, Regulatory factor X, Winged helix transcription factor, Cortex, Midline, Mouse

Introduction

Targeted insertional mutagenesis in mice has become the standard method for uncovering the roles of a specific gene in development. However, several instances of accidental insertion of a transgene into a crucial genomic locus have yielded important information as well. For example, a reeler-like phenotype was observed in one line of transgenic mice harboring an unrelated transgene (Miao et al., 1994). The transgene had interrupted what is now known as the *reeler* locus, and much has since been learned about the function of this gene and its gene product, reelin, in regulating the development of the central nervous system (D'Arcangelo et al., 1995; D'Arcangelo et al., 1996; Rice and Curran, 2001). Several other examples have been described recently (Durkin et al., 2001; Friedman et al., 2000; Overbeek et al., 2001).

The present studies began with the evaluation of a line of transgenic mice in which the cardiac-specific expression of an unrelated transgene (CYP2J2) was associated with the development of head swelling and rapid neurological decline in young adulthood. Anatomical characterization of these mice revealed severe congenital hydrocephalus that was nonetheless compatible with life and fertility in some cases. This obstructive hydrocephalus appeared to be secondary to failure

of development of the subcommissural organ (SCO), a structure that is important for the patency of the aqueduct of Sylvius and normal cerebrospinal fluid flow in the brain (Cifuentes et al., 1994; Perez-Figares et al., 1998; Perez-Figares et al., 2001; Rodriguez et al., 2001; Rodriguez et al., 1998; Vio et al., 2000). Identification of the genomic sequences flanking the inserted transgene led to the discovery that the transgene interfered with the expression of a novel brain-specific isoform of the winged helix transcription factor regulatory factor X4 (RFX4), which has been named RFX4 variant transcript 3, or RFX4_v3.

Fetal mice completely lacking in RFX4_v3 expression exhibited severe defects in the formation of dorsal midline brain structures, and intra-uterine or perinatal death. Thus, this accidental transgene insertion led to the identification of a novel splice variant of RFX4 that is crucial for normal brain development. In addition, disruption of a single allele led to an autosomal dominant pattern of expression of congenital hydrocephalus. Given the 96% identity between the mouse and human protein products of RFX4_v3, it seems possible that abnormalities of expression or primary sequence of the human gene could result in some cases of congenital obstructive hydrocephalus, a common human birth defect.

Materials and methods

Mice

These mice were generated in an unrelated study in which transgenic mice were created for the cardiac-specific expression of human CYP2J2, a cytochrome P450 arachidonic acid epoxygenase, using a mouse α -myosin heavy chain promoter and a human growth hormone 3'-untranslated region (3'-UTR). Details of the transgene construction and methods used to create the transgenic mice will be described elsewhere, but are briefly summarized here. The coding region of the 1.8 kb CYP2J2 cDNA (GenBank Accession Number U37143) was cloned into the *SalI-HindIII* sites of the vector pBS- α MHC-hGH (clone 26), a generous gift from Dr Jeffrey Robbins (University of Cincinnati, OH). This vector contains the 5.5 kb α -myosin heavy chain promoter to drive cardiomyocyte-specific expression of the transgene and 0.6 kb of human growth hormone/polyA sequences to enhance transgene mRNA stability. The linearized transgene (~7.9 kb) was microinjected into pronuclei of single cell C57BL6/J mouse embryos which were implanted into pseudopregnant mice. Founder pups were identified by a combination of PCR and Southern blotting of tail genomic DNAs. PCR reactions used reagents from Applied Biosystems (Foster City, CA) and the following oligonucleotide primers: α MHCF1, 5'-GGCACTCTTAGCAAACCTCAGG-3'; CYP2J2R1, 5'-AGCCAGTAATAAGAACTGCAGA-3'; α MHCF2, 5'-TCTGACAGAGAAGCAGGCACTTTA-3'; and CYP2J2R2, 5'-AAGATATGTTCTCGCATAGGGGTC-3'. All studies used mice from CYP2J2 founder line Tr5 and were approved by the NIEHS Animal Care and Use Committee.

Histology

For routine histology, embryos and tissues from newborn or adult mice were fixed in Bouin's fixative for 12-48 hours, depending on tissue size, and then cleared in 70% (v/v) ethanol. They were then embedded in paraffin wax, sectioned and stained with Hematoxylin/Eosin by standard methods. For immunohistochemistry, paraffin wax-embedded sections were stained with an antibody (Rodriguez et al., 1984) to Reissner's fibers (RF) within the SCO, as described previously for a different antibody (Blackshear et al., 1996). The anti-RF antibody was a generous gift from Dr E. M. Rodriguez (Instituto de Histología y Patología, Facultad de Medicina, Universidad Austral de Chile, Valdivia, Chile).

Identification of the transgene insertion site

A Universal GenomeWalker Kit (Clontech, Palo Alto, CA) was used to identify the mouse genomic sequences adjacent to the transgene insertion site. Briefly, genomic DNA from transgenic mice was digested with *DraI*, *EcoRV*, *PvuII* or *StuI*, and ligated to adaptors supplied by the manufacturer. PCR amplification of 3' adjacent sequences used the Advantage Genomic PCR Kit (Clontech), the universal adaptor primers AP1 and AP2, and the following nested gene-specific primers: 5'-ACAACCTCTGCGATGGGCTCTGCTTT-3' and 5'-CTGACCAATTTGACGGCGCTGCACA-3'. PCR products were cloned into the pCRII vector using the TA Cloning Kit (Invitrogen/Life Technologies, Carlsbad, CA) and sequenced using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA). PCR amplification of 5' adjacent sequences was similarly performed using the following nested gene-specific primers: 5'-GGCCATTGTCAACCACTCGTAA-3' and 5'-CACAAAGTAAAGGCTAACGCGC-3'.

Plasmids

Plasmids containing the indicated human, mouse and zebrafish ESTs were obtained from the IMAGE consortium. A plasmid containing the putative protein coding region of the mouse RFX4_v3 was made by first using Superscript II RNase H⁻ Reverse Transcriptase (Invitrogen/LifeTechnologies, Carlsbad, CA) to reverse transcribe total adult mouse brain RNA template. The resulting cDNA was then

subjected to two rounds of nested PCR using Platinum Pfx DNA polymerase (Invitrogen/LifeTechnologies) and primers based on the 5' and 3' sequences of apparent mouse brain RFX4 sequences from GenBank. The first pair of primers corresponded to bp 255-278 of Accession Number BB873367 and to bp 100-124 of Accession Number BB379807, and the second set of primers corresponded to bp 291-309 of Accession Number BB873367 and bp 99-78 of Accession Number BB379807. The resulting PCR product was sequenced using the ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA).

Probes corresponding to the unique 5'-ends of mouse RFX4_v1 and RFX4_v3 were constructed by PCR amplification of reverse-transcribed mouse testis RNA or brain RNA, respectively. Reverse transcription was carried out using 1 μ g of total RNA, an anchored oligo (dT) primer (T₁₈VN) and Superscript II RNase H⁻ Reverse Transcriptase (Invitrogen Life Technologies, Carlsbad, CA). PCR was performed using primers based on the sequence for human RFX4_v1 (Accession Number NM_032491) or the sequence for mouse RFX4_v3 contained in the mouse brain EST Accession Number BB595996. The forward primer for RFX4_v1 was 5'-AGGTG-GGAAGGCAGTTATGACAG-3' (corresponding to bases 1-23 of NM_032491) and the reverse primer was 5'-TCCGTGATATTT-CTGCTTAGTGGGC-3' (bases 201-177). A second round of PCR was carried out with forward primer 5'-GGCAGTTATGACAGTTGA-GAAGTAGTAG-3' (bases 10-37) and reverse primer 5'-CTGCTT-AGTGGGCATCTCGAATCTATC-3' (bases 189-163). The forward primer for mouse RFX4_v3 was 5'-TTTTGACGGGTTGGCTTTG-3' (bases 118-137 of BB595996) and the reverse primer was 5'-TTCCTCCAGTAACCCACAATGC-3' (bases 447-426). A probe corresponding to the unique 5'-end of RFX4_v2 was isolated by PCR amplification from mouse L cell genomic DNA using primers based on the sequence for human RFX4_v2 (Accession Number NM_002920). PCR was carried out using forward primer 5'-TGGAGAGGCCACAGCTGCTGG-3' (bases 1-21 of NM_002920) and reverse primer 5'-TCGAGGCCTGGTCTGTGCGC-3' (bases 159-140). A second round of PCR was performed with 5'-CACAGCTGCTGGCTTCCTGG-3' (bases 10-29) and the same reverse primer as in the first round of PCR. All three unique 5'-ends of RFX4_v1, RFX4_v2 and RFX4_v3 were sequenced using the ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA).

A cDNA corresponding to human RFX4_v3 was cloned by screening a human fetal brain cDNA library (Stratagene) with the insert from the human IMAGE clone # 46678 (GenBank Accession Number H10145). The resulting cDNA clone was sequenced by dideoxynucleotide techniques (see above). A plasmid (GenBank Accession Number AI657628) containing a zebrafish EST sequence that predicted a protein closely related to the N terminus of mouse and human RFX4_v3 was also obtained from the IMAGE Consortium and sequenced by dideoxynucleotide techniques.

In situ hybridization histochemistry

Embryos were dissected in PBS and fixed in 4% (w/v) paraformaldehyde/PBS at 4°C. Specimens for whole-mount in situ hybridization were gradually dehydrated in methanol/PBS and stored in 100% methanol at -80°C. Specimens for in situ hybridization on frozen sections were cryoprotected in 30% sucrose and embedded in TissueTek (Sakura), and 20 μ m sections were obtained using a cryostat. Whole-mount and section in situ hybridization was performed according to the methods of Wilkinson (Wilkinson, 1992) and Tsuchida et al. (Tsuchida et al., 1994), respectively. The probes used and their sources were as follows: Rfx4 (this paper); Otx2 (Antonio Simeone); Bf1 (Eseng Lai); Fgf8 (Gail Martin); Msx2 (Betham Thomas); Wnt3a and Wnt7b (Andrew McMahon); Lhx2 (Heiner Westphal); Pax6 and Six3 (Peter Gruss); and Emx1, Dlx2 and Nkx2.1 (J.L.R.R.'s laboratory).

Results

Evaluation of transgenic mice

This study was initiated by the observation that a large percentage of mice in one (Tr5) of six transgenic lines exhibited head swelling followed by rapid neurological deterioration and death in young adulthood. The external swelling was apparent by the increased convexity of the head, and the lateral displacement of the ears (Fig. 1A). Histological examination of the brains of symptomatic adult mice revealed severe hydrocephalus in the anterior brain, with extreme dilatation of the lateral ventricles but no apparent effect on the fourth ventricle (Fig. 1B). Although many of the mice developed the severe form of the syndrome within the first 2 months of life, sufficient mice survived to propagate the line. Nonetheless, examination of the brains of successful adult breeders showed severe hydrocephalus, with extreme lateral ventricle dilatation and the formation of false ventricles near the external capsule, as well as midline structural disruption by the extreme hydrocephalus (data not shown). These findings were compatible with an obstructive hydrocephalus, and were consistent with the form of hydrocephalus seen with stenosis of the aqueduct of Sylvius or aqueductal stenosis. It should be noted that CYP2J2 transgene expression did not occur in brains from the transgenic mice, as evaluated with two different CYP2J2-specific antibodies on western blots.

Examination of transgenic mice from the Tr5 line at the time of birth (P0.5) showed that severe hydrocephalus was present in all mice harboring the transgene, indicating that the

hydrocephalus was congenital. By contrast, none of the wild-type littermates had hydrocephalus. The hydrocephalus was most apparent in the olfactory and lateral ventricles, with apparent sparing of the fourth ventricle (Fig. 2). These data supported the possibility of a congenital obstruction in the aqueduct of Sylvius.

Examination of the aqueduct in serial coronal sections from a transgenic mouse and its wild-type littermate at P0.5 showed the apparent absence of the subcommissural organ (SCO) in the transgenic mice (Fig. 3A). This organ produces Reissner's fibers, and both the organ and the fibers have been shown to be important for the patency of the aqueduct, in that destruction of the SCO leads to obstructive hydrocephalus (Perez-Figares et al., 2001). Antibodies specific to Reissner's fibers (Rodriguez et al., 1984; Rodriguez et al., 2001; Rodriguez et al., 1998) strongly and specifically labeled the SCO from the wild-type mice (Fig. 3B), but this label was generally not detected in the same anatomical region in the transgenic mice. Rarely, a small amount of staining could be found in sections from the transgenic mice at the anatomical location that should have contained the SCO (Fig. 3B); however, this staining was always markedly less than that seen in the wild-type mice. Although the SCO appeared to be largely absent in the transgenic mice, other midline structures, such as the pineal body and posterior commissure, were present and appeared to be anatomically normal (data not shown).

We next examined the birth statistics from this line of transgenic animals for Mendelian frequencies. For crosses in which transgenic mice were bred to wild-type mice, there were 6.7 ± 0.4 live births per litter based on data from 47 litters. Of 315 pups born, 46% were transgenic and 54% were wild type. For comparison, transgenic mice originating from another founder line crossed with wild-type mice resulted in 7.0 ± 0.4 live births per litter based on data from 45 litters, with 52% of 317 pups genotyped as transgenic. These data suggest minimal if any prenatal loss of transgenic pups, despite the presence of congenital hydrocephalus. In the transgenic mice, severe hydrocephalus requiring euthanasia developed in about 75% of the mice at an average age of 47 ± 3 days (range 24–84 days). There was no significant difference in frequency of hydrocephalus between males and females. The hydrocephalus phenotype has persisted in transgenic mice through nine generations.

All other non-brain tissues of the transgenic mice appeared to be histologically normal.

Identification of genomic sequences flanking the transgene

Our working hypothesis was that the transgene had interrupted the coding or regulatory regions of an important gene, and we therefore attempted to identify the mouse genomic sequences flanking the transgene. Using PCR based on 5' and 3' transgene sequences, we found that there were at least two tandem copies of the 7.9 kb transgene in genomic DNA from the transgenic mice, indicating that the potential genomic interruption was at least 15 kb in size; Southern analysis using a transgene-specific probe indicated that there was only one copy of this concatenated transgene in the mouse genome (data not shown). Using the 'GenomeWalker' technique with genomic DNA from the transgenic mice and transgene-specific oligonucleotide primers, we identified

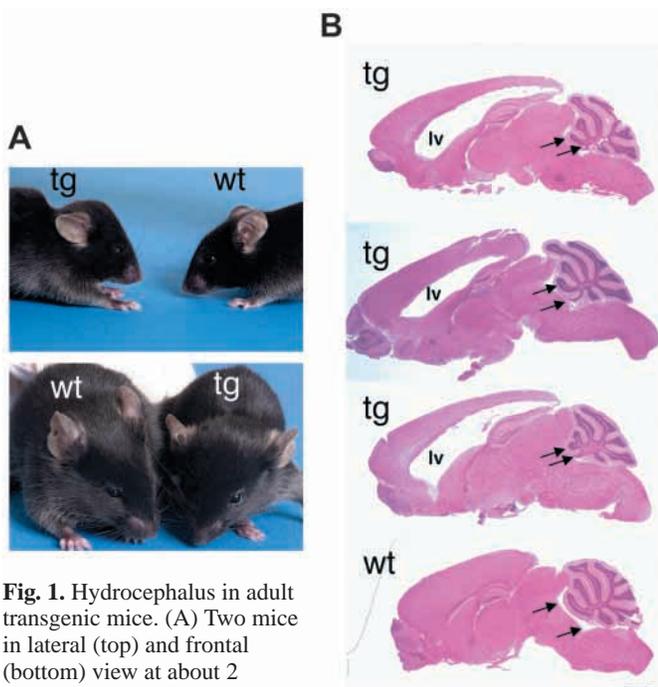


Fig. 1. Hydrocephalus in adult transgenic mice. (A) Two mice in lateral (top) and frontal (bottom) view at about 2 months of age, showing the characteristic domed head and lateral displacement of the ears in the transgenic mouse compared with its wild-type littermate. (B) Parasagittal sections, stained with Hematoxylin and Eosin, of brains from four littermate mice, three transgenic and one wild type, at about seven weeks of age. The marked dilatation of the lateral ventricles (LV) is obvious in the transgenic mice; however, there is no evidence for dilatation of the fourth ventricles (arrows). Scale bar: 1 mm.

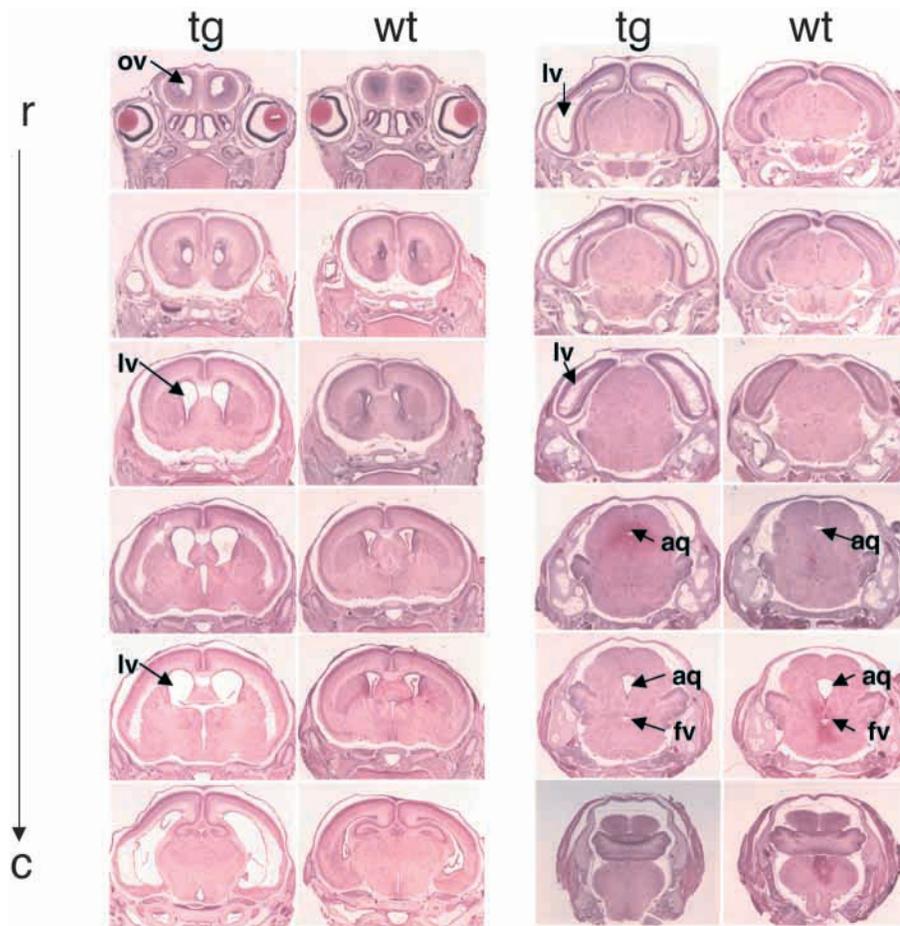


Fig. 2. Hydrocephalus in newborn transgenic mice. Serial rostral (r) to caudal (c) coronal sections, stained with Hematoxylin and Eosin, from newborn (P0.5) transgenic and wild-type littermates are shown, with each pair of sections representing approximately the same coronal plane. Note the extreme hydrocephalus apparent in the olfactory ventricles (OV) and the lateral ventricles (LV) of the transgenic compared with the wild-type mouse. In the more posterior sections, note the similar appearance of the aqueduct of Sylvius (aq) and the fourth ventricle (fv) in the wild-type and transgenic mice.

both the 5' and 3' flanking genomic sequences into which the transgene had been inserted. When these sequences were compared to the mouse genomic sequences in the GenBank trace archives, the transgene insertion site was identified as between bp 528 and 529 in *gnl|t|13973384* and between bp 171 and 172 in *gnl|t|84074979*. The 5' and 3' flanking sequences identified by the GenomeWalker technique were contiguous in the normal mouse genomic sequences in the trace archives, indicating that the transgene insertion was not accompanied by a genomic deletion, as has been seen in some recent examples of accidental transgenic insertional mutagenesis (Durkin et al., 2001; Overbeek et al., 2001). Southern analysis using a 3' insertion site-specific probe demonstrated the presence of single novel bands in restriction enzyme-digested DNA from the transgenic mice, confirming a single transgene insertion site at this location (Fig. 4A).

The flanking sequences identified by the GenomeWalker approach were merged with the available mouse genomic sequence from the trace archives to form a small contig; this did not recognize any cDNAs or expressed sequence tags (ESTs) in the database at that time. Therefore, we used the assembled mouse contig to search the human genome sequences then available in GenBank, using *blastn*. The mouse sequence was highly related ($4e-28$) to a human genomic sequence corresponding to a region of human chromosome 12 (accession number NT_009720.8). When this small region of human genomic sequence was analyzed for expressed sequences, it did not match any deposited in GenBank. However,

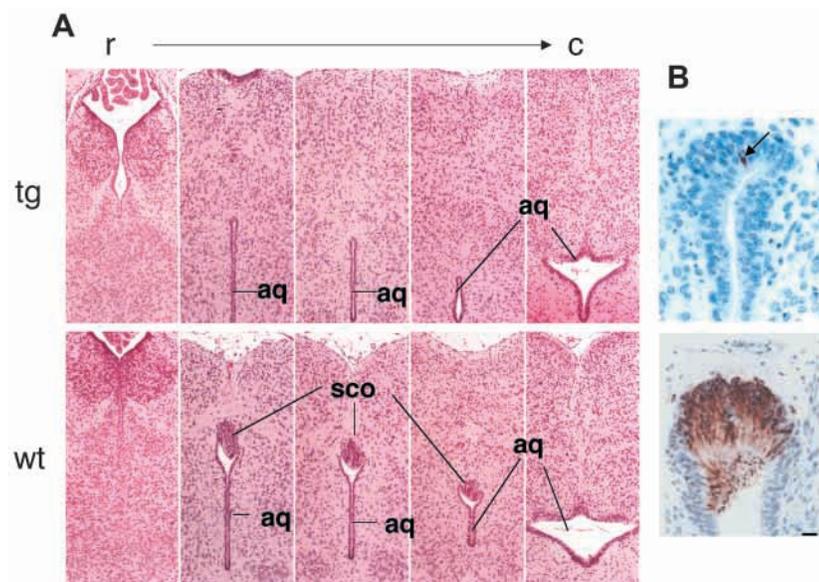


Fig. 3. Aqueduct of Sylvius and SCO in wild-type and transgenic mice. (A) Coronal sections in a rostral (r) to caudal (c) direction from P0.5 wild-type and transgenic littermates stained with Hematoxylin and Eosin, demonstrating the apparent absence of the SCO in the transgenic mouse. (B) Similar sections stained with an antibody to Reissner's fibers. Note the near-absence of antibody staining in the transgenic section (top) compared with the wild-type section (bottom). The arrow in the top section indicates a small amount of antibody staining in one section from the knockout mouse, indicating the presence of the Reissner's fiber antigen. The counterstain was Hematoxylin. Scale bar in B: 50 μ m (bottom); 20 μ m (top).

when a much larger amount of human genomic DNA from this locus was used to search for expressed sequences, genomic DNA within 200 kb of the human sequence corresponding to the transgene insertion site was found to contain all of the exons of two distinct cDNAs in GenBank that correspond to two forms of the human winged helix protein RFX4: one is represented by GenBank Accession Number NM_032491, referred to as RFX4 variant transcript 1 or RFX4_v1, and corresponds to protein Accession Number NP_115880; the other is represented by GenBank Accession Number NM_002920 and is referred to as RFX4 variant transcript 2 or RFX4_v2, corresponding to protein accession number NP_002911. See the nomenclature recommendations of the Human Genome Nomenclature committee (<http://www.gene.ucl.ac.uk/nomenclature/guidelines.html#Appendix>) for the conventions described here.

According to the mouse-human alignments, the site of the transgene insertion within the mouse genome was at a corresponding region within the human chromosomal 12 sequence that would be within the intron between exons 13 and 14 of RFX4_v1 (see below); it would not have affected the exon arrangements of RFX4_v2.

Using PCR primers based on the inserted transgene and the neighboring endogenous mouse genomic DNA, we found that the wild-type (+/+) and transgene-interrupted alleles (+/- for one allele disrupted, -/- for both alleles disrupted) could be readily distinguished in a litter of newborn mice from interbred transgenic mice (Fig. 4B).

To examine the possibility that the transgene insertion had in some way interfered with the expression of a full-length mouse RFX4 transcript in brain, we probed northern blots from brains of neonatal +/+, +/- and -/- mice with a mouse

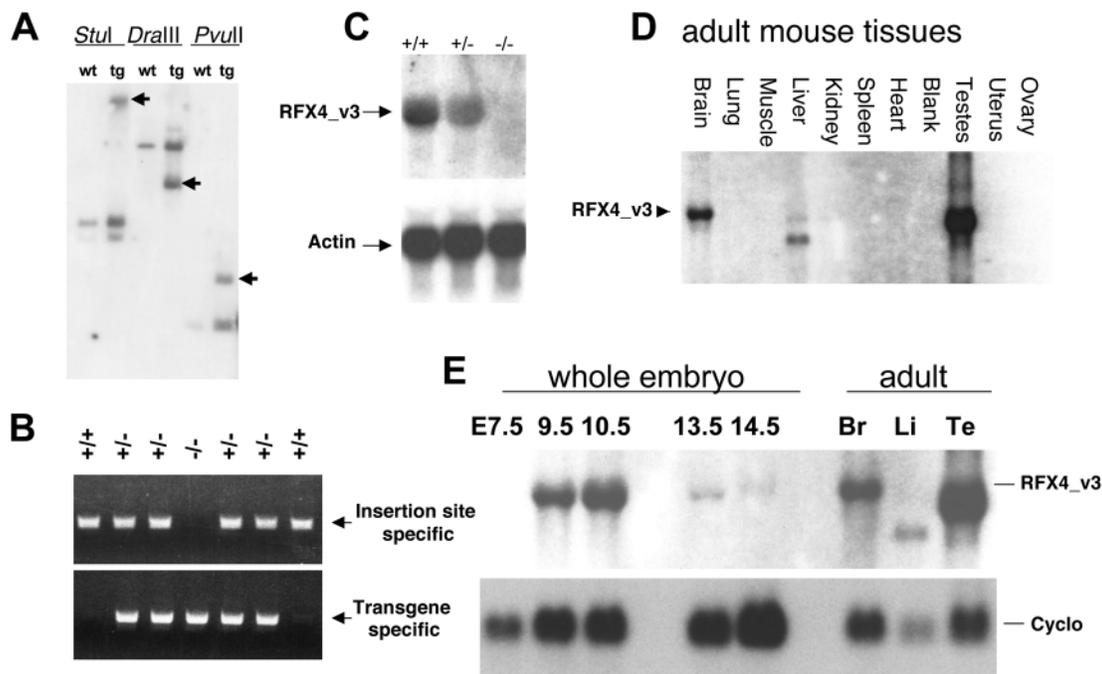


Fig. 4. Identification of transgene insertion site. (A) A Southern blot of genomic DNA from wild-type and transgenic mice, digested with the three restriction enzymes indicated and probed with a 3'-insertion site-specific probe. The arrows indicate the three single, novel bands hybridizing to the probe in the DNA from the transgenic mice, indicating the likelihood of a single transgene insertion site. (B) A PCR-based analysis of genomic DNA from one litter of interbred transgenic mice, indicating the PCR products that were specific for the presence of the transgene (Transgene-specific) and those that were specific for the endogenous sequence that was interrupted by the transgene (Insertion site-specific). The transgene specific primers were 5'-AGCCAGTAATAAGAAGTGCAGA-3' and 5'-GGCACTCTTAGCAAACCTCAGG-3', which correspond to bp 264-285 of the human cytochrome P450 cDNA clone accession number NM_000775.2 and bp 5225-5246 of the mouse α -myosin heavy chain promoter clone Accession Number MMU71441, respectively. The insertion site specific primers were 5'-CATGGAAAGGGCAGAGTGAGC-3' and 5'-GGCCATTGTCACCACTCGTAA-3', which correspond to bp 732-752 and bp 323-343 of mouse trace archive sequence gnl|ti|91911671, respectively. In both cases, the results were confirmed by PCR using different pairs of primers. The DNA is characterized as +/+, +/- and -/- by the presence of the interrupted allele. (C) A northern blot of total brain RNA from newborn mice of the +/+, +/- and -/- genotypes. This blot was probed with a mouse EST clone that was 94% identical over 284 bases to a region corresponding to the 3'-end of the human testis-specific RFX4 transcript H10145. The only visible transcript was of ~4 kb (RFX4_v3); this was decreased in expression in the +/- sample, and undetectable in the -/- sample. Longer exposure of the blot did not reveal the presence of any truncated mRNA species in the +/- and -/- lanes. The same blot was hybridized to an actin cDNA (lower panel), and demonstrates roughly equivalent loading of the three RNA samples. (D) The hybridization of the same probe to adult mouse tissues, revealing an ~4 kb transcript in brain (RFX4_v3), a 3.7 kb transcript in testis and a still smaller transcript in liver. (E) The pattern of developmental expression of the 4 kb transcript, which was undetectable in whole embryos at E7.5, highly expressed in whole embryos at E9.5 and 10.5, and less well expressed at E13.5 and 14.5. The brain, liver and testis lanes from D are juxtaposed in E to illustrate the difference in size between the brain (RFX4_v3), liver and testis transcripts, and the size identity of the adult brain transcript and the embryonic transcript. Also shown is the expression of a control mRNA for cyclophilin (Cyclo.).

brain EST cDNA clone (IMAGE # 763537, GenBank Accession Numbers AA285775 and AI462920) that was highly related (e-124 over 284 aligned bases) to the 3' end of the human cDNA for RFX4_v1. Brains from the +/+ mice expressed a prominent band of ~4 kb that we will refer to as RFX4 variant transcript 3 or RFX4_v3 (Fig. 4C; see below). Brains from the +/- mice expressed ~50% of the normal complement of this transcript, whereas the brains from the -/- mice expressed no detectable transcript of this size (Fig. 4C).

Probing the same blot with an actin cDNA demonstrated that gel loading was similar in the three lanes (Fig. 4C). Similar results were obtained in three separate experiments. There was no evidence for the expression of a truncated mRNA in the brain samples from either the +/- or -/- mice (data not shown). These studies confirmed that an mRNA species of ~4 kb that was recognized by a probe derived from putative mouse 3' RFX4_v1 sequences was decreased in amount in brains of the +/- mice, and absent from the brains of the -/-

mice. These data suggested that the insertion of the transgene interfered with the expression of the putative brain RFX4_v3 transcript.

Using the same probe to examine the tissue-specific and developmental expression of this RFX4 transcript, we found high-level expression of a slightly smaller transcript in normal adult testis, and lower level expression of a considerably smaller transcript in liver (Fig. 4D). The largest species, corresponding to the apparent brain-specific transcript labeled RFX4_v3 in Fig. 4D, was the only one detected in whole embryos early in development (Fig. 4E). These data suggested that an apparently brain-specific isoform of RFX4 in the adult was highly expressed in the whole embryo during early development, initially appearing between embryonic day (E) 7.5 and 9.5 (Fig. 4E).

Identification of the RFX4_v3 transcripts and proteins

Using primers based on mouse brain EST sequences that contained internal sequences highly related to the human RFX4 cDNAs in GenBank, we used PCR and an adult mouse brain cDNA library to generate a ~3 kb plasmid insert that was then sequenced. This cDNA has been designated RFX4 transcript variant 3 (RFX4_v3), and the mouse sequence has been deposited in GenBank (Accession Number AY102010). When this sequence was merged with all available 5' and 3' mouse ESTs from GenBank, the resulting transcript was 3952 bp, closely approximating the transcript size seen on northern blots. While this paper was under revision, a cDNA sequence was deposited in GenBank on 5 December 2002 (GenBank Accession Number AK034131.1) that was 3535 bp in length; over this length, it was more than 99% identical to the putative

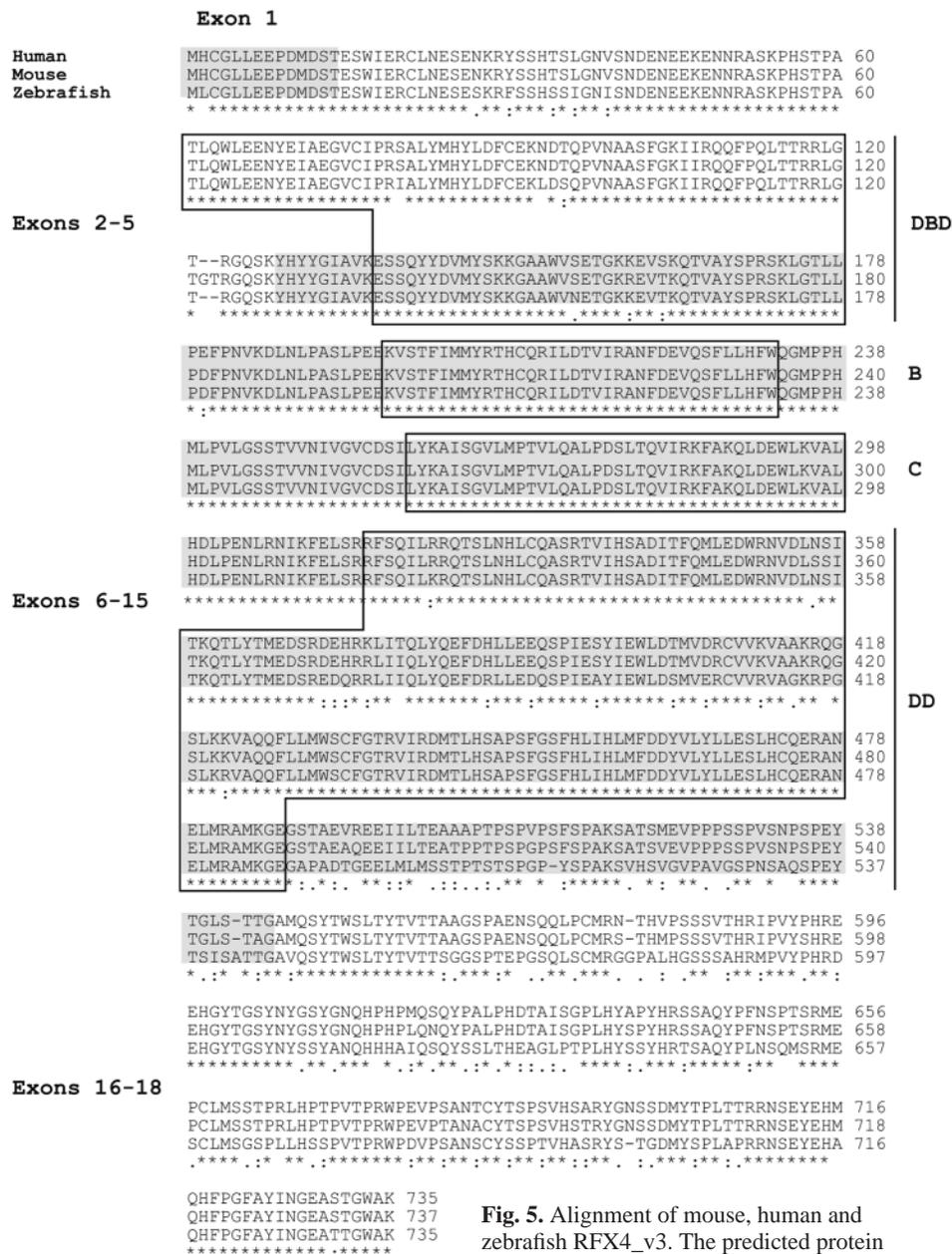


Fig. 5. Alignment of mouse, human and zebrafish RFX4_v3. The predicted protein sequences from these three RFX4_v3 orthologs

were aligned using ClustalW. The position of the characteristic RFX DNA binding domain (DBD) is indicated by the box; other boxes contain the B and C boxes, and the dimerization domain (DD). The shaded first 14 amino acids labeled exon 1 were unique to RFX4_v3 (human); the next unshaded sequences represent exons 2-5 and are identical to sequences from RFX4_v2; the next shaded sequences represent exons 6-15 and are identical to sequences from both RFX4_v1 and RFX4_v2; the next unshaded sequences represent exons 16-18 and are identical to sequences in RFX4_v1. Asterisks indicate amino acid identity; double dots indicate a high degree of amino acid similarity; single dots indicate less similarity.

RFX4_v3 full-length transcript described above, and included the entire putative protein coding region. This cDNA was isolated from an adult male mouse diencephalon library and confirms the existence in brain of at least the protein-coding region of our predicted full-length RFX4_v3 transcript.

Similar probes as used to generate the northern blots shown in Fig. 4 were then used to screen a human brain cDNA library, and positive inserts were sequenced. This cDNA sequence has been deposited in GenBank as human RFX4_v3 (Accession Number AY102009). The predicted unique mouse N-terminal protein sequence (see below) also was used to search the non-human, non-mouse ESTs in GenBank, and a zebrafish EST clone (Accession Number AI657628) with a nearly identical predicted N-terminal protein sequence was obtained from the IMAGE consortium and sequenced. This sequence is referred to as zebrafish RFX4_v3, and the complete insert cDNA sequence has been assigned accession number AY102011.

An alignment of these three predicted amino acid sequences is shown in Fig. 5. There was 96% amino acid identity between the predicted mouse and human proteins, and 83% amino acid identity between the predicted human and zebrafish proteins. The alignment also illustrates several of the characteristic domains of the RFX proteins that are highly conserved in all three orthologs, i.e. the DNA binding domain, boxes B and C, and the dimerization domain (Morotomi-Yano et al., 2002).

We then re-searched the human chromosome 12 sequence

with the mouse and human cDNA sequences, and determined the exons that contributed to the novel human RFX4_v3 isoform described here, in addition to those described above that corresponded to the two previously described human cDNAs. The results of this analysis are shown in Fig. 6. The two previously described human RFX4 cDNAs are composed of both unique and shared exons. In the case of the cDNA represented by Accession Number NM_002920 (RFX4_v2), the first five exons (shown in green in Fig. 6) correspond to five exons within the 90 kb interval between bp 390,000-480,000 of the genomic clone NT_009720.8 (in reverse complement orientation). The next nine exons and part of a tenth (yellow) are common to the other version of RFX4 in GenBank (RFX4_v1), represented by the cDNA NM_032491. These 10 exons are derived from coding sequences in the genomic clone NT_009720.8 between 340,000 and 400,000. As shown in Fig. 6, the final (15th) exon of RFX4_v2 contains a polyadenylation sequence that allows for final processing of the mature mRNA.

The other human cDNA, RFX4_v1 (NM_032491), contains a 5' exon (red hatching) that is encoded by genomic sequences in NT_009720.8 that are located *between* the exons 5 and 6 of RFX4_v2 (Fig. 6) and is unique to that cDNA. RFX4_v1 then shares 10 exons with RFX4_v2 (yellow), followed by three unique 3' exons (red). These last three unique exons are found within the interval between bp 315,000-325,000 of the genomic clone NT_009720.8. Remarkably, exon 11 from RFX4_v1 is the

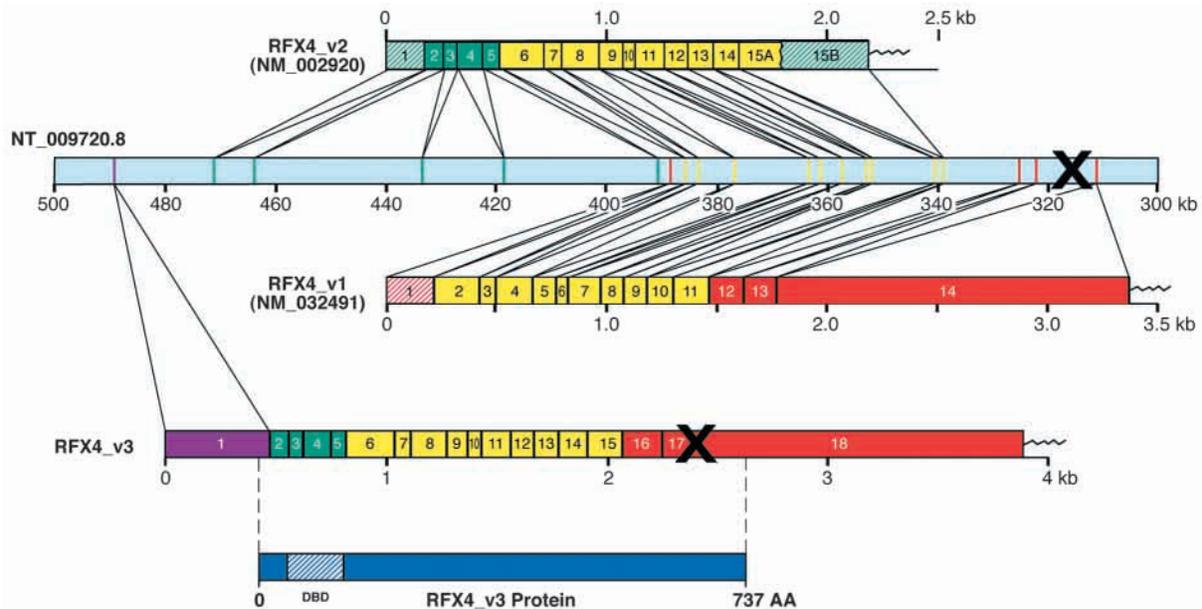


Fig. 6. The human RFX4 locus and its three known transcripts. This figure is a schematic representation of 200 kb of human genomic sequence from NT_009720.8, shown in reverse complement orientation, and of the position within this sequence of the exons that comprise the three indicated RFX4 transcripts. At the top of the figure is shown the transcript corresponding to RFX4_v2 (Accession Number NM_002920). Exon 1 is unique to this transcript (green hatching); exons 2-5 (solid green) are shared with the novel RFX4_v3 transcript described in this paper; exons 6-15A (yellow) are shared with the RFX4_v3 transcript as well as the transcript RFX4_v1; exon 15B (green hatching) is apparently unique to this transcript, and contains a polyadenylation sequence and presumably a polyA tail, as indicated by the wavy line. The location of these exons on the genomic sequence are indicated. Below the genomic sequence is represented the transcript RFX4_v1. It contains a unique exon 1 (red hatching); exons 2-11 (yellow) shared with both RFX4_v2 and RFX4_v3; and exons 12-14 (red) shared only with RFX4_v3. The RFX4_v3 transcript contains a unique exon 1 (purple); exons 2-5 (green) shared only with RFX4_v2; exons 6-15 (yellow) shared with both RFX4_v1 and RFX4_v2; and exons 16-18 (red) shared only with RFX4_v1. The site of transgene insertion is indicated in the genomic clone by the black X in the intron between exons 13 and 14 of RFX4_v1; its position between exons 17 and 18 of RFX4_v3 is also indicated. The regions of the RFX4_v3 transcript coding for the 737 amino acid human RFX4_v3 protein (blue) are indicated, as is the DNA-binding domain (DBD) of the protein. See the text for additional details.

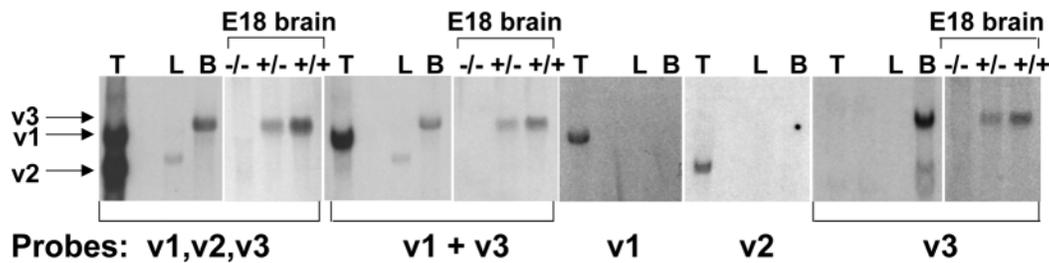


Fig. 7. Northern analysis of RFX4 transcript expression using transcript-specific probes. cDNA probes corresponding to multiple or single RFX4 transcript variants, as indicated on the bottom of the figure, were used to probe northern blots containing total cellular RNA from adult testes (T), liver (L) and brain (B), or from brains of E18 mice of the +/+, +/- and -/- genotypes, as indicated. The blots were aligned to demonstrate the positions of the three hybridizing RFX4 species v1, v2 and v3 (arrows), as well as an uncharacterized transcript seen in adult mouse liver. There was no detectable hybridization of the specific v1 and v2 probes to the E18 brain RNA of any genotype (not shown).

same as the 5' region of exon 15 of RFX4_v2, shown as exon 15A in Fig. 6. Exon 15B in RFX4_v2 (green hatching) is the 5' region of a large intron between exons 11 and 12 in RFX4_v1, representing another presumed alternative splicing event.

The exon pattern that corresponds to the mouse and human RFX4_v3 mRNAs and proteins is illustrated at the bottom of Fig. 6. A novel exon (purple) derived from a sequence between 480,000 and 500,000 of NT_009720.8 was used to form the first 14 amino acids at the N terminus (Fig. 6). The next four exons, 2-5, are composed of the four exons of the same number from RFX4_v2 (green); exon 1 of RFX4_v2 (green hatching) is not present in the RFX4_v3 cDNA. The middle of the RFX4_v3 cDNA and protein are formed by the 10 exons (yellow) held in common between RFX4_v1 and RFX4_v2. The C terminus of RFX4_v3 is composed of the three C-terminal exons present only in RFX4_v1 (red). Thus, the novel RFX4_v3 isoform described here is composed of a unique arrangement of 18 exons derived from almost 200 kb of human genomic sequence. One exon (the first) is unique to this sequence; exons 2-5 are shared with RFX4_v2; exons 6-15 are shared with both RFX4_v1 and RFX4_v2; and exons 16-18 are shared with only RFX4_v1.

The site of transgene interruption is also illustrated in Fig. 6. The >15 kb transgene was inserted into the intron between exons 17 and 18 of RFX4_v3, within the C-terminal end of the protein coding region, and presumably interferes with splicing of the final exon and generation of an intact mature mRNA. We have found no evidence to date that a stable truncated mRNA species results from this transgene insertion.

We next designed and cloned specific cDNA probes corresponding to unique 5' sequences for each of the three RFX4 transcript variants RFX4_v1, v2 and v3. These were then used to probe northern blots of RNA from brains of E18.5 mice as well as from adult testes, liver and brain. As shown in Fig. 7, a probe that spanned regions common to the RFX4_v1, v2 and v3 transcripts hybridized to two major mRNA species in testes, a single transcript of intermediate size in liver, and a single transcript of the largest size (~4 kb) in RNA from adult brain. This probe only hybridized to the 4 kb RNA species in brains from E18.5 mice; the amount of hybridization of this probe decreased from the +/+ to the +/- mouse brain, and was undetectable in brain from the -/- sample. When similar blots were hybridized with a probe specific for v1 and v3, only the larger of the two testes transcripts (v1) was detected, while the largest transcript (v3) was again identified in the adult brain

sample and in the brain from E18.5 +/+ fetal mice. Again, the expression of the transcript hybridizing to this probe decreased with decreasing allelic dosage.

The identities of the various transcripts were determined by the use of transcript-specific probes, which confirmed the assignments of the v1 and v2 transcripts in testes, and the complete absence of hybridization of either probe to transcripts from normal adult brain (Fig. 7), or brain from E18.5 mice of the +/+, +/- and -/- genotypes (not shown). There was no evidence of compensatory expression of either the v1 or v2 transcripts in the E18.5 brains of the -/- mice. The v3-specific probe was used to confirm the identity of the single, large transcript in brain as RFX4_v3, and also confirmed its allelic dose-related expression in E18.5 mouse brain (Fig. 7).

These data indicate that the v3 transcript variant is the only form significantly expressed in the adult and fetal brain, and also confirmed it as the transcript variant expressed in the whole embryo and brain in earlier development (see Fig. 4E).

The identity of the apparently liver-specific transcript is not known, as it does not correspond to any of the three RFX4 transcript variants described above. It could represent a still unknown hypothetical 'RFX4_v4', or it could represent cross hybridization of the longer probes to another member of the RFX transcript family that is highly expressed in liver. We favor the latter possibility, as none of the shorter, specific v1-v3 probes hybridized to this species in our northern blots.

Analysis of RFX_v3 transcript expression during development

The pattern of RFX4_v3 transcript expression in mouse embryos was analyzed using RNA in situ hybridization. The data shown are from experiments in which a probe was used that contained sequences specific to both RFX4_v1 and v3. RFX4_v3 RNA was found primarily in the brain where its regional expression was highly dynamic during development. At E8.5, RFX4_v3 expression was detected in most of the neural plate, but its expression was excluded from the presumptive forebrain region (Fig. 8A,B). By E9.5, most of its expression encompassed two large regions: the caudal diencephalon/mesencephalon and the spinal cord (Fig. 8C). The rostral limit of the diencephalic expression approximated the zona limitans; the only expression extending anterior of this boundary was in the caudodorsal telencephalon (Fig. 8C).

At E10.5, RFX4_v3 expression extended throughout the

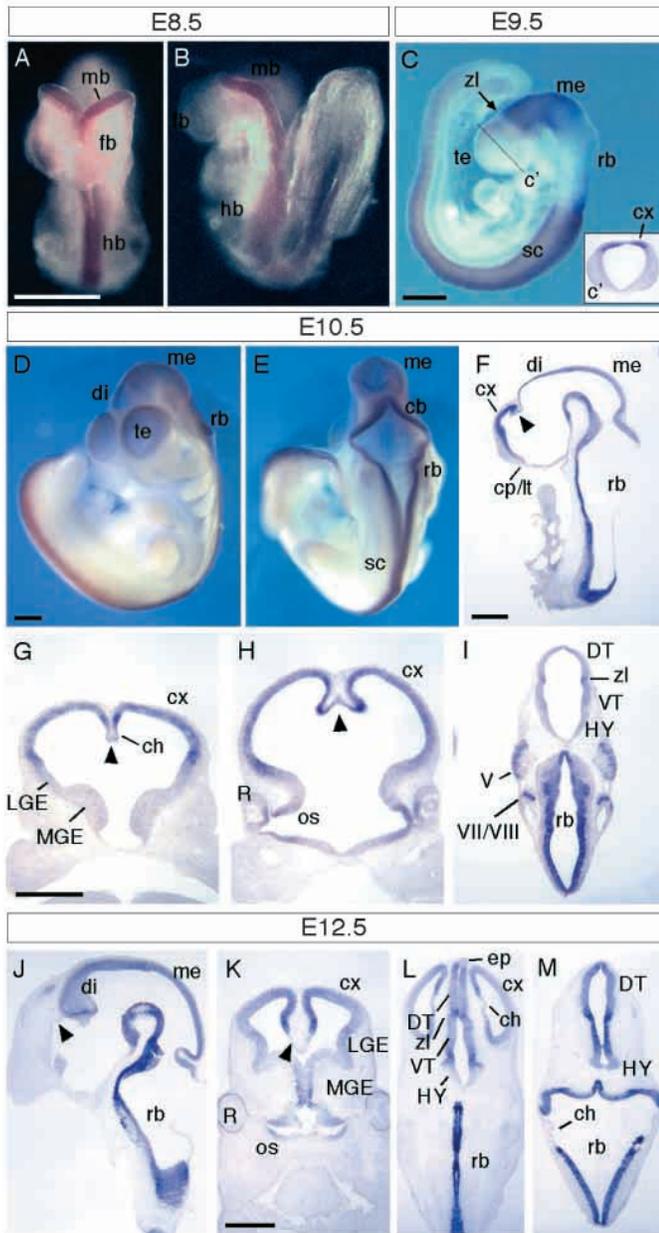


Fig. 8. Developmental expression of RFX4_v3. (A-E) Whole-mount embryos at the indicated embryonic days (E) in which the RFX4_v3 transcript is indicated by the blue digoxigenin staining. (C) Whole-mount suggests minimal staining rostral of the zona limitans (zl); however, a section through the plane indicated as C' shows staining of the dorsal cortex (cx). (D,E) Wholemounts at E10.5; (F) midline sagittal section; (G-I) coronal sections through similar embryos. The arrowheads in F-H indicate the lost expression in the telencephalic dorsal midline at E10.5. (J-M) One sagittal (J) and three rostral-to-caudal coronal sections through the head at E12.5. Note the lack of staining in the telencephalic dorsal midline (arrowheads in J,K) in the epiphysis (ep) in L and in the fourth ventricle choroid plexus (ch) in M. Scale bars: 500 μ m. mb, midbrain; fb, forebrain; hb, hindbrain; te, telencephalon; me, mesencephalon; rb, rhombencephalon; sc, spinal cord; di, diencephalon; cb, cerebellum; cp/lt, commissural plate/lamina terminalis; LGE, lateral ganglionic eminence; MGE, median ganglionic eminence; ch, choroid plexus; R, retina; os, optic stalk; DT, dorsal thalamus; VT, ventral thalamus; HY, hypothalamus; V, trigeminal ganglion; VII/VIII, facial/vestibular ganglion.

neural tube (Fig. 8D-F). In the telencephalon, its expression was limited to the cerebral cortex. Expression in the telencephalic dorsal midline was not detectable (Fig. 8F-H, arrowheads), and remained negative from that time onward during development. Thus, expression in the telencephalic roof plate was temporally restricted to the period just after neural tube closure (~E9.5).

Transient RFX4_v3 expression appeared in the central retina. The lateral optic stalks also exhibited RFX4_v3 expression (Fig. 8H), while the medial optic stalks showed expression at later stages (Fig. 8K).

From E12.5 to birth, the neuroepithelium and later the ependyma of most of the neural tube expressed variable levels of RFX4_v3 transcripts. For example, in the cerebral cortex, RFX4_v3 was expressed in a dorsal-to-ventral gradient (Fig. 8K). The majority of roof plate derivatives of the CNS, including most of the circumventricular organs, had turned off RFX4_v3 expression by this stage (for example, the epiphysis, and the choroid plexus of the lateral and fourth ventricles in Fig. 8L,M). A striking exception to this pattern was the expression of RFX4_v3 in the region of the developing SCO found in the caudal diencephalon, where there was strong expression from E14.5 to birth (Fig. 9C,E-G).

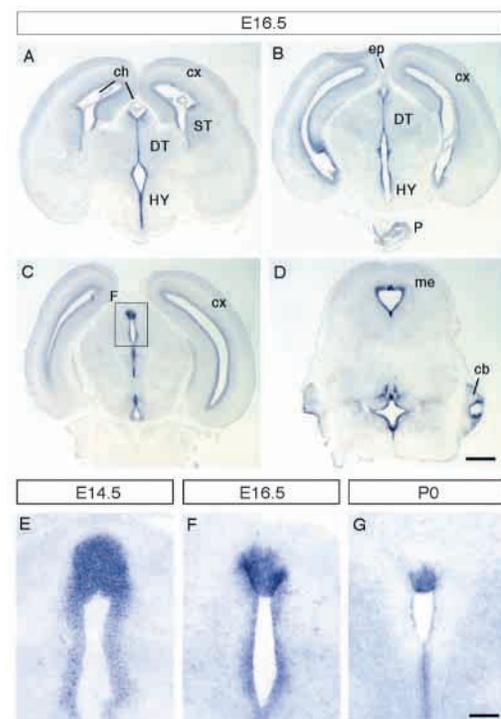


Fig. 9. RFX4_v3 in situ staining in the region of the developing SCO. (A-D) Progressively rostral-to-caudal sections through the brain of a normal embryo at E16.5. The box labeled F in section C contains the SCO and the aqueduct of Sylvius. This region is shown enlarged at E14.5 (E), E16.5 (F) and at the time of birth (P0; G). Note the high level expression of the RFX4_v3 transcript in the region of the developing SCO in E, and in the SCO itself in F,G. Scale bars: 500 μ m for A-D; 100 μ m for E-G. cx, cerebral cortex; ch, choroid plexus; DT, dorsal thalamus; HY, hypothalamus; me, mesencephalon; cb, cerebellum; P, pituitary; ST, striatum; ep, epiphysis.

The only RFX4_v3 positive structures noted outside of the central nervous system were the trigeminal and facial/vestibular ganglia (Fig. 8I) and the anterior pituitary (Fig. 9B).

Phenotype of RFX4_v3-deficient mice

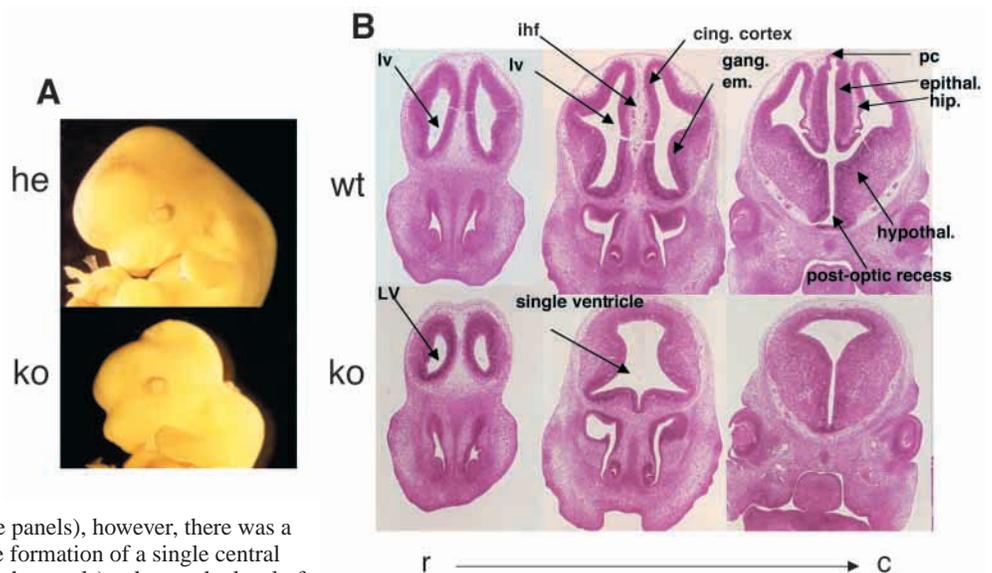
Surviving transgenic mice, which we will now refer to as RFX4_v3 +/- mice, were interbred to generate -/- mice. Ten pregnant +/- mice were allowed to carry to term and deliver; the average litter size of these pregnancies was 5.3±0.6, which was significantly smaller than litters from a control line 7.0±0.4 ($P=0.022$). Of 53 pups born, 19 (36%) were wild type, 28 (53%) +/-, and 6 (11%) -/-, suggesting substantial intrauterine or perinatal loss of the -/- pups. All of the -/- pups born died within 1 hour of birth. Seven additional litters were obtained between E8 and E18. The average size of those litters was 8.7±0.5, which was not significantly different from control litters. Of 61 pups obtained, there were 10 (16%) +/+, 36 (59%) +/- and 15 (25%) -/-, indicating no excess intrauterine mortality.

The brains of the -/- mice at the time of birth and at E16.5 were grossly dysmorphic (data not shown). We therefore examined the -/- mice at an earlier developmental stage, E12.5. The phenotype at this age was striking (Fig. 10). Externally, there were clear abnormalities of head appearance, although the position of the eyes, vibrissae and other facial structures appeared relatively normal (Fig. 10A). Coronal sections suggested that dorsal structures in the rostral brain were hypoplastic and lacked morphological differentiation of medial and paramedial dorsal structures. This was most striking in the forebrain and midbrain (Fig. 10B), but abnormalities persisted into the hindbrain and spinal cord. As in the hemizygotes, the anatomy of the rest of the body in the E12.5 -/- embryos was apparently normal.

To characterize the patterning of the mutant brains we analyzed the expression of genes that play important roles in regionalization (Marin and Rubenstein, 2002; Yun et al., 2001).

Fig. 10. Head morphology from -/- mice at E12.5. (A) Heads from two E12.5 littermates after fixation, one hemizygous (he) and one knockout (ko) as indicated. Note the near normal appearance of the eyes and the facial structures, but the clearly abnormal doming of the skull and the smaller head of the -/- littermate.

(B) Coronal sections from wild-type (top row) and knockout (bottom row) littermate mice at E12.5, stained with Hematoxylin and Eosin. In the most rostral (r) sections (left panels), the brains appear somewhat similar, showing both lateral ventricles (LV) and apparently normal midline structures, although the brains were somewhat smaller in the knockout mice. In more caudal (c) sections (middle panels), however, there was a striking loss of midline structures and the formation of a single central ventricle. In still more caudal sections (right panels), taken at the level of the retinas, there were continued striking abnormalities and loss of essentially all dorsal midline structures in the knockout mice. ihf, interhemispheric fissure; cing. cortex, cingulate cortex; gang. em., ganglionic eminence; pc, posterior commissure; epithal., epithalamus; hip., hippocampus; hypothal., hypothalamus. See the text for further details.



Our analysis was mainly focused on the telencephalon of E12.5 -/- embryos (Fig. 11). The lateral walls of the telencephalic vesicles primarily consist of the basal ganglia (rostroventral) and the cerebral cortex (caudodorsal). The rostral and rostradorsal midline is constituted by the commissural plate and adjacent parts of the septal area; the caudodorsal midline consists of the choroid plexus and the cortical hem. The cortical hem is a Wnt- and Bmp-rich signaling center in the dorsomedial telencephalon that has been shown to be crucial in cortical development (Furuta et al., 1997; Galceran et al., 2000; Grove et al., 1998; Lee et al., 2000).

Expression of the telencephalic marker *Foxg1* (*Bfl*) was maintained in the cortex and basal ganglia of RFX4_v3 mutants (Fig. 11A). The expression of markers specific for midline structures, the cerebral cortex and the basal ganglia revealed that the principal telencephalic defects in RFX4_v3 mutants involved severe hypoplasia of the dorsal midline and adjacent cerebral cortex (Fig. 11). The lack of dorsal midline structures was demonstrated by the loss of *Wnt3a*, *Wnt7b* and *Bmp4* expression in the hem (Fig. 11E,F and not shown) and the reduction of *Msx2* expression in the hem and choroid plexus (Fig. 11D). The cerebral cortex was present, based on the expression of *Wnt7b*, *Emx1*, *Pax6* and *Lhx2* (Fig. 11F-I); however, it was severely hypoplastic. Despite the severe hypoplasia, the cortex did produce post-mitotic cells, based on the mantle zone expression of *Wnt7b* (Fig. 11F).

In wild-type mice, *Lhx2* and *Emx1* are expressed in a dorsoventral gradient in the cortical neuroepithelium. In the RFX4_v3 mutants, *Lhx2* and *Emx1* expression levels were similar to those seen in the ventral part of the normal cortex, suggesting that dorsal parts of the cortex were missing (Fig. 11G,I). An *Emx1*-negative, *Lhx2*-positive territory intercalated between the striatum and the prospective piriform cortex, which develops into parts of the claustramygdaloid complex (Puelles et al., 2000; Yun et al., 2001), was maintained in the mutants (Fig. 11G,I). Finally, *Pax6* is normally detected in a

ventrodorsal gradient. In the mutants, the ventral area where expression was strongest was detected (Fig. 11H). Thus, the most ventral subdivisions of the cortex, located adjacent to the striatum, i.e. the piriform cortex and parts of the claustramygdaloid complex, seemed to be correctly specified, while the most medial cortical subdivisions, located adjacent to the cortical hem, i.e. the hippocampus and the neocortex, are either severely reduced, lost, or mis-specified.

The basal ganglia are formed in mammals by the lateral ganglionic eminence, which develops into the striatum, and the medial ganglionic eminence, which develops into the pallidum (Marin and Rubenstein, 2002). In the mutants, although the size of the basal ganglia was disproportionately large compared with the cortex, it is unclear whether or not there was an absolute increase in the sizes of the lateral and medial ganglionic eminences. The RFX4_v3 mutants exhibited normal expression of *Dlx2* and *Six3* transcription

factors in the lateral and medial ganglionic eminences (Fig. 11J,K). Expression of *Otx2*, *Fgf8* and *Six3* in the septum, a basal ganglia-related structure, was also detected (Fig. 11B,C,J). In addition, the specific expression of the transcription factor *Nkx2.1* in the medial ganglionic eminence and ventral septum was apparently normal in the mutants (Fig. 11L).

Discussion

These findings demonstrate that the transgene-interrupted expression of this novel RFX4 transcript is responsible for a dose-dependent brain phenotype: hydrocephalus associated with hypoplasia or absence of the SCO in the hemizygous mice, and severe and lethal defects of midline brain structure formation in the homozygotes. These data suggest that a quantitative decrease in the expression of the RFX4_v3

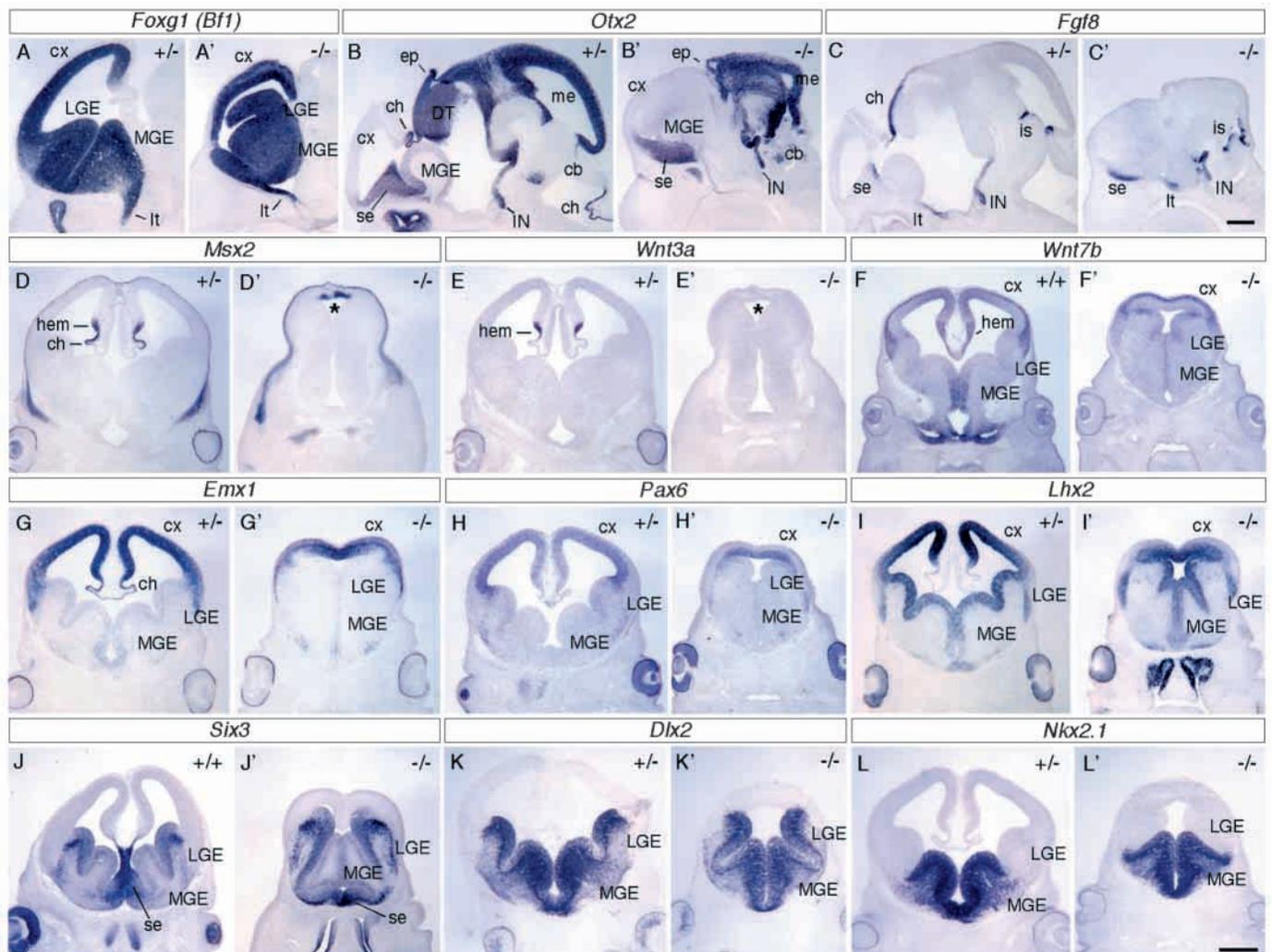


Fig. 11. Expression of molecular markers in wild-type and littermate knockout mice at E12.5. Shown are the in situ hybridization staining patterns of sagittal (A–C) and coronal sections through wild-type (+/– or +/+) and knockout (–/–) heads at E12.5. The blue digoxigenin staining indicates the presence of the specific transcript being evaluated. Note that *Fgf8* expression is maintained in the isthmus, infundibulum, lamina terminalis and septum, but is lost in the choroid plexus of the forebrain (C,C'). The asterisks in D' and E' indicate the decrease in *Msx2* expression (D,D') and the lack of *Wnt3a* expression (E,E') in the dorsal midline of the knockout embryos. Scale bars: 500 μ m. LGE, lateral ganglionic eminence; MGE, medial ganglionic eminence; ch, choroidal plexus; cx, cerebral cortex; ep, epiphysis; IN, infundibulum; lt, lamina terminalis; is, isthmus; hem, cortical hem; DT, dorsal thalamus; se, septum; me, mesencephalon; cb, cerebellum.

transcript is sufficient to interfere rather specifically with the development of the SCO, leading to effective stenosis of the aqueduct of Sylvius and congenital hydrocephalus. This partial RFX4_v3 deficiency is nonetheless compatible with postnatal life, and in some cases with successful fertility. By contrast, complete deficiency of this transcript leads to catastrophic failure of dorsal midline structure formation in early brain development, and universal prenatal or perinatal death. These studies identify RFX4_v3 as a key, early regulator of midline brain structure development in the mouse.

The RFX proteins belong to the winged-helix subfamily of helix-turn-helix transcription factors, and are so named because they bind to 'X-boxes' in target DNA sequences and regulate expression of the target genes. The X-box consensus sequence is 5'-GTNRCC(0-3N)RGYAAAC-3', where N is any nucleotide, R is a purine and Y is a pyrimidine. Five RFX proteins have been described in man (RFX1-RFX5), all of which contain a highly conserved DNA binding domain near the N terminus. A structure has been determined for the binding of this domain from RFX1 to an X-box sequence (Gajiwala et al., 2000); this shows that the 'wing' of this DNA binding domain is used to recognize DNA. Members of this family have been implicated in the transcriptional regulation of a number of important genes.

A partial sequence of a novel family member, termed RFX4, was initially identified by Dotzlaw et al. (Dotzlaw et al., 1992) as part of a fusion cDNA in human breast cancers, in which the N-terminal estrogen binding domain of the estrogen receptor was fused with the RFX DNA binding domain. More recently, two full-length RFX4 cDNAs have been described and categorized, and their relationships and nomenclature were updated by NCBI on 26 March 2003. The new RFX4_v3 variant described here is composed of novel exons as well as exons derived from one or both of these two earlier variants. As illustrated in Fig. 6, the RFX4_v3 cDNA is the largest of the three and is composed of a unique 5' exon of ~476 bp that encodes the first 14 amino acids of RFX4_v3; this is then followed by four exons shared only with RFX4_v2, then 10 exons shared with both RFX4_v1 and RFX4_v2, and finally three 3'-exons shared only with RFX4_v1. The existence of this transcript in mouse brain was confirmed while this paper was under revision by the deposition in GenBank on 5 December 2002, of a 3535 bp cDNA isolated from adult mouse diencephalon that is 417 bp shorter but otherwise essentially identical to the RFX4_v3 transcript described here. However, at the time of this writing (21 March 2003), there were no human cDNA sequences in either the GenBank nr or est collections corresponding to the unique 5'-end of RFX4_v3, other than the sequence described here.

Our data indicate that the novel RFX4_v3 transcript is expressed in the developing central nervous system from the neural plate stages. Its early expression is dynamic, particularly in the telencephalon, where initially it is only expressed in and adjacent to the dorsal midline. Later, its expression is extinguished in the midline, and spreads as a dorsoventral gradient throughout the cortex. It is also expressed in adult brain, although its non-developmental functions and anatomical distribution in this tissue remain to be determined. We did not detect significant levels of this transcript in other organs of the adult mouse. It will be of interest in future studies to identify genes whose expression is directly affected by

RFX4_v3, presumably acting as a transcription factor, as well as other transcription factors influencing the developmental expression of RFX4_v3 itself.

Disruption of both RFX4_v3 alleles severely altered early brain morphogenesis. The reduction of *Msx2*, and the loss of *Wnt3a*, *Wnt7b* and *Bmp4* expression in the cortical hem, strongly suggest that RFX4 is required either to establish or maintain the dorsal patterning center of the telencephalon. Mice deficient in WNT signaling from the cortical hem have defects in growth and patterning of the dorsomedial cerebral cortex (Galceran et al., 2000; Lee et al., 2000). Although these WNT-signaling mutants have hippocampal defects, they can generate a choroid plexus. Mice deficient in BMP-signaling, through the loss of *Bmpr1a* function in the telencephalon, fail to produce choroid plexus (Hebert et al., 2002). Given the loss of the choroid plexus in RFX4_v3 mutants, and the loss of *Bmp4* and the reduction of *Msx2* expression, we suggest that RFX4 has a general role in regulating dorsal patterning that involves both WNT and BMP signaling pathways.

The hypoplasia of the cerebral cortex could be entirely due to defects in the dorsal patterning center. However, RFX4 is expressed throughout the cortical primordium, and therefore could have a role in regulating proliferation and differentiation of the cerebral cortex, similar to *Foxg1* (*Bfl1*), another winged-helix gene (Dou et al., 1999; Hanashima et al., 2002). Future studies should aim to elucidate how RFX4 regulates the dorsal patterning center and to establish its more general role within neural progenitors.

Disruption of a single RFX4_v3 allele led to a quantitative decrease in RFX4_v3 mRNA expression in the brain and non-communicating congenital hydrocephalus. Hydrocephalus is generally being divided into congenital forms, i.e. those present at birth, and acquired forms that develop after birth. Within the spectrum of congenital hydrocephalus are non-genetic causes, such as uterine infection, hemorrhage and meningitis, as well as genetic causes. Congenital hydrocephalus can be subdivided further into communicating and non-communicating forms, in which the latter is associated with stenosis of the aqueduct of Sylvius. In one series of individuals with isolated congenital hydrocephalus, i.e. not associated with other congenital anomalies, about 43% were associated with aqueductal obstruction, 36% had communicating hydrocephalus, 15% had the Dandy-Walker syndrome and about 6% had other lesions (Burton, 1979). Overall, isolated hydrocephalus in man occurs in ~0.6 per 1000 of newborn children (Halliday et al., 1986). It is thought that the X-linked form is present in 7-27% of male cases; this form is now known to be due to abnormalities in the *LICAM* gene, and there are several overlapping neurodevelopmental human syndromes associated with defects in this gene (Weller and Gartner, 2001). Most cases of congenital hydrocephalus, however, have no known genetic cause. The unusual finding that expression of only a single allele leads to congenital hydrocephalus, at least in mice, means that this defect exhibits an autosomal dominant inheritance pattern. As in mice, expression of a single intact allele may be compatible with life and fertility in humans. We are currently exploring the possibility that some human cases of congenital, non-L1 hydrocephalus are due to abnormalities in the expression or sequence of the RFX4_v3 transcript.

Hydrocephalus in these mice was associated with the apparent absence of the SCO. Abnormalities of the SCO have

been associated with hydrocephalus in many studies, as recently reviewed (Perez-Figares et al., 2001). Although there has been some debate in the literature about whether the SCO abnormalities cause or are consequences of the hydrocephalus, the overall consensus seems to be in favor of the SCO abnormalities preceding and causing the hydrocephalus, owing to effective stenosis of the aqueduct of Sylvius. Examples of damage to or abnormalities of the SCO causing hydrocephalus include radiation during fetal life (Takeuchi and Takeuchi, 1986), maternal transfer of antibodies to Reissner's fibers (Vio et al., 2000), congenital absence of the SCO in the MT/HokIdr strain of mice (Takeuchi et al., 1987), hypoplasia of the SCO in SUMS/np mice (Jones et al., 1987) and hypoplasia of the SCO in two strains of rats with congenital hydrocephalus (Jones and Bucknall, 1988; Takeuchi et al., 1988). It therefore seems likely that, in the present case, the aplasia or hypoplasia of the SCO seen in the RFX4_v3 hemizygous mice is the cause of the congenital hydrocephalus, presumably by interfering with cerebrospinal fluid flow through the rostral part of the aqueduct. It will be of interest to determine whether any of these previously described mutants have abnormalities in the expression or sequence of the RFX4_v3 protein. In addition, examination of downstream gene expression at this specific anatomical site in these hemizygous mice may lead to new insights into the formation of this 'enigmatic secretory gland of the brain' (Schoniger et al., 2001).

We thank Laura Miller, J. Alyce Bradbury, Tom Sliwa and Jane Tuttle for technical assistance, and Drs Yuji Mishina and Jean Harry for helpful comments on the manuscript. I.C. was supported by the Spanish Ministry of Education and Culture. J.L.R.R. was supported by NINDS grant RO1 NS34661, NIMH grant K02 MH01046 and Nina Ireland.

References

- Blackshear, P. J., Lai, W. S., Tuttle, J. S., Stumpo, D. J., Kennington, E., Nairn, A. C. and Sulik, K. K. (1996). Developmental expression of MARCKS and protein kinase C in mice in relation to the exencephaly resulting from MARCKS deficiency. *Dev. Brain Res.* **96**, 62-75.
- Burton, M. K. (1979). Recurrence risks for congenital hydrocephalus. *Clin. Genet.* **16**, 47-53.
- Cifuentes, M., Rodriguez, S., Perez, J., Grondona, J. M., Rodriguez, E. M. and Fernandez-Llebrez, P. (1994). Decreased cerebrospinal fluid flow through the central canal of the spinal cord of rats immunologically deprived of Reissner's fibre. *Exp. Brain Res.* **98**, 431-440.
- D'Arcangelo, G., Miao, G. G., Chen, S. C., Soares, H. D., Morgan, J. I. and Curran, T. (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374**, 719-723.
- D'Arcangelo, G., Miao, G. G. and Curran, T. (1996). Detection of the reelin breakpoint in reeler mice. *Mol. Brain Res.* **39**, 234-236.
- Dotzlaw, H., Alkhalaf, M. and Murphy, L. C. (1992). Characterization of estrogen receptor variant mRNAs from human breast cancers. *Mol. Endocrinol.* **6**, 773-785.
- Dou, C. L., Li, S. and Lai, E. (1999). Dual role of brain factor-1 in regulating growth and patterning of the cerebral hemispheres. *Cereb. Cortex* **9**, 543-550.
- Durkin, M. E., Keck-Waggoner, C. L., Popescu, N. C. and Thorgeirsson, S. S. (2001). Integration of a c-myc transgene results in disruption of the mouse Gtf2ird1 gene, the homologue of the human GTF2IRD1 gene hemizygotously deleted in Williams-Beuren syndrome. *Genomics* **73**, 20-27.
- Friedman, R. A., Adir, Y., Crenshaw, E. B., Ryan, A. F. and Rosenfeld, M. G. (2000). A transgenic insertional inner ear mutation on mouse chromosome 1. *Laryngoscope* **110**, 489-496.
- Furuta, Y., Piston, D. W. and Hogan, B. L. (1997). Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development* **124**, 2203-2212.
- Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. and Burley, S. K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916-921.
- Galceran, J., Miyashita-Lin, E. M., Devaney, E., Rubenstein, J. L. and Grosschedl, R. (2000). Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. *Development* **127**, 469-482.
- Grove, E. A., Tole, S., Limon, J., Yip, L. and Ragsdale, C. W. (1998). The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. *Development* **125**, 2315-2325.
- Halliday, J., Chow, C. W., Wallace, D. and Danks, D. M. (1986). X linked hydrocephalus: a survey of a 20 year period in Victoria, Australia. *J. Med. Genet.* **23**, 23-31.
- Hanashima, C., Shen, L., Li, S. C. and Lai, E. (2002). Brain factor-1 controls the proliferation and differentiation of neocortical progenitor cells through independent mechanisms. *J. Neurosci.* **22**, 6526-6536.
- Hebert, J. M., Mishina, Y. and McConnell, S. K. (2002) BMP signaling is required locally to pattern the dorsal telencephalic midline. *Neuron* **35**, 1029-1041.
- Jones, H. C. and Bucknall, R. M. (1988). Inherited prenatal hydrocephalus in the H-Tx rat: a morphological study. *Neuropathol. Appl. Neurobiol.* **14**, 263-274.
- Jones, H. C., Dack, S. and Ellis, C. (1987). Morphological aspects of the development of hydrocephalus in a mouse mutant (SUMS/np). *Acta Neuropathol.* **72**, 268-276.
- Lee, S. M., Tole, S., Grove, E. and McMahon, A. P. (2000). A local Wnt-3a signal is required for development of the mammalian hippocampus. *Development* **127**, 457-467.
- Marin, O. and Rubenstein, J. L. (2002). Patterning, regionalization and cell differentiation in the forebrain. In *Mouse Development* (ed. J. Rossant and P. Tam), pp. 75-106. London: Academic Press.
- Miao, G. G., Smeyne, R. J., D'Arcangelo, G., Copeland, N. G., Jenkins, N. A., Morgan, J. I. and Curran, T. (1994). Isolation of an allele of reeler by insertional mutagenesis. *Proc. Natl. Acad. Sci. USA* **91**, 11050-11054.
- Morotomi-Yano, K., Yano, K., Saito, H., Sun, Z., Iwama, A. and Miki, Y. (2002). Human regulatory factor X 4 (RFX4) is a testis-specific dimeric DNA-binding protein that cooperates with other human RFX members. *J. Biol. Chem.* **277**, 836-842.
- Overbeek, P. A., Gorlov, I. P., Sutherland, R. W., Houston, J. B., Harrison, W. R., Boettger-Tong, H. L., Bishop, C. E. and AgoulNIK, A. I. (2001). A transgenic insertion causing cryptorchidism in mice. *Genesis* **30**, 26-35.
- Perez-Figares, J. M., Jimenez, A. J., Perez-Martin, M., Fernandez-Llebrez, P., Cifuentes, M., Riera, P., Rodriguez, S. and Rodriguez, E. M. (1998). Spontaneous congenital hydrocephalus in the mutant mouse hyl. Changes in the ventricular system and the subcommissural organ. *J. Neuropathol. Exp. Neurol.* **57**, 188-202.
- Perez-Figares, J. M., Jimenez, A. J. and Rodriguez, E. M. (2001). Subcommissural organ, cerebrospinal fluid circulation, and hydrocephalus. *Microsc. Res. Tech.* **52**, 591-607.
- Puelles, L., Kuwana, E., Puelles, E., Bulfone, A., Shimamura, K., Keleher, J., Smiga, S. and Rubenstein, J. L. (2000). Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes *Dlx-2*, *Emx-1*, *Nkx-2.1*, *Pax-6*, and *Tbr-1*. *J. Comp. Neurol.* **424**, 409-438.
- Rice, D. S. and Curran, T. (2001). Role of the reelin signaling pathway in central nervous system development. *Annu. Rev. Neurosci.* **24**, 1005-1039.
- Rodriguez, E. M., Oksche, A., Hein, S., Rodriguez, S. and Yulis, R. (1984). Comparative immunocytochemical study of the subcommissural organ. *Cell Tiss. Res.* **237**, 427-441.
- Rodriguez, E. M., Oksche, A. and Montecinos, H. (2001). Human subcommissural organ, with particular emphasis on its secretory activity during the fetal life. *Microsc. Res. Tech.* **52**, 573-590.
- Rodriguez, E. M., Rodriguez, S. and Hein, S. (1998). The subcommissural organ. *Microsc. Res. Tech.* **41**, 98-123.
- Schoniger, S., Wehming, S., Gonzalez, C., Schobitz, K., Rodriguez, E., Oksche, A., Yulis, C. R. and Nurnberger, F. (2001). The dispersed cell culture as model for functional studies of the subcommissural organ: preparation and characterization of the culture system. *J. Neurosci. Methods* **107**, 47-61.
- Takeuchi, I. K., Kimura, R., Matsuda, M. and Shoji, R. (1987). Absence of subcommissural organ in the cerebral aqueduct of congenital hydrocephalus spontaneously occurring in MT/HokIdr mice. *Acta Neuropathol.* **73**, 320-322.

- Takeuchi, I. K., Kimura, R. and Shoji, R.** (1988). Dysplasia of subcommissural organ in congenital hydrocephalus spontaneously occurring in CWS/Idr rats. *Experientia* **44**, 338-340.
- Takeuchi, I. K. and Takeuchi, Y. K.** (1986). Congenital hydrocephalus following X-irradiation of pregnant rats on an early gestational day. *Neurobehav. Toxicol. Teratol.* **8**, 143-150.
- Tsuchida, T., Ensini, M., Morton, S. B., Baldassare, M., Edlund, T., Jessell, T. M. and Pfaff, S. L.** (1994). Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957-970.
- Vio, K., Rodriguez, S., Navarrete, E. H., Perez-Figares, J. M., Jimenez, A. J. and Rodriguez, E. M.** (2000). Hydrocephalus induced by immunological blockage of the subcommissural organ-Reissner's fiber (RF) complex by maternal transfer of anti-RF antibodies. *Exp. Brain Res.* **135**, 41-52.
- Weller, S. and Gartner, J.** (2001). Genetic and clinical aspects of X-linked hydrocephalus (L1 disease): Mutations in the L1CAM gene. *Hum. Mutat.* **18**, 1-12.
- Wilkinson, D. G.** (1992). Whole mount in situ hybridization on vertebrate embryos. In *In Situ Hybridization: A Practical Approach* (ed. D. G. Wilkinson), pp. 75-83. Oxford: IRL Press.
- Yun, K., Potter, S. and Rubenstein, J. L.** (2001). Gsh2 and Pax6 play complementary roles in dorsoventral patterning of the mammalian telencephalon. *Development* **128**, 193-205.