Available online at www.sciencedirect.com

SCIENCE DIRECT

Developmental Biology 282 (2005) 509 – 523

DEVELOPMENTAL BIOLOGY

www.elsevier.com/locate/ydbio

Genomes & Developmental Control

A consensus Oct1 binding site is required for the activity of the *Xenopus Cdx4* promoter

John S. Reece-Hoyes[1], Iain D. Keenan[2], Mary Elizabeth Pownall, Harry V. Isaacs*

Area 11, Department of Biology, University of York, York YO10 5YW, UK

Received for publication 22 November 2004, revised 9 March 2005, accepted 14 March 2005
Available online 25 April 2005

Abstract

Cdx homeodomain transcription factors have multiple roles in early vertebrate development. Furthermore, mis-regulation of Cdx expression has been demonstrated in metaplasias and cancers of the gut epithelium. Given the importance of Cdx genes in development and disease, the mechanisms underlying their expression are of considerable interest. We report an analysis of the upstream regulatory regions from the amphibian *Xenopus laevis Cdx4* gene. We show that a GFP reporter containing 2.8 kb upstream of the transcription start site is expressed in the posterior of transgenic embryos. Deletion analysis of the upstream sequence reveals that a 247-bp proximal promoter fragment will drive posterior expression in transgenic embryos. We show that 63 bp of upstream sequence, that includes a consensus site for POU-domain octamer-binding proteins, retains significant promoter activity. Co-expression of the octamer-binding protein Oct1 induces expression from a *Cdx4* reporter and mutation of the octamer site abolishes activity of the same reporter. We show that the octamer site is highly conserved in the promoters of the human, mouse, chicken, and zebrafish *Cdx4* genes and within the promoters of amphibian *Cdx1* and *Cdx2*. These data suggest a conserved function for octamer-binding proteins in the regulation of Cdx family members.

Keywords: Xenopus; Cdx4; POU-domain; Oct1

Introduction

The Cdx genes code for a family of homeodomain transcription factors that have been shown to be present in all animal groups examined. The genomes of most vertebrate species, including human, mouse, chick, and frog, contain three Cdx genes. In humans and mice, these genes are designated *Cdx1*, *Cdx2*, and *Cdx4*. The Cdx genes of the amphibian *Xenopus laevis* were originally designated *Xcad1*, *Xcad2*, and *Xcad3* (Pillemer et al., 1998a).

Cdx transcription factors have been shown to play vital roles in the assignment of regional identity along the main body axis of both invertebrates (Dick and Buss, 1994; Hunter and Kenyon, 1996; McDonald and Struhl, 1986) and vertebrates (Dick and Buss, 1994; Epstein et al., 1997; Hunter and Kenyon, 1996; MacDonald and Struhl, 1986; Marom et al., 1997; Pownall et al., 1996; Subramanian et al., 1995). An important finding is that Cdx factors regulate the expression of a subset of vertebrate Hox genes (Charité et al., 1998; Epstein et al., 1997; Isaacs et al., 1998; Pownall et al., 1996, 1998; Subramanian et al., 1995), which are themselves highly conserved, developmental regulatory genes with functions associated with the assignment of segmental identity in arthropods and vertebrates (Krumlauf, 1994).

In addition to a role in early axial patterning, Cdx genes are involved in regulating the proliferation and differentiation of the epithelial lining of the mature vertebrate gut (Chawengsaksophak et al., 1997; Ishii et al., 1997; Levy et al., 1995; Lynch et al., 2000; Tamai et al., 1999). Recent evidence suggests that abnormal Cdx gene expression is associated with a number of disease conditions. Ectopic expression of *Cdx1* and *Cdx2* has been detected in intestinal

* Corresponding author. Fax: +44 1904 328505.
E-mail address: hvi1@york.ac.uk (H.V. Isaacs).
[1] Present address: Department of Biology, University of Leeds, Leeds LS2 9JT, UK.
[2] Present address: Molecular Pharmacology Unit, Cancer Research UK, Biomedical Research Centre, Ninewells Hospital, Dundee DD1 5SY, UK.

0012-1606/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2005.03.035

metaplasias of the esophagus and stomach (Bai et al., 2002; Phillips et al., 2003; Silberg et al., 1997). Furthermore, reduction in expression of *Cdx1* and *Cdx2* has been reported in intestinal-type adenocarcinomas of the stomach (Guo et al., 2004), suggesting that Cdx genes act as tumor suppressors in this context. Ectopic expression of *Cdx2* has also been implicated as the transforming event in a mouse model of acute myeloid leukemia (Rawat et al., 2004).

Given the medical importance of Cdx genes and their crucial roles in early development, the mechanisms involved in regulating the expression of this gene family are of considerable interest. The control of Cdx gene expression is complex, and a number of signaling pathways, which are active during early development, have been implicated in their regulation (Lohnes, 2003). Studies in mice have shown the importance of both Wnt and retinoic acid (RA) signaling in the regulation of *Cdx1* (Ikeya and Takada, 2001; Prinos et al., 2001). Work in *Xenopus* indicates a role for signaling by BMP, FGF, RA, and Wnt in regulating the expression of *Xcad2* and *Xcad3* (Haremaki et al., 2003; Northrop and Kimelman, 1994; Pillemer et al., 1998b; Pownall et al., 1996; Shiotsugu et al., 2004).

The analysis of the promoter regions from murine *Cdx1* and *Cdx2* has revealed a number of features. Consistent with a regulatory role for RA and Wnt signals, retinoic acid response elements (RAREs) and elements responsive to Tcf/Lef, a downstream component of the canonical Wnt signaling pathway, have been identified in the proximal promoter region of *Cdx1* (Houle et al., 2000; Lickert et al., 2000, 2002; Prinos et al., 2001).

Data indicate a role for auto- and cross-regulatory interactions among the Cdx genes during the establishment and maintenance of the Cdx gene expression pattern (Chawengsaksophak et al., 2004; Isaacs et al., 1998; Prinos et al., 2001). The presence of Cdx responsive elements in the regulatory regions of *Cdx1* and *Cdx2* supports this notion (Beland et al., 2004; Chawengsaksophak et al., 2004; Xu et al., 1999). As yet, there has been no reported analysis of the gene regulatory regions of murine *Cdx4*.

A recent study of the *Xenopus Xcad3* gene indicates that FGF and Wnt responsive elements, located within the first intron, are required for the expression of this gene in the developing posterior nervous system (Haremaki et al., 2003). A previous phylogenetic analysis of Cdx peptide sequences indicated that Xcad3 might represent the amphibian orthologue of amniote Cdx4 (Pillemer et al., 1998a). Here, we present an analysis of the amphibian *Xcad3* genomic locus which strongly supports this notion, and for consistency with studies in other vertebrate species, we will now refer to *Xcad3* as *Cdx4*.

In this study, we report an analysis of the upstream regulatory regions of the *X. laevis Cdx4* promoter. We analyze a 2.8-kb region, immediately 5′ to the transcriptional start site, which is capable of driving robust expression of a reporter in the posterior of transgenic embryos. We show that the *Cdx4* promoter lacks a TATA box but contains a consensus initiator sequence and a downstream promoter element (DPE) close to the putative transcription start site. Our analysis in transgenic animals reveals that a fragment containing 237 bp of proximal promoter sequence can drive posterior expression in the tail-forming region from early neurula stages and that 92 bases bp of upstream sequence will drive posterior expression in tailbud embryos. A smaller deletion construct containing just 63 bp of upstream sequence loses the ability to drive posterior expression but retains significant promoter activity, which is dependent upon the presence of a consensus binding motif for the POU domain protein Oct1. We show that the Oct1 binding site within the *Cdx4* proximal promoter region has been highly conserved during vertebrate evolution. We also show that a consensus Oct binding motif is found in the promoters of *Xenopus Cdx1* and *Cdx2*. Our data indicate that POU domain proteins are likely to play crucial roles in regulating promoter activity of the vertebrate Cdx gene family.

Materials and methods

Embryo culture

Injected and transgenic embryos were cultured to appropriate stages in either NAM/3 + 5% Ficoll or NAM/10 + 5% Ficoll. Embryos were staged according to Nieuwkoop and Faber (1967). Animal cap explants were removed from embryos at stages 8–10 and cultured in NAM/2 until the appropriate stage.

Isolation of the upstream regulatory region of Cdx4

15 μg of *X. laevis* genomic DNA prepared from red blood cells was digested with a number of restriction enzymes including *Pvu*II and Southern blotted onto Hybond-N membrane (Amersham). A ^{32}P-labeled exon 1 specific probe consisting of a 400-bp *Eco*RI fragment from the 5′ end of the *X. laevis Cdx4* (*Xcad3*) cDNA reported by Northrop and Kimelman (1994) was used to probe the blot. Hybridization was carried out at 65°C overnight and washes carried out according to the method of Church and Gilbert (1984). Autoradiography of the membrane was carried out overnight. Accurate assessment of the molecular weight of hybridizing DNA bands was made using appropriate markers. A sample of similarly digested genomic DNA was run out on another agarose gel and the regions corresponding to the size of hybridizing bands of interest were removed. The DNA from the gel slice was isolated and subcloned into *Eco*RV-digested Bluescript KS+ (Stratagene). Ligations were used to transform electrocompetent cells of the JM109 strain of *E. coli*. Recombinant colonies were screened for the presence of the insert of interest by colony hybridization using the same probe used for the

original Southern blot (Sambrook et al., 1989). A 3400-bp hybridizing clone (pBS4c3.4) was fully sequenced on both strands. The sequence has been deposited with GenBank (accession number AY818613).

Production of synthetic mRNA for injections

Capped mRNAs were synthesized using the SP6 Megascript transcription kit (Ambion) and a modified protocol using 0.5 mM GTP and 5 mM $m^7G(5')Gppp(5')G$ cap. Following synthesis, RNAs were subject to sequential precipitation with 0.5 M ammonium acetate and 2.5 M LiCl to remove unincorporated nucleotides. The Cs2-FGF8 plasmid was linearized with *Not*I and transcribed with SP6 (Christen and Slack, 1997). The *Xenopus tropicalis Oct1* clone was obtained from MRC Geneservice (GenBank accession number AL849985). CS107 *Oct1* plasmid was linearized with *Not*I and transcribed with SP6 polymerase.

Construction of reporter plasmids

4GO was constructed by removing 400 bp, corresponding to the *Cdx4* ORF, from the 3′ end of pBS-4c3.4. The GFP2 ORF was subsequently cloned downstream of the promoter fragment to generate *4GO*. The *Cdx4* ORF was excised using *Sty*I/*Bam*HI restriction digest, end-filling with Klenow enzyme and religating the plasmid. The GFP2 ORF and a downstream SV40 polyadenylation site was then inserted downstream of the *Cdx4* 3-kb promoter and 5′ UTR fragment. The GFP2/SV40 cassette was removed from Cs2+-GFP2 by digesting with *Bam*HI and *Not*I. The *Bam*HI overhang was blunted with Klenow enzyme and the fragment directionally cloned into the *Sty*I (blunted)/*Not*I site of pB-4c3.4, now designated *4GO*.

4LO was constructed by transferring the 3.4-kb insert from pBS-4c3.4 into pGL3-BASIC (Promega). The insert was excised using *Kpn*I/*Xba*I and the vector digested using *Kpn*I/*Nhe*I. The *Cdx4* ORF was removed by the PCR-based protocol described below, using the following primers:

4-T7A	5′-CTATACTAGATCCTTGGTGG-3′
4-T3G	5′-ACACTCTGTAGATTTGACTC-3′

A PCR-based approach was used to construct the various deletions and mutations of both GFP and Luciferase reporter plasmids. PCR was carried out with a high-fidelity, long-range DNA polymerase (Pfu Turbo, Stratagene). Primers were designed to amplify a PCR product of plasmid backbone with varying amounts of insert at either end. The product was kinased and religated to generate the relevant reporter deletion or mutation. Following transformation of bacteria with plasmid DNA, the sequence of the promoter fragment was confirmed. All deletions of both *4GO* and *4LO* used the anchor primer 4-T7H, while the position of other primer defined the extent of the deletion. These are the primers used in the generation of both 4GFP (*4G*) and 4Luciferase (*4L*) deletions.

4-T7H (anchor)	5′-CAGTCATGTTATTTTGTGTGC-3′
4GI	5′-ATATAGTAAATAGGGGATGCC-3′
4GH	5′-TAGAGAGCATTTATATGCCC-3′
4GG	5′-TACCTCGAATAAACTCATACCCTC-3′
4GF	5′-TATTTCCATGCACGGGTTCC-3′
4GE and *4LE*	5′-CCGGTAATCAAGGTTTCTG-3′
4GD and *4LD*	5′-TCCATAATAGACTGAAACA-3′
4GC and *4LC*	5′-CTGAAAGTAGCAACAAGG-3′
4GB and *4LB*	5′-GGATGTCCAGCCTTTTATTTGAATGC
4GA and *4LA*	5′-GGGCAGGCAGTTATCTCTG-3′

PCR for generating the POU site mutants of *4LE* and *4LB* (*4LE-ΔPOU* and *4LB-ΔPOU*) used the appropriate plasmid as template and the primers 4-T3P and 4-T7P.A The underlined section of 4-T3P indicates the mutation of the POU site ATGCAAAT to **C**TGCA**G**AT.

4-T3P	5′-CTGCAGATGTATTTGGGGGGC-3′
4-T7P.A	5′-TCAAATAAAAGGCTGGACATC-3′

Transgenesis

Transgenic embryos containing the *X. laevis Cdx4* GFP reporter constructs were produced either using the *R*estriction *E*nzyme *M*ediated *I*ntegration (REMI) method (Kroll and Amaya, 1996) or the non-REMI transgenic protocol (Sparrow et al., 2000). In both cases, the DNA constructs were linearized with either *Acc*65I or *Not*I restriction enzymes before being added to isolated sperm heads. Successfully cleaving embryos were isolated and cultured to the appropriate stage before being examined for GFP fluorescence and processed for in situ hybridization.

Luciferase assays on embryos injected with DNA and RNA constructs

Prior to injection, embryos were transferred to NAM/3 + 5% Ficoll. Embryos were injected with the stated amounts of DNA and RNA in a total volume of 20 nl per embryo at either the 2- or 4-cell stage. 1 pg of the Renilla luciferase CMV promoter plasmid was coinjected into each embryo in all experiments and was used to normalize the output from the firefly luciferase assays. 20 μl/embryo or 2 μl/animal cap of passive lysis buffer was used to homogenize samples on ice. Firefly and Renilla luciferase assays were carried out as per manufacturer's protocol (Promega). Assays were conducted using a Berthold Lumat 9501 luminometer. At least three replicates, consisting of 5–10 embryos or animal caps, were analyzed for each condition. Outputs from the firefly assays were normalized to the outputs from the Renilla assays for each replicate to produce a figure for relative luminescence.The outputs for each replicate assay were

used to calculate the mean and standard deviation for each condition, which were used for the construction of the presented bar charts. Each experiment was repeated at least twice. The data presented are from representative experiments.

RNAase protection

RNA from samples was purified by extracting up to 6 embryos in 0.4 ml of 0.1 M NaCl, 50 mM Tris (pH 8.0), 5 mM EDTA, and 0.5% SDS followed by phenol/chloroform extraction and ethanol precipitation. RNAase protection analysis was carried using the RPA III kit (Ambion). ^{32}P UTP-labeled antisense probes for detection of *Cdx4* and *ODC* were synthesized and hybridized to sample RNA at 45°C overnight and digested with RNAase T1 at 700 u/ml for 1 h before running on 6% acrylamide/urea gels. The *Cdx4* plasmid containing a 300-bp PCR-generated subclone from the reported *Xcad3* sequence (Northrop and Kimelman, 1994) was linearized with *Eco*RI and transcribed with T7 polymerase. The ODC plasmid was linearized with *Bgl*II and transcribed with T7 (Isaacs et al., 1992). The Cdx4 and ODC probes give protected fragments of 300 and 91 bases, respectively.

In situ hybridization

A DIG-labeled in situ probe for the detection of *GFP* transgene expression was produced by linearizing pBS-GFP400 with *Bam*HI and transcribing with T3 RNA polymerase (Pownall et al., 1998). *Cdx4* probe for the detection of endogenous gene expression was made by linearizing with *Eco*RV and transcribing with T3 RNA polymerase. Pigmented wild-type and transgenic embryos were cultured to appropriate stages, and fixed in MEMFA (0.1 M MOPS, 2 mM EDTA, 1 mM $MgSO_4$, 3.7% formaldehyde) for 1 h at room temperature and stored in 100% ethanol at −20°C until further processing. Embryos were rehydrated through a graded series of ethanol and then rinsed in PBS with 0.1% Tween. In situ hybridization was carried out as per the method of Harland (1991). Hybridization was at 60°C overnight and embryos were washed to 0.2 × SSC with the elimination of RNAase A treatment. Expression was visualized using the BM purple alkaline phosphatase substrate (Roche). Following color development, embryos were fixed and bleached in PBS-A + 5% H_2O_2 for several hours.

Photography

Digital photography of specimens was carried out using a Spot Junior CCD camera (Diagnostic Instruments). Image manipulation was carried out using Adobe Photoshop.

Results

Xcad3 is the orthologue of Cdx4

The three *Xenopus* Cdx genes were originally named *Xcad1*, *Xcad2*, and *Xcad3*. Based upon a phylogenetic analysis of the conceptual peptide sequence of the three *Xenopus* proteins, it has been proposed that Xcad1, 2, and 3 represent the orthologues of amniote Cdx2, 1, and 4, respectively (Marom et al., 1997). However, peptide sequence identity between Cdx4 and Xcad3 is only 48%. This is lower than is common between amphibian and amniote orthologues, which typically share between 65% to 85% identity at the protein level. This raises the possibility that *Xcad3* and *Cdx4* might have arisen from independent duplication and divergence events in the amphibian and amniote lineages.

In order to clarify the orthology of *Xcad3* and *Cdx4*, we have used publicly available genome sequence data to compare the genomic loci of *Cdx4*-related genes in zebrafish, *X. tropicalis*, chicken, and human. Our analysis reveals that genomic contexts of the potential *Cdx4* orthologues share common features in all four organisms (Fig. 1A). The Zebrafish *Cdx4*-related gene *Kugelig* and *X. tropicalis cad3* (*Xtcad3*) are flanked by homologues of the *Chic1* and *VEGF-receptor*-related genes. This arrangement is also conserved in the chicken *Cdx-B*/*Cdx4* orthologue found on chromosome 4. Analysis of the human *Cdx4* locus on the X-chromosome shows that the *Chic1* gene is also closely linked. However, close linkage with the *VEGF-receptor*-related gene has been lost during mammalian evolution, with the most closely related sequence to the *VEGF-receptor*-related gene being present on chromosome 3.

These data indicate that the *Xcad3* does represent the orthologue of amniote *Cdx4* and that the duplication event giving rise to *Cdx4* predates the divergence of fish and land tetrapods.

Isolation of the upstream region of the X. laevis Cdx4 gene

A size-directed mini-genomic library approach was used to isolate a 3.4-kbp fragment of the *X. laevis Cdx4* gene. A 400-bp *Eco*RI fragment from the 5′ end of the *Xcad3* cDNA (Northrop and Kimelman, 1994) was used to probe a Southern blot of *X. laevis* genomic DNA digested with a number of restriction enzymes. The DNA digested with *Pvu*II showed strongly hybridizing bands of approximately 3.4 kbp and 2.5 kbp (Fig. 1B). It is likely that the two bands correspond to *Pvu*II fragments from the two pseudoalleles of *Cdx4* which are presumed to be present in the allotetraploid *X. laevis* genome. DNA corresponding to the larger hybridizing band was subcloned and sequenced fully on both strands.

Sequence analysis revealed that the 3.4-kb fragment contained 400 bp of the 5′ end of the *Cdx4* cDNA and 3015 bp upstream of the initiating ATG. We used an RT-PCR-

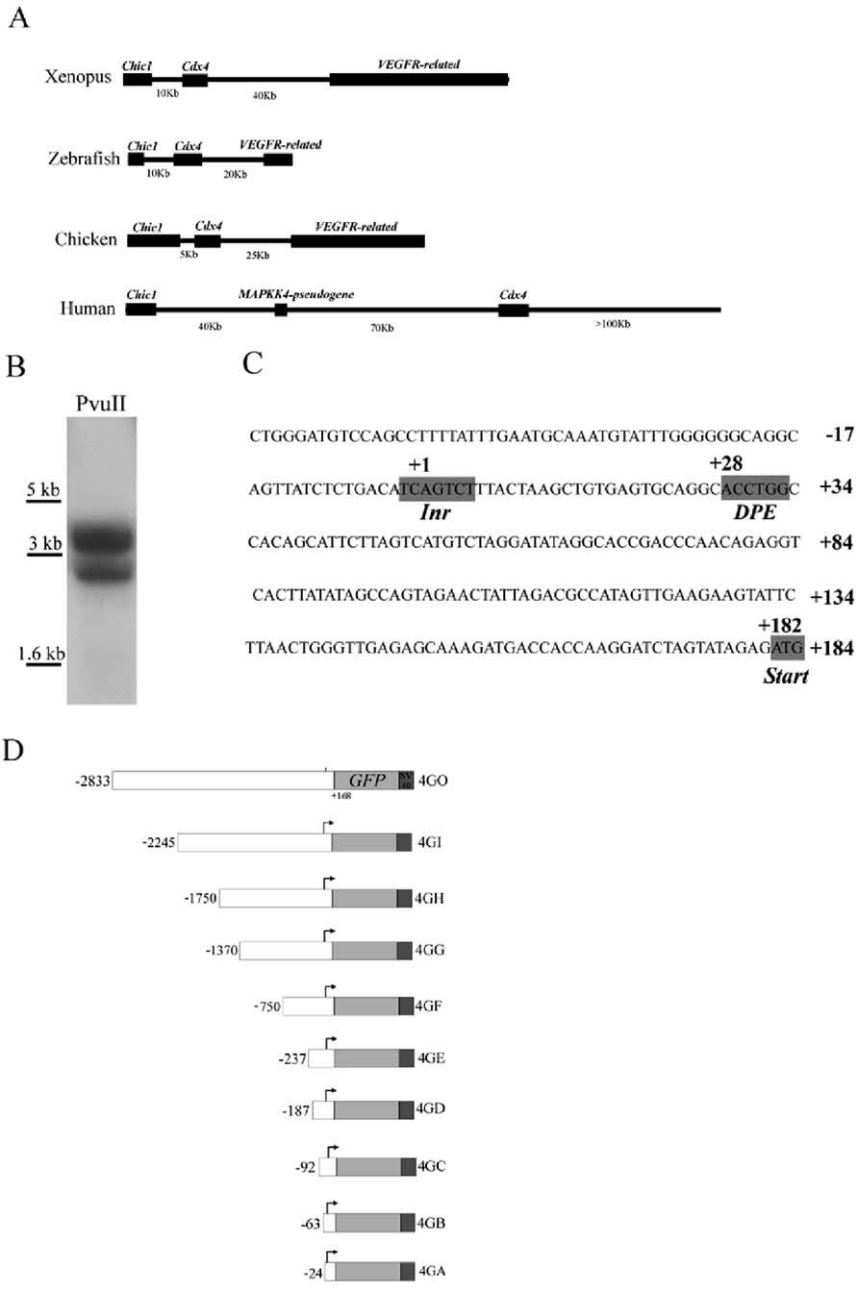

Fig. 1. Identification of the *X. laevis Cdx4* promoter. Panel A is a schematic diagram showing the genomic context of *Cdx4*-related loci in *X. tropicalis*, zebrafish, chicken, and human. Panel B shows an autoradiograph of a Southern blot of *X. laevis* genomic DNA digested with *Pvu*II restriction enzyme and probed with a 400-bp probe from the 5′ end of the *Cdx4* cDNA. Note the two bands arising from sequence polymorphism of the two pseudoallelic *Cdx4* genes. Panel C shows the sequence from the proximal promoter region of *X. laevis Cdx4*. Transcriptional initiator sequence (***Inr***), downstream promoter element (***DPE***), and translational start (***start***) are boxed in gray. Sequence is numbered relative to the transcription start site +1. Panel D shows diagrams of the GFP reporter plasmids made from the isolated 3.4-kb *Cdx4* promoter fragment. The base construct *4GO* contains 2833 bp upstream of the transcription start site and 168 bp of 5′ untranslated region (UTR) cloned upstream of a nuclear GFP reporter. The construct also contains a 3′ SV40 adenylation sequence. Constructs *4GI* to *4GA* represent a deletion series (2245 bp to 24 bp) of the sequence upstream of the transcriptional start site.

based approach to map the transcription start of the *Cdx4* gene to a position between 185 and 177 upstream of the initiating ATG codon (data not shown). This is in close agreement with the position of 175 upstream of the ATG previously reported for the *X. laevis Cdx4* transcription initiation site (Haremaki et al., 2003). Our analysis of the sequence surrounding the transcription start showed that the consensus sequence for a TATA box was absent but that the sequence from position −183 to −177, relative to the ATG, adheres closely to the (C/T)(C/T)A(A/C/G/T)(T/A)(C/T)(C/T) consensus for a transcription initiator (Inr) sequence (Fig. 1C). This would predict that the A at position −181, relative to the ATG, represents the transcription start site. Further support for this being the legitimate transcription start site

comes from the identification of a sequence matching 5 out 6 bases for the downstream promoter element (DPE) consensus (A/G/T) (C/G) (A/T) (C/T) (A/C/G) (C/T) at positions +28 to +33 relative to the proposed transcriptional start site (Fig. 1C). The DPE is a common feature of other TATA-less RNA polymerase II promoters and is involved in the binding of TFIID (Burke et al., 1998; Butler and Kadonaga, 2001).

Expression from the 2.8-kb upstream fragment in transgenic embryo

A GFP reporter construct was produced which consists of the whole of the sequence upstream of transcription start site (−2834 to −1) together with most of the 5′ UTR (+1 to +168). This clone designated *4GO* was then used to generate a series of 5′ deletions designated *4GF* to *4GA* (Fig. 1D).

In order to aid comparison between endogenous *Cdx4* and transgenic reporter expression, we present the normal expression pattern of the *Xenopus Cdx4* gene (Fig. 2). Expression is initiated during gastrula stages in a circumblastoporal ring within the nascent mesoderm (Fig. 2A). At the end of gastrulation, expression is seen in a posterior domain around the closed blastopore (Fig. 2B). A sagittal section through an embryo at the end of gastrulation shows that *Cdx4* expression is present in all three germ layers in the posterior (Fig. 2C). During neurula stages, expression persists in the posterior mesoderm and becomes prominent in the posterior neural tube (Fig. 2D). During tailbud stages, expression becomes restricted to the posterior neural tube and mesoderm of the tailbud (Fig. 2E). A transverse section through the posterior neural tube region of a tailbud stage embryo shows that *Cdx4* expression is restricted to the neural tube situated above the notochord (Fig. 2F).

Fig. 3 shows the expression of the *4GO* reporter construct in embryos at different stages of development. In this study, we have visualized reporter gene expression using in situ hybridization to detect expressed GFP mRNA, rather than using direct GFP fluorescence because this provides a permanent record of expression and in our hands is more sensitive at detecting reporter expression, particularly in pre-tailbud stage embryos. In gastrula stage embryos (Figs. 3A to F), the highest level of expression is detected in the mesoderm in a similar pattern to the endogenous gene. However, there is typically more reporter expression in the animal hemisphere early gastrula stage than is seen with endogenous *Cdx4* (Fig. 3B). A section through an early gastrula stage embryo along the dorsoventral axis reveals that, in the dorsal blastopore region, reporter expression is restricted to superficial cells, whereas in ventral regions, both superficial and deep cells show expression. A similar difference between dorsal and ventral blastopore expression of *Cdx4* has previously been reported (Isaacs et al., 1999) (Fig. 3C). Diffuse expression is also noted in the anterior regions of mid-gastrula stage transgenics (Fig. 3E). This suggests that some negative regulatory elements necessary for refining *Cdx4* expression during these stages are absent from the sequences present in *4GO*. However, it is clear that *4GO* is able to drive robust reporter expression in the posterior mesoderm of gastrula stage embryos (Figs. 3G, H, and I). Posterior mesodermal expression persists through neurula stages; however, the expression in the posterior neural tube is very reduced in *4GO* transgenic embryos as compared to endogenous *Cdx4* (Figs. 3J and K).

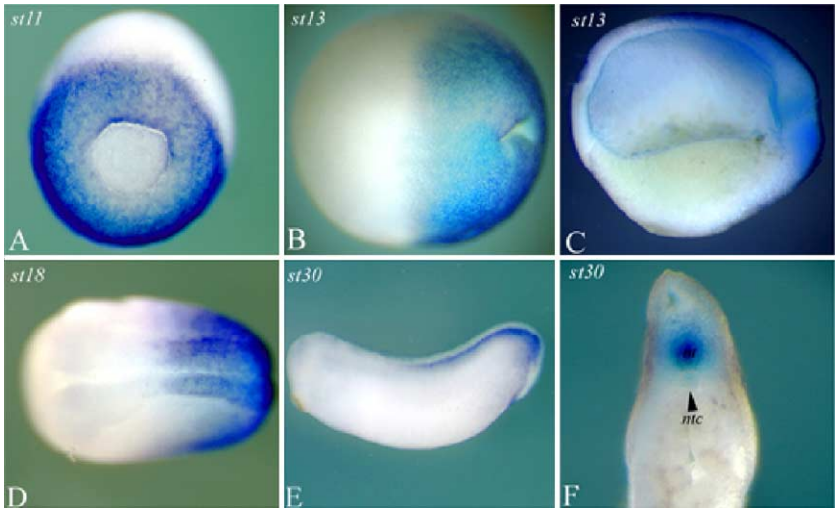

Fig. 2. Normal expression of *X. laevis Cdx4*. This figure shows a whole-mount in situ hybridization analysis of the expression of *Cdx4* at various stages of development. Panel A is a vegetal view of a mid-gastrula stage 11 embryo. Dorsal is to the top. Panel B is a dorsal view of a early neurula stage 13 embryo. Anterior is to the left. Panel C is a sagittal section through a neurula stage 13 embryo. Anterior is to the left. Panel D is a dorsal view of a late neurula stage 18 embryo. Anterior is to the left. Panel E is a lateral view of an early tailbud stage 30 embryo. Anterior is to the left. Panel F is a transverse section through the posterior trunk of an early tailbud stage 30 embryo. Dorsal is to the top and the black arrow indicates the notochord. ntc = notochord, nt = neural tube.

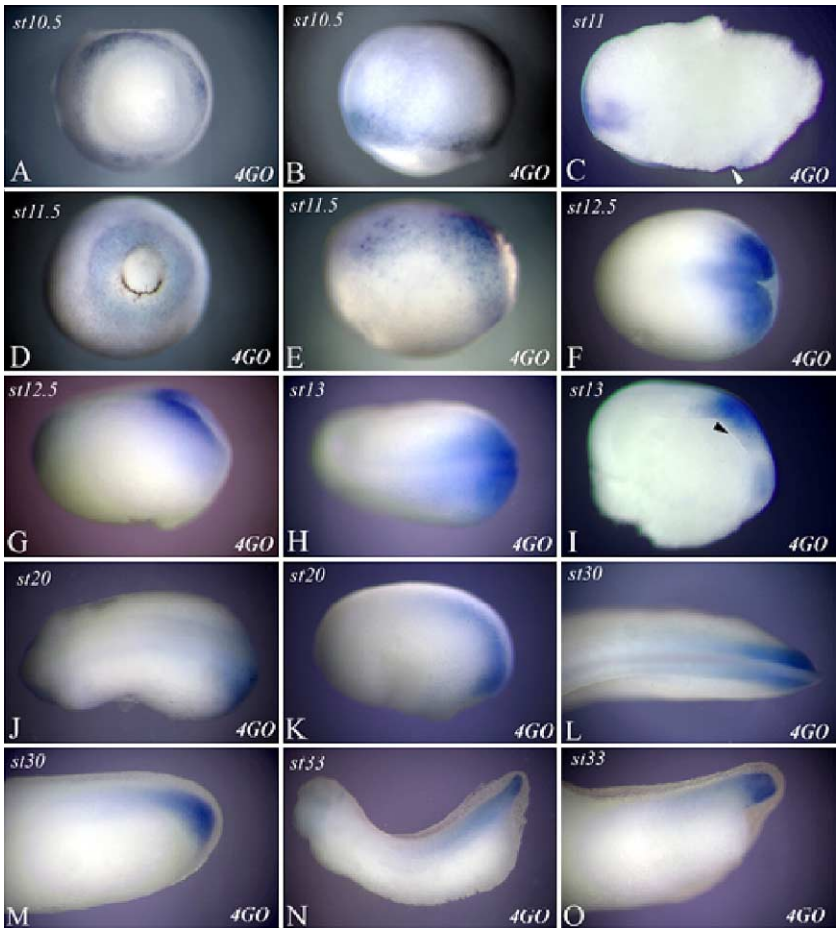

Fig. 3. Expression of the *4GO* reporter in transgenic embryos during development. This figure shows the expression of the *4GO* reporter construct in transgenic embryos at a number of developmental stages as visualized by in situ hybridization against GFP RNA. The 4GO construct contains 2.8 kb upstream of the transcription start site. In Panels E to O, anterior is to the left. Panel A is a vegetal view of a gastrula stage 10.5 embryo. Dorsal is to the top. Panel B is a lateral view of a stage 10.5 embryo. The animal hemisphere is to the top. Panel C is a section of a gastrula stage 11 embryo bisected along the dorsoventral axis. Animal is to the top dorsal to the right. White arrow indicates the blastpore lip. Panel D is a vegetal view of a stage 11.5 embryo. Dorsal is to the top. Panel E is a lateral view of a stage 11.5 embryo. Dorsal is to the top. Panel F is a dorsal view of a gastrula stage 12.5 embryo. Panel G is a lateral view of a stage 12.5 embryo. Panel H is a dorsal view of an early neurula stage 13 embryo. Panel I is a sagittal section of a stage 13 embryo. Back arrow indicates forming archenteron cavity. Panel J is a dorsal view of a late neurula stage 20 embryo. Panel K is a lateral view of a stage 20 embryo. Panel L is a dorsal view of the posterior region of tailbud stage 30 embryo. Panel M is a lateral view of the posterior region of a stage 30 embryo. Panels N and O are lateral views of tailbud stage 33 embryos.

In tailbud stage embryos, expression from the *4GO* is found in the tail-forming region (Figs. 3L–O). There are some differences between the expression of the *4GO* reporter and the endogenous *Cdx4* gene, indicating that some positive and negative regulatory elements are absent from *4GO*. For example, reporter expression is more widespread in somatic tissue and less prominent in the posterior neural tube relative to *Cdx4*. However, we conclude that the 2.8-kb upstream fragment is able to recapitulate much of the expression of the endogenous *Cdx4* gene.

Tailbud expression from a deletion series of the 2.8-kb upstream fragment in transgenic embryos

In order to begin to identify elements necessary for expression from the 2.8-kb promoter fragment, we generated a series of 5′ deletions, designated *4GI* to *4GE*, containing between 2245 and 237 bp upstream of the proposed transcription start site (Fig. 1D). Fig. 4 shows the expression in tailbud stage embryos from these constructs. Robust tailbud expression is seen with constructs *4GI* to *4GE*.

A more detailed analysis of expression at different stages from the *4GE* reporter which contains 237 bp of the proximal promoter sequence was undertaken (Fig. 5). Unlike the *4GO* construct, *4GE* does not drive mesoderm-specific expression in gastrula stages. Fig. 5A shows general animal hemisphere expression in an early gastrula stage 4GE transgenic embryo. However, *4GE* is capable of driving posterior expression in neurula and tailbud stage embryos. Fig. 5B is a sagittal section through an early neurula stage embryo showing strong posterior expression all three germ layers. Diffuse anterior

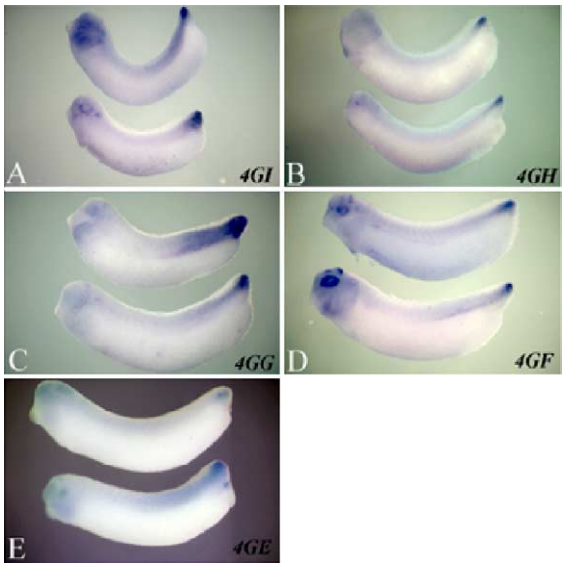

Fig. 4. Expression from a 5′ deletions series in transgenic embryos. This figure shows the expression in transgenic tailbud stage embryos from a series of 5′ deletion constructs visualized by hybridization against GFP RNA. The *4GI*, *4GH*, *4GG*, *4GF*, and *4GE* constructs contain, respectively, 2245, 1750, 1370, 750, and 237 bp of sequence upstream of the transcriptional start site. In all panels, anterior is to the left. Panels A, B, C, D, and E show the expression of constructs *4GI*, *4GH*, *4GG*, *4GF*, and *4GE*, respectively.

ectodermal expression is also frequently detected during neurula stages (Figs. 5C and D).

During tailbud stages, *4GE* drives robust tailbud expression. As has been previously noted for *4GO*, *4GE* expression is detected in the posterior neural tube, although this is accompanied by more general expression in the posterior mesoderm (Figs. 5E–G).

We note with reporter *4GE* that there are additional anterior regions of expression in the eye, the hindbrain, and at the midbrain/hindbrain junction during tailbud stages (Fig. 5H). Our data indicate that, although the elements required to drive posterior expression in neurula and tailbud stages are present in the 237-bp proximal promoter fragment, other important negative regulatory elements necessary for refining normal *Cdx4* expression are missing.

Quantitative analysis of expression from the upstream fragments

Having shown that *4GO*, containing 2.8 kb of upstream sequence, and *4GE*, containing 237 bp of upstream sequence, were able to drive reporter expression in transgenic embryos, we decided to use a quantitative method to analyze the expression from these constructs. In order to do this, the promoter fragments from *4GO* and *4GE* were cloned upstream of a firefly luciferase reporter to produce the *4LO* and *4LE* constructs. We used these constructs to show that the profiles of expression from a luciferase reporter driven by these fragments are similar to that of the endogenous gene.

Fig. 6A is an RNAse protection analysis showing the expression of *Cdx4* at several stages from blastula stage 8 to tailbud stage 35. Fig. 6B is a bar chart showing the relative expression of *Cdx4* at these stages, as determined by densitometry of the autoradiograph in Fig. 6A. *Cdx4* expression was normalized to the *ornithine decarboxylase* (*ODC*) loading control. It can be seen that the highest level of expression from *Cdx4* is at late gastrula stage 13 and that overall expression levels diminish through to tailbud stages.

Fig. 6C is a bar chart showing the relative luminescence produced by expression from the *4LO* and *4LE* reporter constructs at a number of stages from blastula stage 7 until tailbud stage 30. The plasmids were injected into 2- or 4-cell stage embryos and were cultured until the relevant stage. Standard luciferase assays were undertaken and the luminescence produced by the firefly luciferase was normalized to the expression produced from a coinjected Renilla luciferase plasmid. As is seen with the endogenous gene, highest levels of expression from the *4LO* and *4LE*

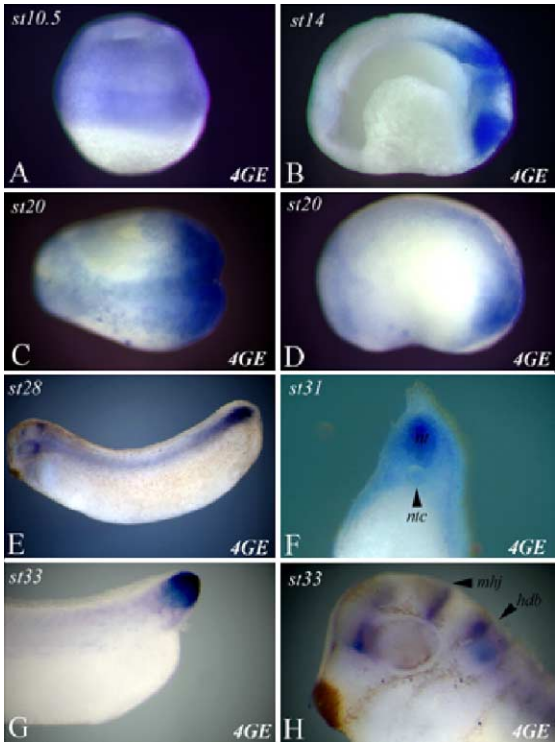

Fig. 5. Expression from the *4GE* 5′ deletion construct in transgenic embryos. This figure shows the expression in transgenic embryos of the *4GE* 5′ deletion construct at various stages visualized by hybridization against GFP RNA. *4GE* contains 237 bp of sequence upstream of the transcriptional start site. In panels B, C, D, E, G, and H, anterior is to the left. Panel A is a lateral view of a gastrula stage 10.5 embryo. Animal is to the top. Panel B is a sagittal section through an early neurula stage 14 embryos. Panel C is a dorsal view of a late neurula stage 20 embryo. Panel D is a lateral view of stage 20 embryo. Dorsal is to the top. Panel E is a lateral view of a tailbud stage 28 embryo. Panel F is a transverse section through the posterior trunk of a tailbud stage 31 embryo. Panel G is a lateral view of the tail-forming region from a stage 33 embryo. Panel H is a lateral view of the head of a tailbud stage 33 embryo. mhj = midbrain/hindbrain junction, hdb = hindbrain, ntc = notochord.

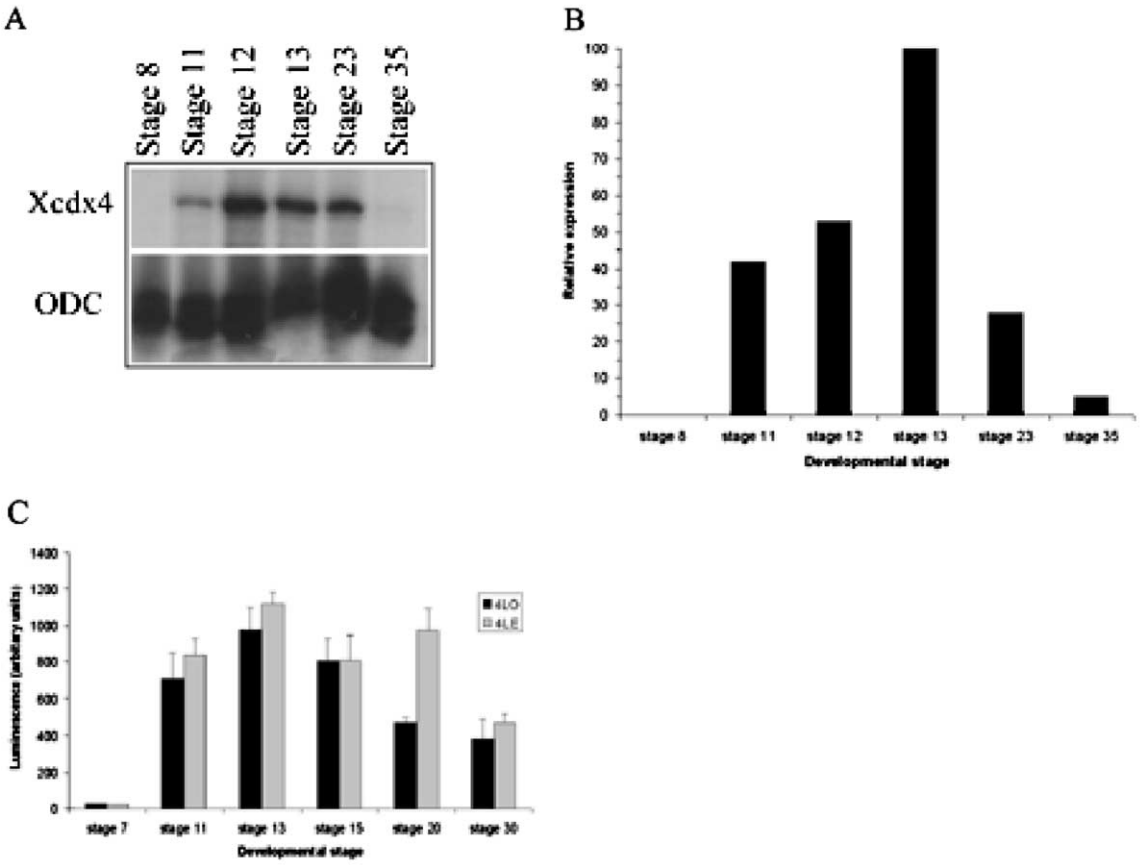

Fig. 6. Quantitative analysis of *Cdx4* and reporter construct expression. Panel A is an autoradiograph of an RNAse protection analysis of *Cdx4* gene expression during development. 10 μg total RNA from each stage was hybridized with probes for *Cdx4* and *ODC*. Panel B is a chart showing the relative expression of Cdx4 at the same stages as determined by densitometry. *Cdx4* expression was normalized to *ODC* expression and represented as a percentage of the maximum expression at stage 13. Panel C shows the relative luminescence detected from embryos injected with the *4LO* and *4LE* firefly luciferase reporter plasmids during development. The *4LO* and *4LE* constructs, respectively, contain 2833 bp and 237 bp upstream of the transcriptional start site. The firefly luminescence was normalized to the luminescence produced by the coinjected CMV-Renilla luciferase control plasmid.

constructs are detected in gastrula stage 13 embryos. Expression from both *4LO* and *4LE* falls to similar levels during tailbud stages. However, it is clear that the levels of expression from both reporters in tailbud embryos are much higher relative to gastrula stages than is seen with the endogenous genes, suggesting some elements necessary for normal expression are missing.

Identification of regulatory elements with the proximal promoter region

Given that a 237-bp fragment of the proximal promoter region was capable of driving reporter expression, we were interested to identify elements necessary for the activity of the *Cdx4* promoter. A further set of GFP reporter deletions, designated *4GD* to *4GA*, was produced (Fig. 1D). Constructs *4GD* and *4GC*, containing 187 and 92 bases upstream of the transcription start site, respectively, both gave tailbud expression (Figs. 7A and B). In addition to posterior expression in the tail-forming region, ectopic domains of expression were observed in the anterior nervous system.

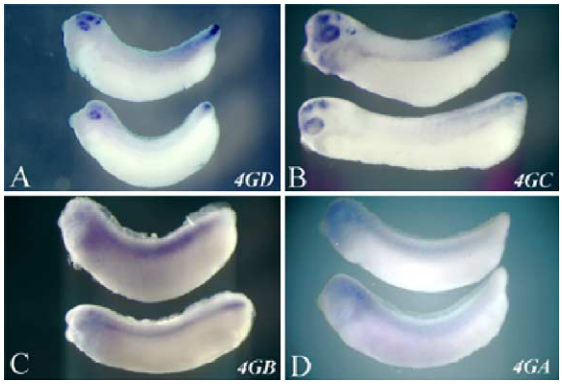

Fig. 7. Expression from a 5′ deletion series of the proximal promoter region in transgenic embryos. This figure shows the expression in transgenic tailbud stage embryos from a series of reporter constructs containing 5′ deletions of the proximal promoter region of *Cdx4* visualized by hybridization against GFP RNA. The *4GD*, *4GC*, *4GB*, and *4GA* constructs contain, respectively, 187, 92, 63, and 24 bp of sequence upstream of the transcriptional start site. In all panels, anterior is to the left. Panels A, B, C, and D show the expression of constructs *4GD*, *4GC*, *4GB*, and *4GA*, respectively.

These domains were similar to those previously noted for constructs *4GF* and *4GE* (Figs. 4C and D).

The next deletion *4GB*, containing just 63 bases of upstream sequence, gives diffuse axial expression but no specific tailbud expression is detected (Fig. 7C). The smallest deletion construct *4GA* containing 24 bases of upstream sequence drives no detectable transgenic expression.

Quantitative analysis of deletions from the proximal promoter region

From our transgenic 5′ deletion study, we hoped to gain insight as to the sequences required for tailbud expression and promoter activity. A clear conclusion from the data showing non-localized reporter expression from *4GB* and tailbud expression from *4GC* is that at least some of the elements necessary for tailbud expression are present between bases −64 and −92. Furthermore, lack of expression from *4GA* and generalized expression from *4GB* suggests that elements between −25 and −63 are involved in general promoter activity rather than regulating regional expression. Here, we provide further evidence for this latter conclusion using a quantitative analysis of expression from a set proximal promoter deletions driving luciferase expression in neurula stage embryos. Neurula stages were chosen for these and subsequent quantitative analysis of promoter deletions and mutations because this stage is the earliest that posterior-specific expression is detected from the *4GE* proximal promoter fragment (Figs. 5B–D).

Fig. 8A shows the 237-bp sequence of the *Cdx4* proximal promoter region contained within the *4LE* reporter. Also shown are the positions of deletions used to generate the luciferase constructs *4LD* to *4LA*. The sequence of these deletions corresponds with those contained in the transgenic GFP reporters *4GE* to *4GA*.

The activity of the firefly luciferase reporters was assayed in early neurula stage 13 embryos. Fig. 8B shows that fragments *4LD* to *4LB* have enhanced promoter activity relative to *4LE*. However, the activity of construct *4LA*, which is deleted at position −24, dramatically reduces activity to 7% of the activity of *4LE* and to 5% of the activity of the next smallest promoter fragment (*4LB*), which is deleted at position −63. This confirms our previous conclusion that an important positive regulatory element necessary for *Cdx4* promoter activity is found between positions −63 to −24. A search for potential transcription factor binding sites revealed that a consensus binding site for the POU-domain transcription factor Oct1 is present at positions −41 to −34. Interestingly, Oct1 has previously been implicated in the regulation of the murine *Cdx2* gene (Jin and Li, 2001). We note that the potential Oct1 binding site of ATGCAAAT found in the *X. laevis Cdx4* promoter exactly matches the sequence in murine *Cdx2* proximal promoter region that has been shown to physically interact with Oct1 protein.

The Oct1 binding motif and promoter activity

Having identified an Oct1 binding site in the proximal promoter region of *Cdx4*, we determined if mutating this site interferes with the activity of the *Cdx4* luciferase reporter constructs. In order to do this, mutant versions of the *4LE* and *4LB* (designated *4LE-ΔPOU* and *4LB-ΔPOU*) were constructed in which the Oct1 binding site is mutated from ATGCAAAT to **C**TGCA**G**AT. This mutation has previously been shown to abolish Oct1-dependent reporter activation (Jin and Li, 2001). Fig. 9A shows that mutation of the Oct1 binding site in *4LE* and *4LB* greatly reduces promoter activity in embryos. The activity of the *4LE-ΔPOU* reporter is 10% that of *4LE*, and in the case of *4LB-ΔPOU*, the activity of the reporter is reduced to 8% of *4LB*.

We provide additional support for the role of Oct1 in regulating the *Cdx4* promoter by showing that ectopic expression of Oct1 in tissue explants from embryos injected with the *4LE* plasmid greatly increases reporter activity and that the *4LE-ΔPOU* plasmid does not respond in a similar

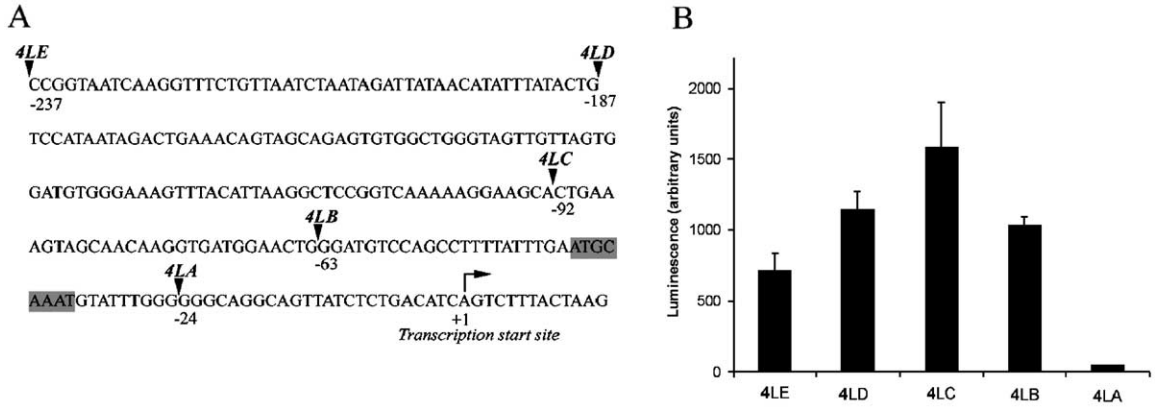

Fig. 8. Quantitative analysis of expression from *Cdx4* proximal promoter deletion series. Panel A shows the sequence of the proximal promoter region of *X. laevis Cdx4*. The positions, relative to the transcriptional start site, of the deletions used to make luciferase reporter constructs *4LE* to *4LA* are indicated. The consensus Oct binding site is boxed in gray. Panel B compares the relative luminescence detected in neurula stage 13 embryos injected with the *4LE* to *4LA* deletion series. The firefly luminescence was normalized to the luminescence produced by the coinjected CMV-Renilla luciferase control plasmid.

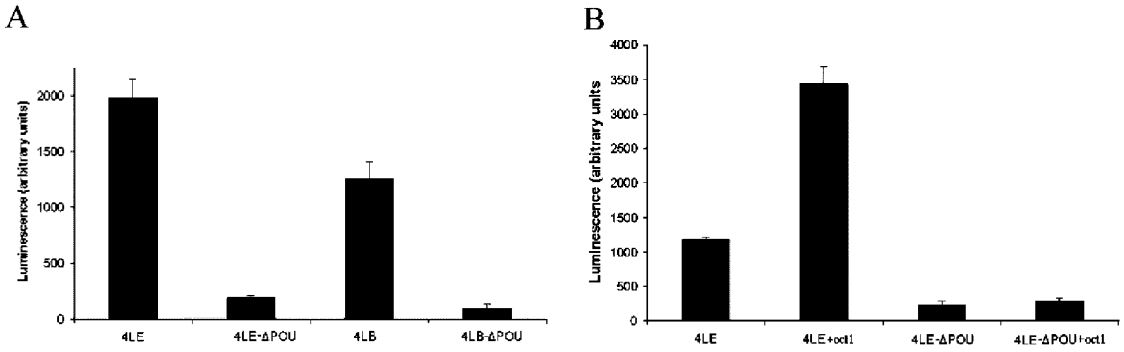

Fig. 9. Quantitative analysis of effects on expression resulting from mutating a consensus Oct binding motif in the proximal promoter region. Panel A compares the relative levels of luminescence detected in neurula stage 15 embryos injected with reporters *4LE* and *4LB* with that detected in embryos injected with reporters 4LE-ΔPOU and 4LB-ΔPOU, in which the consensus Oct binding site ATGCAAAT has been mutated to **C**TGCA**G**AT. Panel B shows the effects coinjection of *Oct1* mRNA has on expression from the *4LE* and 4LE-ΔPOU plasmids in animal cap explants at neurula stage 18. The firefly luminescence was normalized to the luminescence produced by the coinjected CMV-Renilla luciferase control plasmid.

fashion to ectopic *Oct1* expression. Fig. 9B shows that the coinjection of 1 ng *Oct1* mRNA with the *4LE* plasmid elevates reporter activity by nearly 3-fold. Furthermore, we demonstrate the importance of the Oct1 site in mediating this effect by showing that coinjection of *Oct1* mRNA with *4LE-ΔPOU* does not result in any increase in reporter activity. We conclude that the Oct1 site is critical for *Cdx4* promoter activity.

Conservation of Cdx4 promoter elements in other species

The recent availability of genome sequence from the closely related amphibian species *X. tropicalis* provides an opportunity to determine if the proximal promoter elements of *Cdx4* have been conserved during evolution. An alignment of the *X. tropicalis* region corresponding to −237 and +184 of the *X. laevis* proximal promoter region shows a high degree of sequence identity (Fig. 10A). We note that the Oct1 site is perfectly conserved, as is the DPE, and the Inr differs by only one base, but still conforms to the Inr consensus. The overall sequence identity between *X. laevis* and *X. tropicalis* within this region, which corresponds to the regions included in the *4LE* and *4GE* constructs, is 85%. However, in a region corresponding to −238 to −1180 of the *X. laevis* upstream sequence, this sequence identity drops to 49%.

We also show that the Oct1 binding motif and surrounding sequence has been highly conserved during vertebrate evolution. Fig. 10B shows an alignment of the proximal promoter region of *Cdx4* from the *Xenopus* species, mouse, human, chicken, and zebrafish. The Oct1 motif is contained within a highly conserved 22-bp block. The core Oct1 binding motif is identical in all species except chicken where a C to T substitution is present at position 5. In this highly conserved region, 16 out of 22 bases are invariant between all species analyzed and only 2 differences are present between *X. laevis* and humans. The spacing of the Oct1 motif relative to the conserved Inr has also been maintained during evolution.

As discussed above, an Oct1 motif has been previously identified in the murine *Cdx2* promoter. We were therefore interested to see if this motif is present in the promoters of the other *Xenopus* Cdx genes. The sequence of the *Cdx1* and *Cdx2* promoters from *X. laevis* has not been reported. However, using sequences derived from the second draft of the *X. tropicalis* genome, we show that an identical Oct1 binding motif is present in the proximal promoter region of *X. tropicalis Cdx1* and that a sequence differing by only one base is present in *Cdx2* (Fig. 10C). The context of the Oct1 motif is also similar in all three genes, with upstream pyrimidine-rich and downstream GC-rich regions. We conclude that octamer-binding proteins are likely to play a critical role in the regulation of the Cdx gene family.

Discussion

Elements required for posterior expression are present in sequences upstream of the Cdx4 promoter

The Cdx caudal-related gene family have roles in the development of posterior structures in many animal groups and characteristically have dynamic patterns of expression within the developing posterior axis. Until recently, very little was known about the gene regulatory elements involved in regulating expression of vertebrate *Cdx4*. However, a recent study undertook a detailed analysis of regulatory elements within intron 1 of *Cdx4* that are required for expression within posterior neural tissue (Haremaki et al., 2003). In contrast, the present study addresses the function of sequences upstream of the *Cdx4* transcription start site in regulating promoter activity and spatially restricted expression.

Our analysis of the proximal promoter region of *Cdx4* shows that it lacks a recognized TATA box but does contain sequences which conform to the initiator consensus near to the mapped transcription start site. A transgenic analysis of expression from a 5′ deletion series indicates that elements

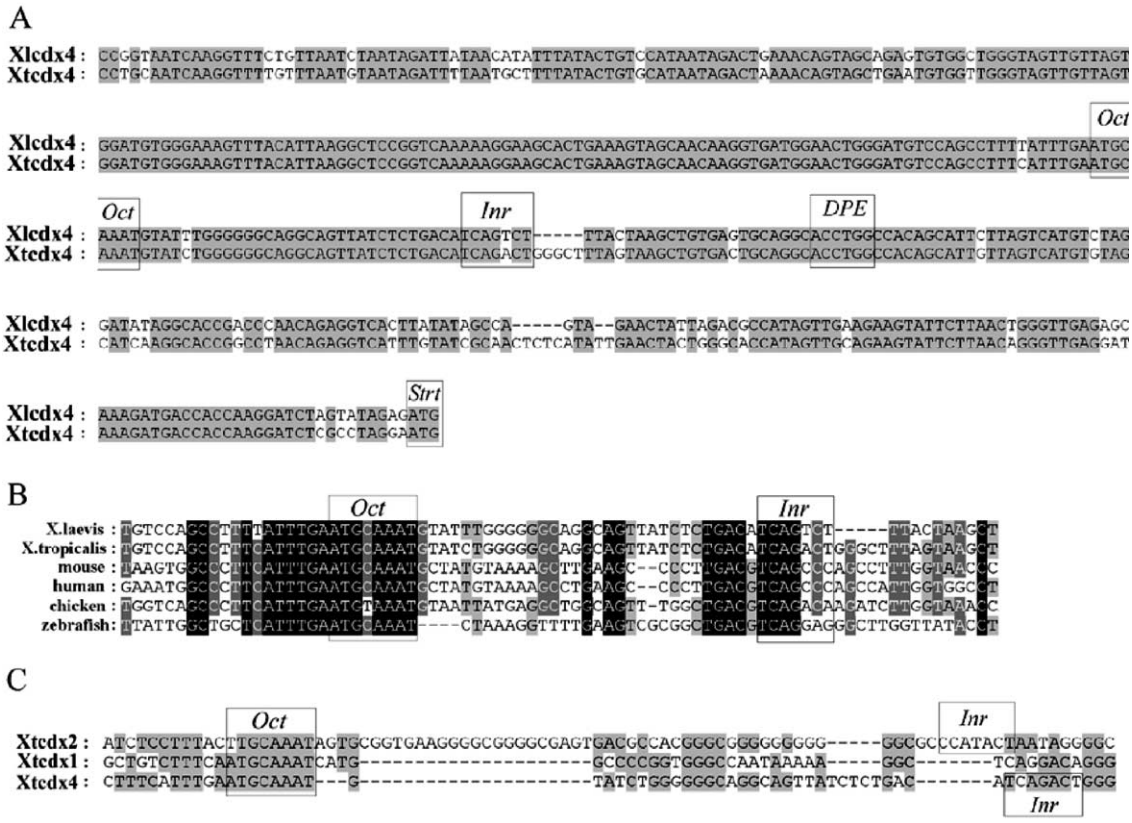

Fig. 10. Conservation of an Oct binding site during evolution of the Cdx gene family. Panel A shows an alignment of the *Cdx4* proximal promoter regions from *X. laevis* and *X. tropicalis*. Conserved sequences are boxed in gray. Panel B is an alignment of the *Cdx4* proximal promoter from *X. laevis*, *X. tropicalis*, mouse, human, chicken, and zebrafish. Sequence identical in all six species is boxed in black. Sequence conserved in three species is boxed in gray. Panel C shows an alignment of the proximal promoter regions from *X. tropicalis Cdx1*, *Cdx2*, and *Cdx4* genes. Sequence identical in all three genes is boxed in gray. The Oct binding motif (**Oct**), initiator sequence (**Inr**), and downstream promoter element (**DPE**) are indicated where appropriate. Alignments were generated with the Megalign program using the Clustal method.

capable of driving posterior expression are present within the upstream region of *Cdx4*. We show that a 2.8-kb upstream fragment will drive reporter expression within the early mesoderm and subsequently in the posterior axis until at least tailbud stages. A smaller 237-bp fragment will also drive posterior expression in the mesoderm and neuroectoderm from neurula stages but has a reduced ability to drive mesoderm-specific expression during gastrula stages. Elements required for the early mesoderm expression are therefore likely to be present in the distal 0.24-kb to 2.8-kb region. The presence of ectopic anterior domains of expression from the smaller deletion constructs, including the 0.24-kb fragments, indicates that repressive elements present in the extended upstream region are required for the posterior restricted expression of *Cdx4*.

A more detailed deletion analysis of the 237-bp region has revealed some interesting features regarding *Cdx4* regulation. We show that a reporter containing 93 bp of upstream sequence will direct posterior expression, while deletion at position −63 eliminates posterior-specific expression but retains significant promoter activity. Thus, we have been able to identify two regions of the proximal promoter, one of which is required for posterior specific expression (bases −64 to −92) and one which is required for general promoter activity (−25 to −63). In the present study, we have not identified any specific elements within the region −64 to −93 involved in driving posterior expression but we have been able to identify an element critical for promoter activity within the −25 to −63 region of the proximal promoter.

A conserved site for octamer-binding proteins in the proximal promoter region of Cdx genes

Deletion analysis indicates that 63 bp upstream of the *Cdx4* transcription start site is sufficient for promoter activity in *Xenopus* embryos. Within this 63-bp region, we have identified a consensus site for octamer-binding proteins and have shown that, when this site is mutated, promoter activity and responsiveness to the Oct1 protein are lost. This same core 8-bp sequence has previously been shown to be present at position −125 to −118 of the murine *Cdx2* promoter and DNA oligonucleotides containing this sequence bind Oct1 protein in gel shift assays (Jin and Li, 2001). The *Cdx2* Oct1 site was also shown to be necessary for *Cdx2* reporter activity in the Caco-2 intestinal cell line.

We present evidence that an Oct binding site has been highly conserved during the evolution of the Cdx gene family and is present in the proximal promoter region of *Cdx4* genes of vertebrate groups from fish to humans. This indicates an ancient origin for a regulatory pathway involving the Cdx and POU-domain families of transcription factors. Furthermore, the presence of conserved sites for octamer-binding proteins in amphibian *Cdx1*, and both amphibian and murine *Cdx2*, would indicate that the origin of this network predates the expansion of the Cdx family that occurred between primitive chordates, which have one Cdx gene, and vertebrates which have three Cdx genes (Brooke et al., 1998). Analysis of Cdx genomic loci from basal chordates will be illuminating in this regard.

POU-domain transcription factors and Cdx gene expression

Oct1 belongs to the POU family of transcription factors which are characterized by the presence a bi-partite DNA binding domain, consisting of a POU-specific (POU-S) subdomain of 75 amino acids, present at the N-terminus of the domain, separated by a variable length linker region from a 60 amino acid variant homeodomain (POU-H) (Andersen and Rosenfeld, 2001). 15 members of the POU family have been identified in mammals and they have been classified into classes I to VI based upon the sequence homology within their POU domains. The POU proteins bind with high affinity to an octamer element of sequence ATGCAAAT or variations on this core sequence. Functional octamer elements have been shown to be present in the regulatory regions of a wide range of genes required for development, differentiation, and homeostasis of the adult organism (Andersen and Rosenfeld, 2001).

The mechanisms by which POU domain transcription factors such as Oct1 are able to activate transcription of target genes is of some interest. Oct1 functions as an activator when bound close to a promoter and POU domain factors contain both an N- and a C-terminal activation domain (Andersen and Rosenfeld, 2001). It has also been proposed that Oct1 can recruit TATA binding protein to TATA-less promoters (Zwilling et al., 1994) and is able to stimulate the formation of the preinitiation complex through interactions with TFIID (Arnosti et al., 1993).

An important question with regard to POU domain regulation of Cdx activity is: Does the octamer-binding site act as a spatial enhancer of Cdx expression? Our data indicate that, in the context of the *Cdx4* promoter, the answer is no; rather, the octamer-binding site is required for the general activity of the *Cdx4* promoter. In this regard, it is worth noting that the octamer element in the murine *Cdx2* proximal promoter region does not have enhancer-like activity and does not stimulate transcription when fused to a basal promoter (Jin and Li, 2001).

Amphibian *Cdx4* is predominantly expressed in ectodermal and mesodormal lineages. It is interesting to note that the murine *Cdx2* study examined the role of the Oct site in regulating expression in endodermal cell types (Jin and Li, 2001). These data indicate that octamer-binding proteins are likely to be part of a general mechanism for Cdx gene expression that is not restricted to specific germ layers or cell types. The elucidation of the exact mechanisms by which octamer-binding proteins regulate expression from Cdx promoters will require further studies.

A model for Cdx4 regulation

It has been known for sometime that the normal expression of *Xenopus Cdx4* requires the activity of the FGF signaling pathway (Northrop and Kimelman, 1994; Pownall et al., 1996) but little was known regarding the gene regulatory elements involved in mediating the ability of FGFs to activate *Cdx4* expression. However, a recent study clearly demonstrated that binding sites for the Ets family of transcription factors present within intron 1 play a crucial role in mediating the FGF responsiveness of *Cdx4* (Haremaki et al., 2003). Interestingly, the same study showed that many of consensus Ets binding sites were in close proximity to functional binding sites for the HMG transcription factors Sox2 and Tcf/Lef. It is suggested that Ets proteins and HMG factors might physically interact and bind as a complex to the composite Ets/HMG binding sites. Tcf/Lef is a nuclear effector of the canonical Wnt signaling pathway and the presence of the Ets/HMG binding sites indicates that interactions between the Wnt and FGF signaling pathways are likely to be crucial for normal Cdx regulation. The early mesoderm and the later posterior axis, including the ectodermal and mesodermal derivatives, represent regions where FGF and Wnt signaling pathways are active. It therefore seems likely that the FGF and Wnt responsive intron enhancers play a role in restricting the spatial expression of *Cdx4* during early development, although the study of Haremaki and Okamoto (2003) does not address this issue directly.

The composite Ets/HMG enhancer sequences present in intron 1 are likely to play a part in regulating *Cdx4* expression in the posterior axis. However, the present study indicates that other elements present in the upstream region of the gene are also likely to be involved in regulating the spatial extent of *Cdx4* expression. In the present study, we show that as little as 93 bp of the *Cdx4* upstream region is capable of driving posterior *Cdx4* expression in transgenic embryos.

Our transgenic analysis of a 2.8-kb fragment also indicates that repressor elements are required to refine *Cdx4* gene expression and inhibit its expression in anterior regions of the embryo. A common site of ectopic expression from our reporter constructs is at the midbrain/hindbrain junction, which is an active site of Wnt and FGF signaling. This might indicate that FGF/Wnt responsive elements are found in this upstream sequence. Indeed, our unpublished observations indicate that both the 2.8-kb and 0.24-kb upstream fragments are responsive to FGF signaling,

although the present study gives no indication of what these elements might be.

Knowledge of the various signaling pathways involved in regulating *Cdx4* gene expression allows us to propose a model for the regulation from the amphibian *Cdx4* locus. Our data indicate that POU proteins are unlikely to be involved in the tight spatial regulation of *Cdx4* expression in the embryo but are factors intimately involved in the activity of the basal promoter. This is further supported by data showing that POU domain transcription factors, such as *Oct1* and *XlPOU-60*, have rather widespread expression in amphibian embryos (Veenstra et al., 1995; Whitfield et al., 1993). Although it has been recently reported that *Xenopus Oct-25* gene has a prominent domain of expression in the tail-forming region (Cao et al., 2004).

Data presented in this paper, together with the study of Haremaki et al. (2003), indicate that the restricted spatial expresson of *Cdx4* in normal development requires *cis*-regulatory elements present in both the upstream and intron 1 regions of the *Cdx4* gene. These *cis*-acting elements serve to integrate regulatory information from multiple signaling pathways, including the FGF and Wnt pathways. Elements within intron 1 have been shown to be active in posterior neural tissue. Data from the present study indicate that the upstream elements drive reduced expression within posterior neural tube relative to the endogenous gene. It is interesting to speculate that the upstream and intron elements might have predominant roles in regulating expression within the mesoderm and neuroectoderm, respectively.

Vertebrate Cdx genes exhibit overlapping but distinct patterns of expression in early development (Marom et al., 1997; Pillemer et al., 1998a). The identification of conserved regulatory elements, such as the Oct, ets, and Tcf responsive elements, is beginning to elucidate the pathways common to the regulation of the Cdx genes. These regulatory pathways are likely to underlie the shared features of Cdx expression. An emerging challenge will be to determine what mechanisms are involved elaborating the aspects of expression distinct to each Cdx family member.

Acknowledgments

Grant sponsors: Isaacs and Reece-Hoyes: Wellcome Trust, University Award; Keenan: Biotechnology and Biological Sciences Research Council research studentship; Pownall: Wellcome Trust project grant.

References

Andersen, B., Rosenfeld, M.G., 2001. POU domain factors in the neuroendocrine system: lessons from developmental biology provide insights into human disease. Endocr. Rev. 22, 2–35.

Arnosti, D.N., Merino, A., Reinberg, D., Schaffner, W., 1993. Oct-2 facilitates functional preinitiation complex assembly and is continuously required at the promoter for multiple rounds of transcription. EMBO J. 12, 157–166.

Bai, Y.Q., Yamamoto, H., Akiyama, Y., Tanaka, H., Takizawa, T., Koike, M., Kenji Yagi, O., Saitoh, K., Takeshita, K., Iwai, T., Yuasa, Y., 2002. Ectopic expression of homeodomain protein CDX2 in intestinal metaplasia and carcinomas of the stomach. Cancer Lett. 176, 47–55.

Beland, M., Pilon, N., Houle, M., Oh, K., Sylvestre, J.R., Prinos, P., Lohnes, D., 2004. Cdx1 autoregulation is governed by a novel Cdx1-LEF1 transcription complex. Mol. Cell. Biol. 24, 5028–5038.

Brooke, N.M., Garcia-Fernàndez, J., Holland, P.W., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922.

Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E., Kadonaga, J.T., 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harbor Symp. Quant. Biol. 63, 75–82.

Butler, J.E., Kadonaga, J.T., 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev. 15, 2515–2519.

Cao, Y., Knochel, S., Donow, C., Miethe, J., Kaufmann, E., Knochel, W., 2004. The POU factor Oct-25 regulates the Xvent-2B gene and counteracts terminal differentiation in *Xenopus* embryos. J. Biol. Chem. 279, 43735–43743.

Charité, J., de Graaff, W., Consten, D., Reijnen, M.J., Korving, J., Deschamps, J., 1998. Transducing positional information to the Hox genes: critical interaction of cdx gene products with position-sensitive regulatory elements. Development 125, 4349–4358.

Chawengsaksophak, K., James, R., Hammond, V.E., Kontgen, F., Beck, F., 1997. Homeosis and intestinal tumours in Cdx2 mutant mice. Nature 386, 84–87.

Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J., Beck, F., 2004. Cdx2 is essential for axial elongation in mouse development. Proc. Natl. Acad. Sci. U. S. A. 101, 7641–7645.

Christen, B., Slack, J., 1997. FGF-8 is associated with anterioposterior patterning and limb regeneration in *Xenopus*. Dev. Biol. 192, 455–466.

Church, G.M., Gilbert, W., 1984. Genomic sequencing. Proc. Natl. Acad. Sci. U. S. A. 81, 1991–1995.

Dick, M.H., Buss, L.W., 1994. A PCR-based survey of homeobox genes in *Ctenodrilus serratus* (Annelida: Polychaeta). Mol. Phylogenet. Evol. 3, 146–158.

Epstein, M., Pillemer, G., Yelin, R., Yisraeli, J.K., Fainsod, A., 1997. Patterning of the embryo along the anterior–posterior axis: the role of the caudal genes. Development 124, 3805–3814.

Guo, R.J., Suh, E.R., Lynch, J.P., 2004. The role of Cdx proteins in intestinal development and cancer. Cancer Biol., 3.

Haremaki, T., Tanaka, Y., Hongo, I., Yuge, M., Okamoto, H., 2003. Integration of multiple signal transducing pathways on Fgf response elements of the *Xenopus* caudal homologue Xcad3. Development 130, 4907–4917.

Harland, R.M., 1991. In situ hybridization: an improved whole-mount method for *Xenopus* embryos. Methods Cell Biol. 36, 685–695.

Houle, M., Prinos, P., Iulianella, A., Bouchard, N., Lohnes, D., 2000. Retinoic acid regulation of Cdx1: an indirect mechanism for retinoids and vertebral specification. Mol. Cell. Biol. 20, 6579–6586.

Hunter, C.P., Kenyon, C., 1996. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in *C. elegans* embryos. Cell 87, 217–226.

Ikeya, M., Takada, S., 2001. Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of cdx-1 expression. Mech. Dev. 103, 27–33.

Isaacs, H.V., Tannahill, D., Slack, J.M.W., 1992. Expression of a novel FGF in the *Xenopus* embryo. A new candidate inducing factor for mesoderm formation and anteroposterior specification. Development 114, 711–720.

Isaacs, H.V., Pownall, M.E., Slack, J.M.W., 1998. Regulation of Hox gene expression and posterior development by the *Xenopus* caudal homologue Xcad3. EMBO J. 17, 3413–3427.

Isaacs, H.V., Andreazzoli, M., Slack, J.M.W., 1999. Anteroposterior patterning by mutual repression of orthodenticle and caudal-type transcription factors. Evol. Dev. 1, 143–152.

Ishii, Y., Fukuda, K., Saiga, H., Matsushita, S., Yasugi, S., 1997. Early specification of intestinal epithelium in the chicken embryo: a study on the localization and regulation of CdxA expression. Dev. Growth Differ. 39, 643–653.

Jin, T., Li, H., 2001. Pou homeodomain protein OCT1 is implicated in the expression of the caudal-related homeobox gene Cdx-2. J. Biol. Chem. 276, 14752–14758.

Kroll, K.L., Amaya, E., 1996. Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. Development 122, 3173–3183.

Krumlauf, R., 1994. Hox genes in vertebrate development. Cell 78, 191–201.

Levy, T., Suh, E.R., Siang, S., Traber, P.G., 1995. The intestine-specific homeobox gene Cdx2 activates intestinal gene-transcription via both promoter and enhancer elements. Gastroenterology 108, A301.

Lickert, H., Kemler, R., 2002. Functional analysis of cis-regulatory elements controlling initiation and maintenance of early Cdx1 gene expression in the mouse. Dev. Dyn. 225, 216–220.

Lickert, H., Domon, C., Huls, G., Wehrle, C., Duluc, I., Clevers, H., Meyer, B.I., Freund, J.N., Kemler, R., 2000. Wnt/(beta)-catenin signaling regulates the expression of the homeobox gene Cdx1 in embryonic intestine. Development 127, 3805–3813.

Lohnes, D., 2003. The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. Bioessays 25, 971–980.

Lynch, J., Suh, E.R., Silberg, D.G., Rulyak, S., Blanchard, N., Traber, P.G., 2000. The caudal-related homeodomain protein Cdx1 inhibits proliferation of intestinal epithelial cells by down-regulation of D-type cyclins. J. Biol. Chem. 275, 4499–4506.

MacDonald, P.M., Struhl, G., 1986. A molecular gradient in early *Drosophila* embryos. Nature 324, 537–545.

Marom, K., Shapira, E., Fainsod, A., 1997. The chicken *caudal* genes establish an anterior–posterior gradient by partially overlapping temporal and spatial patterns of expression. Mech. Dev. 64, 41–52.

McDonald, P.M., Struhl, G., 1986. Molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. Nature 324, 537–545.

Nieuwkoop, P.D., Faber, J., 1967. Normal table of *Xenopus laevis* (Daudin). North-Holland, Amsterdam.

Northrop, J.L., Kimelman, D., 1994. Dorsal–ventral differences in Xcad-3 expression in response to FGF-mediated induction in *Xenopus*. Dev. Biol. 161, 490–503.

Phillips, R.W., Frierson Jr., H.F., Moskaluk, C.A., 2003. Cdx2 as a marker of epithelial intestinal differentiation in the esophagus. Am. J. Surg. Pathol. 27, 1442–1447.

Pillemer, G., Epstein, M., Blumberg, B., Yisraeli, J.K., De Robertis, E.M., Steinbeisser, H., Fainsod, A., 1998a. Nested expression and sequential downregulation of the *Xenopus* caudal genes along the anterior–posterior axis. Mech. Dev. 71, 193–196.

Pillemer, G., Yelin, R., Epstein, M., Gont, L., Frumkin, Y., Yisraeli, J.K., Steinbeisser, H., Fainsod, A., 1998b. The Xcad-2 gene can provide a ventral signal independent of BMP-4. Mech. Dev. 74, 133–143.

Pownall, M.E., Tucker, A.S., Slack, J.M.W., Isaacs, H.V., 1996. eFGF, Xcad3 and Hox genes form a molecular pathway that establishes the anteroposterior axis in *Xenopus*. Development 122, 3881–3892.

Pownall, M.E., Isaacs, H.V., Slack, J.M., 1998. Two phases of Hox gene regulation during early *Xenopus* development. Curr. Biol. 8, 673–676.

Prinos, P., Joseph, S., Oh, K., Meyer, B.I., Gruss, P., Lohnes, D., 2001. Multiple pathways governing Cdx1 expression during murine development. Dev. Biol. 239, 257–269.

Rawat, V.P., Cusan, M., Deshpande, A., Hiddemann, W., Quintanilla-Martinez, L., Humphries, R.K., Bohlander, S.K., Feuring-Buske, M., Buske, C., 2004. Ectopic expression of the homeobox gene Cdx2 is the transforming event in a mouse model of t(12;13)(p13;q12) acute myeloid leukemia. Proc. Natl. Acad. Sci. U. S. A. 101, 817–822.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Shiotsugu, J., Katsuyama, Y., Arima, K., Baxter, A., Koide, T., Song, J., Chandraratna, R.A., Blumberg, B., 2004. Multiple points of interaction between retinoic acid and FGF signaling during embryonic axis formation. Development 131, 2653–2667.

Silberg, D., Furth, E., Taylor, J., Schuck, T., Chiou, T., Traber, P., 1997. CDX1 protein expression in normal, metaplastic, and neoplastic human alimentary tract epithelium [see comments]. Gastroenterology 113, 478–486.

Sparrow, D.B., Latinkic, B., Mohun, T.J., 2000. A simplified method of generating transgenic *Xenopus*. Nucleic Acids Res. 28, E12.

Subramanian, V., Meyer, B.I., Gruss, P., 1995. Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. Cell 83, 641–653.

Tamai, Y., Nakajima, R., Ishikawa, T., Takaku, K., Seldin, M.F., Taketo, M.M., 1999. Colonic hamartoma development by anomalous duplication in Cdx2 knockout mice. Cancer Res. 59, 2965–2970.

Veenstra, G., Beumer, T., PetersonMaduro, J., Stegeman, B., Karg, H., Van der Vliet, P., Destree, O., 1995. Dynamic and differential Oct-1 expression during early *Xenopus* embryogenesis: persistence of Oct-1 protein following down-regulation of the RNA. Mech. Dev. 50, 103–117.

Whitfield, T., Heasman, J., Wylie, C., 1993. XLPOU-60, a *Xenopus* POU-domain mRNA, is oocyte-specific from very early stages of oogenesis, and localised to presumptive mesoderm and ectoderm in the blastula. Dev. Biol. 155, 361–370.

Xu, F., Li, H., Jin, T., 1999. Cell type-specific autoregulation of the Caudal-related homeobox gene Cdx-2/3. J. Biol. Chem. 274, 34310–34316.

Zwilling, S., Annweiler, A., Wirth, T., 1994. The POU domains of the Oct1 and Oct2 transcription factors mediate specific interaction with TBP. Nucleic Acids Res. 22, 1655–1662.