Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 341 (2006) 225–231
BBRC
www.elsevier.com/locate/ybbrc

Cloning and characterization of the 5′-flanking region of the *Ehox* gene

Woon Kyu Lee [a,c,*], Yong-Man Kim [b,1], Nasir Malik [a], Chang Ma [a], Heiner Westphal [a]

[a] *Laboratory of Mammalian Genes and Development, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA*
[b] *Genetic Pharmacology Unit, Experimental Therapeutic Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA*
[c] *Department of Laboratory Animal Medicine, Medical Research Center, College of Medicine, Yonsei University, 134 Sinchon-dong, Seodaemun-gu, Seoul 120-752, Republic of Korea*

Received 5 December 2005
Available online 9 January 2006

Abstract

The paired-like homeobox-containing gene *Ehox* plays a role in embryonic stem cell differentiation and is highly expressed in the developing placenta and thymus. To understand the mechanisms of regulation of *Ehox* gene expression, the 5′-flanking region of the *Ehox* gene was isolated from a mouse BAC library. 5′-RACE analysis revealed a single transcriptional start site 130 nucleotides upstream of the translation initiation codon. Transient transfection with a luciferase reporter gene under the control of serially deleted 5′-flanking sequences revealed that the nt −84 to −68 region contained a positive *cis*-acting element for efficient expression of the *Ehox* gene. Mutational analysis of this region and oligonucleotide competition in the electrophoretic mobility shift assay revealed the presence of a CCAAT box, which is a target for transcription nuclear factor Y (NFY). NFY is essential for positive gene regulation. No tissue-specific enhancer was identified in the 1.9-kb 5′-flanking region of the *Ehox* gene. *Ehox* is expressed during the early stages of embryo development, specifically in the brain at 9.5 dpc, as well as during the late stages of embryo development. These results suggest that NFY is an essential regulatory factor for *Ehox* transcriptional activity, which is important for the post-implantation stage of the developing embryo.

Keywords: *Ehox*; Nuclear factor Y; 5′-Flanking

Homeobox genes, which are transcription factors that are involved in normal differentiation and development, were originally identified as genes in which mutations caused body segment transformation in *Drosophila* [1]. Our BLAST database search using conserved homeodomain sequences showed the existence of more than 225 vertebrata homeobox genes in a variety of species. These genes share a common sequence motif, the homeodomain, which is 183 nucleotides (nt) in length and encodes a 61-amino acid domain that is responsible for DNA binding [2]. While most homeobox genes can be classified based on sequence homologies within the homeodomain and elsewhere in the encoded protein, the remaining genes show more divergence.

The ES cell-derived homeobox (*Ehox*) gene was originally identified by subtractive hybridization as a homeobox-containing gene that is essential during the earliest stages of embryonic stem (ES) cell differentiation [3]. *Ehox* is an X-linked paired-like homeobox gene that is proposed to have a dual role in trophoblast stem cells and compartments of the developing placenta, as well as during development of the pharyngeal pouches, and possibly delineates the area that becomes the thymus [4]. To verify the function of this gene, we cloned the 5′-flanking region of the *Ehox* gene and characterized it in terms of transcriptional regulation.

* Corresponding author. Fax: +1 82 2 362 8647.
E-mail address: wklee@yumc.yonsei.ac.kr (W.K. Lee).
[1] Present address: CB-Pharmicell Co., Ltd, Stem Cell Therapy Institute, #920 Sicox Tower, Jungwon-gu, Sungnam-si, Kyungki-do 462-120, Republic of Korea.

0006-291X/$ - see front matter
doi:10.1016/j.bbrc.2005.12.176

Materials and methods

Isolation of the mouse Ehox gene 5′-flanking region. To screen the mouse bacterial artificial chromosome (BAC) library (ResGen) containing the 5′-flanking region of *Ehox* gene, the following gene-specific primers were used: 5′-GCTTTCGGTTTTCACAACCTC-3′ (primer 1; corresponding to nucleotides +57 to +77 relative to the transcription start site) and 5′-TCTTCCCACCATTCAACTTTTC-3′ (primer 2; +185 to +206); primer specificity was confirmed by reverse transcription polymerase chain reaction (RT-PCR) and genomic PCR (data not shown). PCR amplification was performed in a 50 μl reaction mixture containing 1 μl of the BAC library, 20 pmol of each of the primers, 0.4 mM dNTPs, 3 mM $MgCl_2$, 1.25 U *Taq* polymerase (Roche), and PCR buffer. The following PCR conditions were used: 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 46 °C from 30 s, extension at 72 °C for 45 s, and the final extension step was prolonged to 5 min. DNA from three positive clones was isolated, analyzed by restriction enzyme mapping and Southern blot, and then completely sequenced in both directions.

5′-Rapid amplification of cDNA ends. To determine the transcription initiation site of the Ehox gene, 5′-rapid amplification of cDNA end (5′-RACE) was carried out according to the manufacturer's protocol using Marathon-Ready cDNA from Testis tissue (Clontech). The first PCR was performed using a gene-specific primer 1 (5′-TCTTCTGCTTTGTCCCA TTCCACTGCTG-3′; nt 268–295 of the *Ehox* cDNA; GenBank Accession No. AF265350) and Adaptor primer 1, included in the kit, as primers. The second PCR was performed using a gene-specific primer 2 (5′-TGAAAGCCGACAGCCGACAGCCGAAAG-3′; nt 19–45) and Adaptor primer 2, included in the kit, as primers and the diluted (1/100) reaction mixture of the first PCR as a template. PCR cycling conditions were an initial denaturation step at 94 °C for 1 min followed by 35 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72°C for 1 min. PCR products were separated in 2% agarose gel, cloned into pCRII-TOPO (Invitrogen), and sequenced.

Construction of promoter-luciferase chimeric plasmids. A series of plasmids that contained fragments of various sizes from the 5′-flanking region of the mouse *Ehox* gene were constructed by inserting DNA fragments between the *Xho*I and *Hin*dIII sites of the pGL3-basic vector (Promega). Briefly, 5′-end-deleted DNA fragments were obtained by PCR using the following synthetic oligonucleotides that incorporated a 5′-end *Xho*I site and 3′-end *Hin*dIII site: for the 5′-ends of the inserts, 5′-CCC GCTCGAGATCATGGATTCATTCC-3′ (nt −1987 to −1967, as designated from the transcription start site), 5′-CCCGCTCGAGAACAT TATGTCTGTTAG-3′ (nt −1243 to −1224), 5′-CCCGCTCGAGT AACAGGCCGCTAGTAT-3′ (nt −533 to −515), 5′-CCCGCTCGAGA TCCTGCTCCTTGCAGCA-3′ (nt −176 to −159), 5′-CCCGCTCGAG CCCATAGCTGCAGGAGATC-3′ (nt −126 to −108), 5′-CCCGCTCG AGAAATGGAGGTCAGGATAGC-3′ (nt −107 to −89), 5′-CCCG CTCGAGCTATGATTGGACACTTGGTAAC-3′ (nt −84 to −63), 5′-C CCGCTCGAGGTAACCTGCACTACCGCCCCC-3′ (nt −68 to −47), and 5′-CCCGCTCGAGTTGCAAAGCTAGGTTTTC-3′ (nt −20 to −2); and for the 3′-ends of the inserts, 5′-GACAAGATCTAAGCTTCCCCA TCATATTC-3′ (nt +91 to +108 as designated from the transcription start site). To make the mutant construct of pm-84, the following nucleotide sequences were used as a sense strand for PCR: 5′-CCCGCTCGAGCTATGATAAGACACTTGGTAAC-3′; the NFY-binding sites are boxed and mutations therein are italicized. The amplified DNA fragments were digested with *Xho*I and *Hin*dIII, and inserted immediately upstream of the luciferase reporter gene. The sequences and orientations of the constructs were verified by nucleotide sequencing.

Cell culture and transient transfection assay. Murine NS20Y neuroblastoma cells (obtained from Dr. M. Maral Mouradian, NIH) were grown in Dulbecco's modified Eagle's medium that was supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 μg/ml streptomycin (Gibco-BRL). Cells were plated at 5×10^4 cells per well into 12-well plates. Following 24 h of incubation, transient transfection of plasmids for the luciferase reporter assay was carried out with 0.5 μg of each promoter-reporter construct together with 3 μl of Fugene (Roche) per μg of plasmid DNA. Co-transfection with 0.1 μg of the β-galactosidase expression vector pSV-β (Promega) was used to normalize for transfection efficiency. After incubation for 2 days, cell extracts were prepared and assayed for luciferase and β-galactosidase activities, according to the manufacturer's protocols (Promega). The transfections were performed in triplicate on a minimum of three separate occasions. The data were normalized by calculating the ratio of the luciferase and β-galactosidase activities. The average normalized luciferase activity is presented as fold-activation relative to the activity of the pGL3-basic vector.

Electrophoretic mobility shift assay. The oligonucleotides used as probes for electrophoretic mobility shift assay (EMSA) were annealed to their respective antisense strands and end-labeled with [γ-^{32}P]dATP using T4 polynucleotide kinase. Nuclear extracts for EMSA were prepared as described previously [5]. For the binding reactions, 5 μg of nuclear extract was incubated with reaction buffer [10 mM Tris–HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, 2 μg poly(dI–dC) · poly(dI–dC), and 1 μg BSA] in the presence or absence of competitor for 10 min at room temperature. The radiolabeled double-stranded oligomers (1 ng) were added to the reaction mixture for an additional 20 min at room temperature. In the competition experiments, unlabeled competitors containing intact or mutant NFY consensus sequences were added in 100-fold excess. The binding products were electrophoresed at 4–5 V/cm in a 6% polyacrylamide gel in 0.5× TBE. The gel was dried and analyzed by autoradiography. The DNA sequences of the sense strands of oligonucleotides were as follows: NF-Y, 5′-CACCTATGATTGGACACTTGGTA-3′; and mNFY, 5′-CACCTATGATAAGACACTTGGTA-3′; the NFY-binding sites are underlined and mutations therein are italicized.

Gene expression analysis. In situ hybridization to DNA from 9.5-day post-coitus (dpc) whole embryos was performed as described previously [6] with a 260-bp probe from the *Ehox* cDNA (nt 606–866). The probe was labeled with digoxigenin using the T7 and T3 RNA polymerase (Roche) for the antisense and sense strands, respectively. PCR with mouse Marathon-Ready cDNA (Clontech) was carried out on various tissues. The primer set for the *Ehox* gene was designed to span one intron, thereby eliminating the possibility of genomic contamination. The following nucleotide sequences were used: 5′-AGATGGATAGGTGTGAGTGAA G-3′ and 5′-GTCAAGGGTGGTACACCAG-3′. The PCR conditions were as follows: 94 °C for 2 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 63 °C for 45 s, extension at 72 °C for 45 s, and the final extension step was prolonged to 3 min. The *GAPDH* gene was used as an internal control with the following primer sequences: 5′-TG GCCAAGGTCATCCATGACAACTTTG-3′ and 5′-GCCTGCTTCAC CACCTTCTTGATGTCA-3′. For *GAPDH*, the following PCR conditions were used: 94 °C for 2 min, and then 30 cycles of denaturation at 94 °C for 30 s, annealing/extension at 68 °C for 60 s, and the final extension step was prolonged to 5 min.

Results and discussion

Genomic structure of the Ehox DNA

The mouse BAC DNA pool was screened by PCR using a primer set for the extreme 5′-end of the *Ehox* cDNA. Three positive clones were identified, the DNA inserts of which were subjected to restriction endonuclease mapping and Southern blot analysis. A total of 9.4 kb of the mouse 5′-UTR and full-reading frame of the *Ehox* gene was subcloned into the pBluescript II SK(+) vector at the *Bam*HI and *Xho*I sites, and sequenced. The *Ehox* gene spans ~4.8 kb of the genome and is organized into four exons and three introns (Table 1). We did not find any comparable sequences in any other species, with the exception of the rat (GenBank Accession No. NM_001024889).

Table 1
Exon and intron junctions of the Ehox gene[a]

Exon number	Exon size (bp)	Sequence of exon–intron junction		Intron size (bp)
		5′ splice donor	3′ splice donor	
1	197	**GAATG**GTGAG	TACAG**GTGGG**	190
2	406	**GAAAG**GTAAG	CCCAG**AAGAC**	2,401
3	46	**TTAAG**GTAAG	GGTAG**AGATG**	1,313
4	237			

[a] Exon sequences are shown in bold. The 5′ donors, GT, and the 3′ donors, AG, are underlined. The sites of exon 1 and exon 4 are calculated from position +1 and to the end of the cDNA, respectively.

Transcription initiation site and upstream promoter region

To determine the transcription initiation sites of the *Ehox* gene, 5′-RACE was performed, as described in Materials and methods. Three positive clones were sequenced, all of which had a transcript that was initiated from the guanidine residue 130 bp upstream of the ATG codon (Fig. 1). In the proximal upstream region from this

Fig. 1. Nucleotide sequences of the 5′-flanking region of the mouse *Ehox* gene. The major transcription initiation site (identified in our 5′-RACE analysis), a guanidine, is designated as the +1 nucleotide (bold). Potential *cis*-acting elements are underlined with the names indicated below the line. The ATG translation initiation codon is boxed.

guanidine (designated as +1), several consensus-binding sites for transcription factors, which included GATA [7], DeltaEF1 [8], NKX25 [9], USF [10], AP4, AP1, NF-Y [11], and SP1 [12], were identified.

Transcriptional regulation of the Ehox gene

The transcriptional activity of the *Ehox* gene was examined using a transient transfection method. To determine the regulatory elements responsible for *Ehox* gene expression, a series of deletion constructs of the 5′-region of the gene was generated (Fig. 2A). These deletion constructs were transiently transfected into NS20Y cells, and the levels of luciferase activity were determined. Fig. 2B shows that deletion up to position −84 did not result in any significant changes in the promoter activity, relative to the luciferase activity of the longest *Ehox* promoter-reporter construct (nt −1987 to +108). The promoter activity was lost in the construct that contained nt −68 to +108. These results indicate that the minimal promoter comprises a region that spans from nt −68 to +108, and potential positive regulatory elements seem to be located in the region from nt −84 to −68, a region that is essential for high-level expression of the *Ehox* gene. The region from nt −84 to −68 contains the CCAAT box, a target for the transcription factor NFY, which is involved in the activation of the *Ehox* promoter. For further analysis, a 2-bp mutation of the NFY-binding site (CCAAT → CTTAT) was introduced into the p-84 plasmid, to generate p-m84 (Fig. 3A). As shown in Fig. 3B, NS20Y cells transfected with p-m84 showed markedly lower luciferase levels than cells transfected with the wild-type p-84.

In an attempt to identify transcription factors that have a potential to interact with the *Ehox* promoter region, a synthetic double-stranded DNA probe for NFY was subjected to a gel retardation assay. As shown in Fig. 4, DNA–protein complexes were formed with the nuclear extracts. The specificities of these complexes for the sequence were shown in a competition experiment, in which the complexes were competed by a 100-fold excess of unlabeled wild-type probe, but not by a 100-fold excess of unlabeled mutant-type probe. Typically,

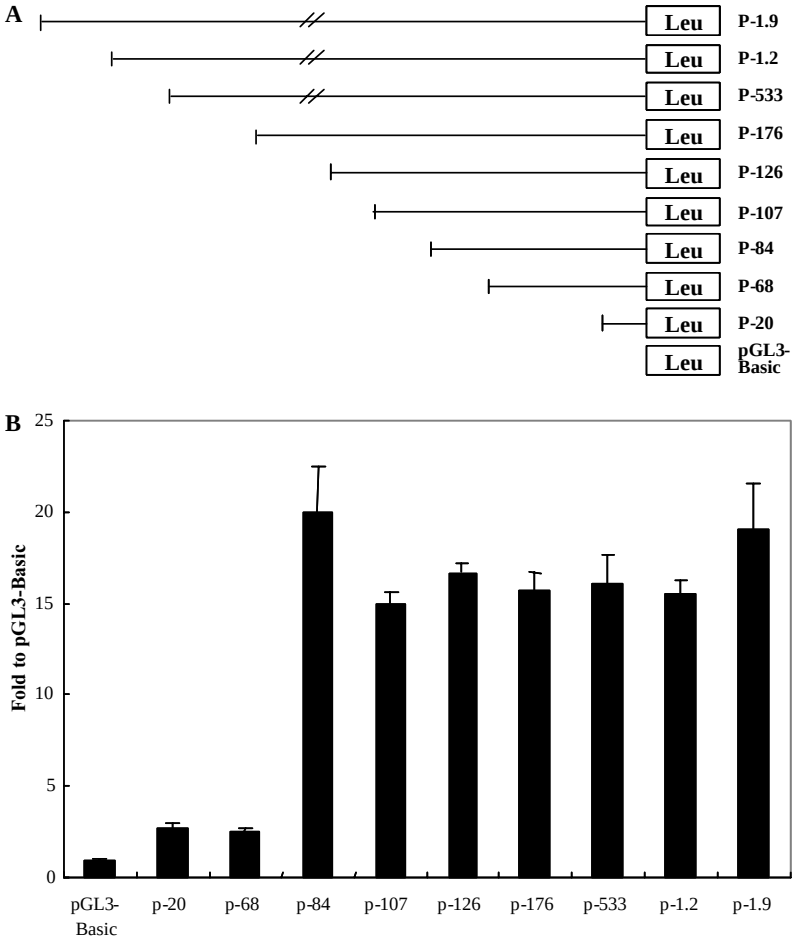

Fig. 2. Deletion analysis of the *Ehox* gene promoter. Reporter plasmids, which contain the luciferase gene under the control of various fragments of the Ehox promoter, were transfected into NS20Y cells. The promoter regions used for the assay are shown (A). The activity of the pGL3-basic vector is given the value of 1, and the relative fold-activity of the *Ehox* promoter is shown (B). Each activity value represents the average of at least three independent experiments and is normalized by the activity of the co-transfected pSV-β plasmid.

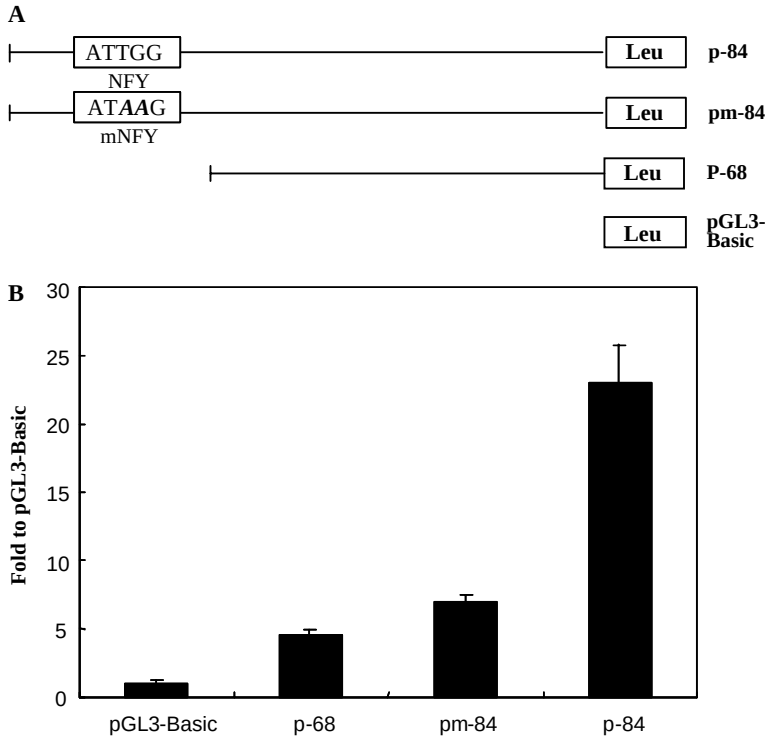

Fig. 3. Effects of site-directed mutations within the NFY-binding region of the *Ehox* promoter. (A) Reporter constructs that contain the NFY site or a mutation in the NFY site. The altered nucleotides in the NFY site are shown as bold and italicized characters. (B) The indicated constructs were transfected into NS20Y cells and cultured for 2 days. Each activity value represents the average of at least three independent experiments and is normalized by the activity of the co-transfected pSV-β plasmid.

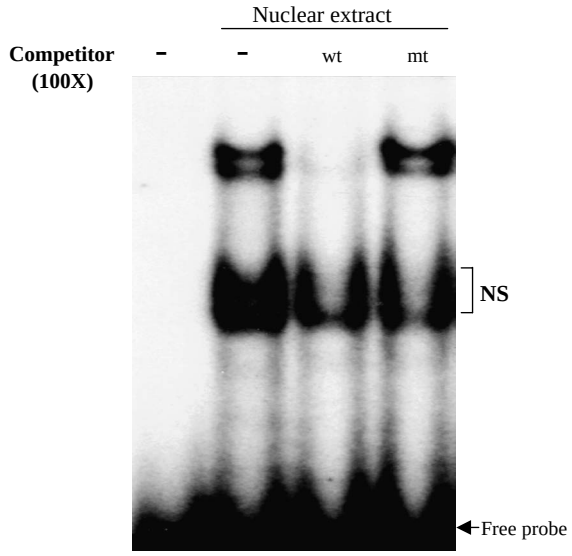

Fig. 4. Binding analysis of the NFY probe. EMSA with radiolabeled oligomers that contain the NFY site was performed in the absence or presence of an excess of each specific oligomer or mutated oligomer. NS, non-specific.

NFY occupies a relatively fixed position in the promoter, usually between nt −80 and −60 upstream of the transcription start site [13], which matches exactly our mapping of the *Ehox* promoter. The interaction of NFY with the CCAAT element is necessary for basal constitutive promoter activation, although NFY also mediates cell-specific and inducible gene expression [14]. It has been reported that *Ehox* is highly expressed in the thymus, skin, and testis [3]. To verify NFY regulation of tissue specificity by binding to the 1.9-kb *Ehox* 5′-flanking region, we examined the *Ehox* promoter activities of cell lines 2B4 (thymus), F9 (testis), and NS20Y (brain) (Fig. 5). The cell lines did not differ with respect to *Ehox* expression, which suggests that additional control elements govern the tissue specificity of NFY-mediated basal control of gene expression.

Ehox expression patterns

To examine *Ehox* expression during development, we performed whole mount in situ hybridization and PCR. As shown in Fig. 6A, *Ehox* expression was widespread in the brain at 9.5 dpc. In addition, *Ehox* was expressed in the adult brain, albeit at very low levels, compared with the testis and developing embryo (Fig. 6B). *Ehox* is known to be expressed in trophoblast stem cells, compartments of the developing placenta, and the pharyngeal pouch during embryo development [3,4]. Recently, it has been reported that *Ehox* is expressed in the ovary and testis as a reproductive homeobox on the X chromosome (Rhox) family member [15]. It is also known that *Ehox* is not expressed in the brain, either during development or adulthood [3,4,15]. The reasons for the observed

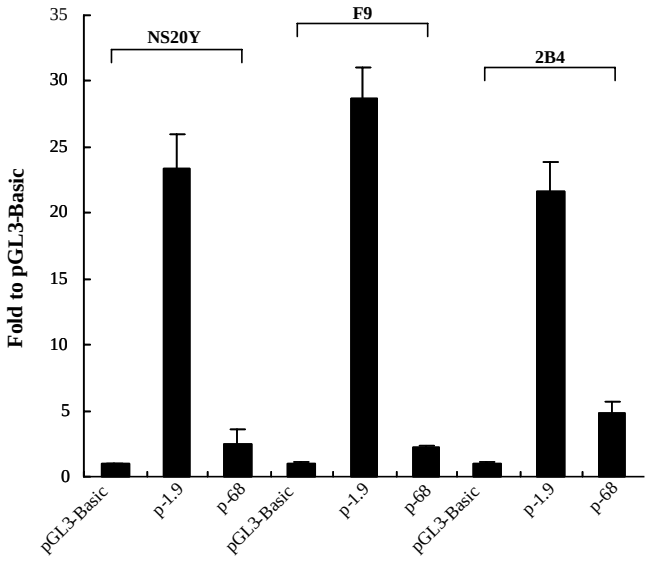

Fig. 5. *Ehox* promoter activities of different cell lines. The luciferase reporter plasmid was transfected into the NS20Y, F9, and 2B4 cell lines, and the cells were cultured for 2 days. Each activity value represents the average of at least three independent experiments and is normalized by the activity of the co-transfected pSV-β plasmid.

discrepancies are unclear. We designed the PCR primers to be specific for *Ehox* cDNA and sequenced the PCR products, to reduce the risk of genomic contamination. Furthermore, the RNA 260-bp pRNA probe, which encompasses part of exon 4 and the 3′-UTR, was narrowly defined, as we did not obtain any specific in situ results using any other region, with the exception of part of the 5′-UTR (data not shown).

On several occasions, we attempted to generate chimeric mice from the *Ehox*-hemizygous mutant ES cells. However, we obtained only three pups, all of which died within 2 months without obvious crucial defects. This suggests that *Ehox* plays a critical functional role during development. How early *Ehox* is expressed remains unclear. We noted *Ehox* expression in blastocyst embryos as well as during ES cell differentiation (data not shown), as reported by other groups [3,4]. *Ehox* was expressed in 7.5-dpc embryos, as well as in 9.5-dpc and 11.5-dpc embryos (Fig. 6), which indicates that Ehox is expressed throughout the post-implantation stage of the developing embryo. It is clear that NFY is an important regulator of *Ehox* transcriptional activity, since NFY binding to the CCAAT box is often essential for gene transcription, particularly that of TATA-less genes [16]. Taken together, our results suggest that NFY is an essential regulatory factor for the *Ehox* transcriptional activity that occurs throughout the post-implantation stage of the developing embryo. Further studies are required to define the factors and elements necessary for tissue-specific *Ehox* gene transcription.

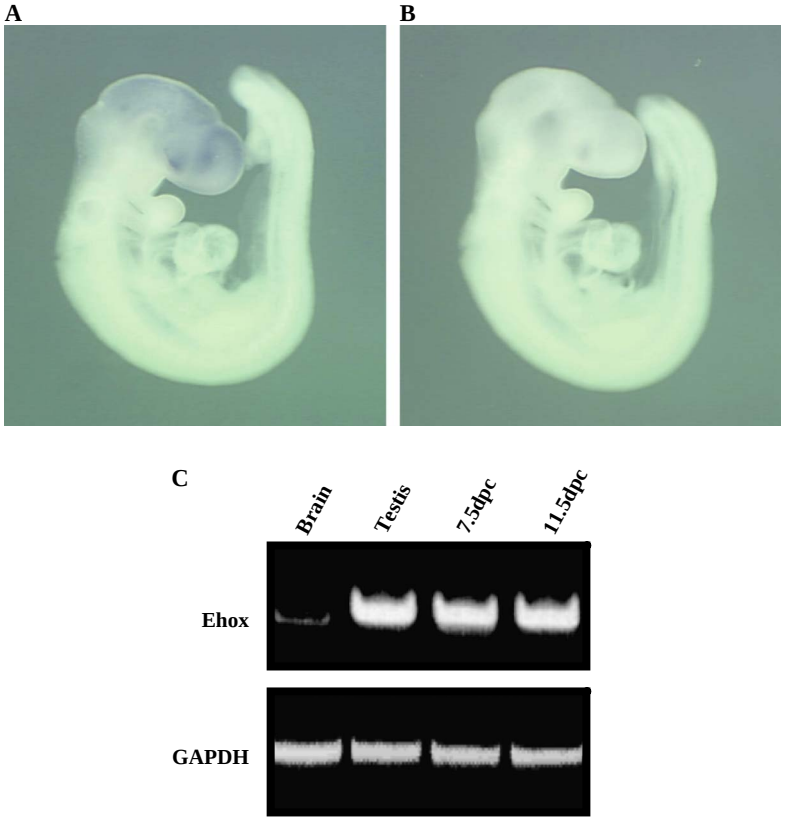

Fig. 6. *Ehox* gene expression patterns. In situ hybridization to 9.5-dpc whole embryos was performed with antisense (A) and sense (B) probes for the *Ehox* cDNA. PCR for the *Ehox* gene and the internal control gene *GAPDH* was performed on samples from various tissues (C).

References

[1] W.J. Gehring, M. Affolter, T. Burglin, Homeodomain proteins, Annu. Rev. Biochem. 63 (1994) 487–526.

[2] W. McGinnis, R. Krumlauff, Homeobox genes and axial patterning, Cell 68 (1992) 283–302.

[3] M. Jackson, J.W. Baird, N. Cambray, J.D. Ansell, L.M. Forrester, G.J. Graham, Cloning and characterization of Ehox, a novel homeobox gene essential for ES cell differentiation, J. Biol. Chem. 277 (2002) 38683–38692.

[4] M. Jackson, J.W. Baird, J. Nichols, R. Wilkie, J.D. Ansell, G. Graham, L.M. Forrester, Expression of a novel homeobox gene *Ehox* in trophoblast stem cells and pharyngeal pouch endoderm, Dev. Dyn. 228 (2003) 740–744.

[5] Y.M. Kim, B.S. Lee, K.Y. Yi, S.G. Paik, Upstream NF-κB site is required for the maximal expression of mouse inducible nitric oxide synthase gene in interferon-γ plus lipopolysaccharide-induced Raw 264.7 macrophages, Biochem. Biophys. Res. Commun. 236 (1997) 655–660.

[6] M. Mukhopadhyay, A. Teufel, T. Yamashita, A.D. Agulnick, L. Chen, K.M. Downs, A. Schindler, A. Grinberg, S.P. Huang, D. Dorward, H. Westphal, Functional ablation of the mouse *Ldb1* gene results in severe patterning defects during gastrulation, Development 130 (2003) 495–505.

[7] L.J. Ko, J.D. Engel, DNA-binding specificities of the GATA transcription factor family, Mol. Cell. Biol. 13 (1993) 4011–4022.

[8] R. Sekido, K. Murai, J.I. Funahashi, Y. Kamachi, A. Fujisawa-Sehara, Y.I. Nabeshima, H. Kondoh, The delta-crystallin enhancer-binding protein deltaEF1 is a repressor of E2-box-mediated gene activation, Mol. Cell. Biol. 14 (1994) 5692–5700.

[9] C.Y. Chen, R.J. Schwartz, Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, Nkx-2.5, J. Biol. Chem. 270 (1995) 15628–15633.

[10] A.J. Bendall, P.L. Molloy, Base preference for DNA binding by the bHLH-Zip protein USF: effects of magnesium chloride on specificity and comparison with binding of Myc family members, Nucleic Acids Res. 22 (1994) 2801–2810.

[11] R. Mantovani, A survey of 178 NF-Y binding CCAAT boxes, Nucleic Acids Res. 26 (1998) 1135–1143.

[12] E.A. Wimmer, H. Jackle, C. Pfeifle, S.M. Cohen, A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene, Nature 366 (1993) 690–694.

[13] P. Bucher, Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences, J. Mol. Biol. 212 (1990) 563–578.

[14] Y. Peng, N. Jahroudi, The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter, Blood 99 (2002) 2408–2417.

[15] J.A. MacLean II, M.A. Chen, C.M. Wayne, S.R. Bruce, M. Rao, M.L. Meistrich, C. Macleod, M.F. Wilkinson, Rhox: A new homeobox gene cluster, Cell 120 (2005) 369–382.

[16] M. Nicolas, V. Noe, C.J. Ciudad, Transcriptional regulation of the human Sp1 gene promoter by the specificity protein (Sp) family members nuclear factor Y (NFY) and E2F, Biochem. J. 371 (2003) 266–275.