Isolation and mapping of EVX1, a human homeobox gene homologous to *even-skipped*, localized at the 5' end of HOX1 locus on chromosome 7

Antonio Faiella, Maurizio D'Esposito, Marco Rambaldi, Dario Acampora, Silvia Balsofiore, Anna Stornaiuolo, Antonio Mallamaci, Enrica Migliaccio, Massimo Gulisano, Antonio Simeone and Edoardo Boncinelli*

International Institute of Genetics and Biophysics, CNR, via Marconi, 10, 80125 Naples, Italy

ABSTRACT

We isolated and mapped the human homeobox gene EVX1. This gene encodes a protein of 407 amino acid residues containing a homeodomain closely related to the Drosophila *even-skipped* (*eve*) segmentation gene of the pair-rule class. EVX1 belongs to a small family of vertebrate *eve*-related homeobox genes including human EVX1 and EVX2 genes, their murine homologs, *Evx 1* and *Evx 2*, and the frog Xhox-3 gene. We previously reported that EVX2 is localized at the 5' end of the HOX4 locus on chromosome 2. We show here that EVX1 is localized at the 5' end of the HOX1 locus on chromosome 7, 48 kb upstream from the most 5' of the eleven HOX1 genes, namely HOX1J. Both EVX genes are transcribed in an opposite orientation as compared to that of adjacent HOX genes. Human HOX1 and HOX4 complex loci appear to be both closely linked to a homeobox gene of the EVX family.

INTRODUCTION

Studies on developmental mutants of *D. melanogaster* have identified three groups of genes controlling body plan formation (1–3): (a) maternal-effect genes, which dictate the structure and the spatial coordinates of the egg, (b) segmentation genes, which determine the number and polarity of body segments and (c) homeotic genes, which specify the identity of each segment. Many of these genes code for transcription factors containing a homeodomain, the DNA-binding domain encoded by the homeobox (4–6). Homeodomains encoded by fly genes belong to at least seven classes according to their primary sequence (7): the *Antennapedia* (*Antp*), *bicoid* (*bcd*), *caudal* (*cad*), *engrailed* (*en*), *even-skipped* (*eve*), *muscle segment* (*msh*) and *paired* (*prd*) type homeodomains (7, 8).

It is now well established that genes containing homeobox sequences encode nuclear proteins with regulatory functions in a wide variety of organisms, from yeast to man (9–13). In particular, murine homologs of all *Drosophila* homeodomain classes have been described, with the exception of the *bcd* type (7). Class I homeoboxes are defined as those encoding homeodomains most closely related to the *Antp* homeodomain. In *Drosophila*, homeotic genes containing a class I homeobox are clustered in two complex loci, the *Antennapedia*-complex (ANT-C) and the *bithorax*-complex (BX-C) (14) and there is a correlation between the physical order of these genes within the complexes and their expression along the antero-posterior body axis (1). Mouse (15,16) and human (17–19) class I homeobox genes appear to be clustered in a similar way in restricted genomic regions (HOX loci) of at least four chromosomes. These HOX loci appear to be homologous to each other and to the *Drosophila* homeotic gene complexes. Moreover, homeobox genes in the four HOX loci display a similar correlation between their position in the cluster and the relative expression along the antero-posterior axis in the neuroectoderm and mesoderm. In fact, they are expressed along the embryonic antero-posterior axis following a 5'-posterior/3'-anterior rule, even if their expression domains largely overlap (15,16).

We previously reported the isolation of 38 human HOX homeoboxes (18–20) clustered in four HOX loci. 11 homeobox genes are located at HOX1 on chromosome 7 and 9 homeobox genes belong to HOX2, HOX3 and HOX4 on chromosome 17, 12 and 2, respectively.

The isolation of two murine genes, *Evx 1* and *Evx 2*, has recently been reported (21). These genes contain a homeobox related to that present in a fly segmentation gene, namely *eve* (22). The *Evx 1* expression pattern during mouse embryogenesis has been shown to be consistent with a role in establishing the antero-posterior axis. We previously reported the isolation and mapping of EVX2, the human homolog of *Evx 2* (23) and showed that it is located at the 5' end of HOX4 on chromosome 2 even if with a transcriptional orientation which is opposite to that of the genes of the HOX4 cluster. Southern blot analysis on human DNA using EVX2 as a probe provided evidence for the presence

---

* To whom correspondence should be addressed

of a second homologous gene. We now report the isolation of EVX1, the human homolog of *Evx 1*. We show that also this second EVX gene is tightly linked to a HOX locus in a similar arrangement. It maps at the 5' end of HOX1 on chromosome 7.

MATERIALS AND METHODS

Genomic and cDNA clones

Human genomic libraries in pcos2EMBL cosmid vector and in EMBL3 phage derivative were screened as previously reported (18). We previously published genomic maps of the four HOX loci (18,20). These were obtained by chromosome walking around the homeobox sequences first isolated. Every clone overlaps for at least 7 kb adjacent clones. Genomic maps obtained from isolated clones were confirmed by restriction analysis of human DNA extracted from peripheral lympho-cytes. A cDNA library from human fetal brain constructed in lambdagt10 was obtained from Clontech (Palo Alto, California). DNA fragments of interest were subcloned and nucleotide sequences were determined according to Sanger et al. (24).

RNA isolation and analysis

NTERA-2 clone D1 (NT2/D1) human EC cells (25) were maintained at high density in Dulbecco's modified minimal essential medium supplemented with 10% fetal calf serum. Cultures to be exposed to retinoic acid (RA) were established by seeding cells at a density of 10^6 cells per 75 cm^2 tissue culture flask. RA (10mM solution in dimethyl sulfoxide, all-*trans*; Eastman Kodak, Rockester, NY) was added to a final concentration of 10μM. Cells were refed every 48 hr with fresh medium containing RA. Dif-ferentiation was monitored by immunochemical staining (26). Undifferentiated cells, as monitored by immuno-chemical staining, decrease steadily after the addition of RA, are <2% after 7 days and very few after 14 days of treatment. Differentiation was marked by the appearance of several cell types, including neurons. For cycloheximide (CHX) treatment, cells were cultured for some hours in medium containing 20 μg/ml cycloheximide (Sigma). Seven-week human embryos were obtained virtually intact by legal curettage abortions (17).

Total RNA was extracted by the guanidinium thiocyanate technique (27) and poly(A)$^+$-selected by one passage on oligo(dT)-cellulose columns. Poly(A)$^+$ RNA was run on 1.0% agarose-formaldehyde gels, transferred to nitrocellulose (Schleicher & Schuell, BA-85) membranes by Northern capillary blotting, and hybridized to 10^7 cpm of DNA probe labeled by nick translation to a specific activity of $3-8\times10^8$ dpm/μg. Prehybridization and hybridization were carried out as described elsewhere (20). After a washing under stringent conditions (15mM NaCl/1.5 mM sodium citrate (1×SSC), 0.1% sodium dodecyl sulfate (SDS) at 65°C), the blots were exposed for 1−7 days at −70°C to Kodak XR-5 films in an X-omatic intensifying screen cassette.

For RNase protection analysis, a 418 bp Not I DNA fragment including the 3' half of the EVX2 homeobox was subcloned in a pGEM3 vector (Promega Biotec). An antisense strand RNA probe was synthesized with SP6 polymerase and hybridized to 15 μg RNA at appropriate temperature. RNase digestion and electrophoresis on 7% urea-polyacrilamide gels were as described previously (28).
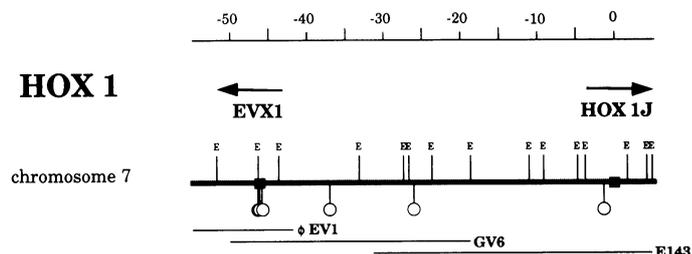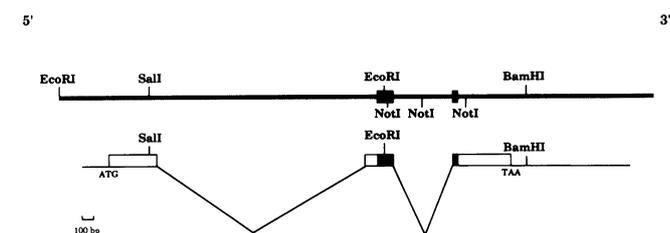
Figure 1. A) Genomic organization of mapped regions at the 5' end of the human HOX1 locus on chromosome 7. Identified homeobox sequences are shown as filled boxes. Transcription of HOX1J and of the other ten HOX1 genes is from left to right, whereas EVX1 appears to be transcribed in opposite orientation (arrows). The map derives from the analysis of overlapping phage and cosmid clones reported below. Open small circles below the map indicate non-methylated Not I sites. E=EcoRI. B) Genomic organization of EVX1 obtained from the comparison of a cDNA clone (below) with the corresponding genomic sequences. The 2500 bp cDNA clone does not extend up to the poly-A tail. Boxes indicate coding regions and filled boxes homeobox regions. The EVX1 homeobox is interrupted by an intron between the 46th and the 47th codon.

RESULTS AND DISCUSSION

Isolation of EVX1

We screened at low stringency a human genomic library in EMBL3 phage derivative with EVX2 sequences. Restriction analysis of positive clones revealed that one of these, EV1, did not contain EVX2 sequences. Subcloning and sequencing showed that it actually contained sequences different from EVX2 and closely related to the murine *Evx 1* reported sequences (21). Terminal sequences of the genomic DNA fragment present in EV1 were used to screen a cosmid library and several positive clones were found. Among these one, GV6, was found to have sequences in common with E143, the cosmid clone already known to contain the HOX1J homeobox (18). The overlapping of the three genomic clones was subsequently confirmed by sequencing and Southern analyis of human genomic DNA. The EVX1 homeobox lies about 48 kb upstream from the HOX1J homeobox but in an opposite orientation (Fig. 1A). The paralogous EVX2 genomic region has been shown to be particularly GC-rich (23). Fig. 1A shows the location of non-methylated Not I sites in the EVX1 genomic region analysed. These sites belong to CpG rich regions (29) consistently hupomethylated in several tissues (not shown). The entire EVX1 region appears to be unusually GC-rich.

The EVX1 sequence

We further studied the intron-exon organization of EVX1. Using the EVX1 homeobox as a probe we screened a cDNA library

```
GGCGGTGCGGCACAGAGGAGCGCTCGCTTCACAAGGTGACCCTAGCTCCCACCGCCACCGCCGCGGTCGCGCTCCAGACGCCCTCCAGCAGCTCCGCGCC    100
CTCCCAGGCACCCGGCCTTTCTTTCTCCCTCTTGCAACCAAGATCCGTCCGGCCGCTGGAGACCCAGGGAGCCGGGGTTAGGAACTCACTTGGGGCTTTC    200

                          MetGluSerArgLysAspMetValVALPheLeuAspGlyGlyGlnLeuGlyThrLeuValGlyLysArgValSer     25
CCCTCCCCCACCGGAGAGCCCCGGGATGGAGAGCCGAAAGGACATGGTTGTGTTTCTGGATGGGGGTCAGCTTGGCACTCTGGTTGGCAAGAGAGTCTCA    300

AsnLeuSerGluAlaValGLYSerProLeuProGluProProGluLysMetValProARGGlyCysLeuSerProArgAlaVALProProALATHRArgG      58
AATTTGTCCGAAGCCGTGGGCAGCCCGCTGCCGGAGCCGCCCGAGAAAATGGTGCCCCGTGGTTGCCTGAGCCCTCGGGCCGTCCCTCCGGCCACCCGGG   400

luArgGlyGlyGlyGlyPROGluGluGluProValAspGlyLeuAlaGlySerAlaAlaGlyPROGlyAlaGluProGLNVALAlaGlyAlaAlaMetLe     91
AGCGCGGCGGGGGAGGCCCGGAGGAGGAGCCGGTAGATGGACTCGCAGGCAGCGCGGCGGGGCCGGGCGCCGAGCCCCAGGTAGCTGGGGCGGCCATGCT   500

uGlyProGlyProProALAProSerVALAspSerLeuSerGlyGlnGlyGlnProSerSerSerAspThrGluSerAspPheTyrGluGluIleGluVal    125
CGGCCCAGGACCCCCGGCCCCCTCAGTCGACAGCCTCTCCGGACAGGGGCAACCCAGTAGCTCGGACACCGAGTCGGATTTCTATGAAGAAATCGAGGTG   600

SerCysThrProAspCysAlaThrGlyAsnAlaGluTyrGlnHisSerLys                                                     142
AGCTGCACCCCGGACTGCGCCACCGGGAACGCCGAGTACCAGCACAGCAAAGgtagccaccgtgcccctccgcct.......1500bp.......tgtgc   652

                     GLYSERGlySerGLUAlaLeuVALGLYSerProASNGLYGlySerGluTHRProLysSerAsnGlyGlySerGlyGlyGLYG    169
tctggacttcgctgtgcagGGTCCGGCTCCGAGGCGCTGGTCGGCAGTCCGAACGGAGGGAGCGAGACCCCCAAGAGCAACGGCGGCAGTGGTGGGGGCG   733

lySerGlnGlyThrLeuAlaCysSerAlaSerAspGlnMetArgArgTyrArgThrAlaPheThrArgGluGlnIleAlaArgLeuGluGluLysGluPheTy   202
GCTCGCAAGGCACCCTGGCGTGCAGCGCCAGTGACCAGATGCGTCGTTACCGCACCGCCTTCACCCGAGAGCAGATTGCGCGGCTGGAGAAGGAATTCTA   833

rArgGluAsnTyrValSerArgProArgArgCysGluLeuAlaAlaAlaAlaLeuAsnLeuProGluThrThrIleLys                         228
CCGGGAGAACTACGTATCCAGGCCGCGGAGATGTGAGCTGGCGGCCGCCCTAAACCTGCCGGAAACCACCATCAAGgtatgcggggtccaggctggggag   933

gcgggtgtgcacctatttagcgggaagtaaatgccaactgccaactccctgaaacgcaggccaggaatctgggcctggggtctccctgcccggcgtcgtg   1033
cagattgaccctcgtgacagctcctaggcagcattgctgccatgtggctgactctgtcccttccctggttgatagacaaggggtgggcgtggggaaggg   1133
gatagagtgcctgtgcgggtgacaaggaagtttctggggacacgctctctgcggccgcagaccaattgagtccatgtcctttcactgctcctcccataca   1233
cacatggcctctggcaccccoggggggcctggccacctgggcagaaggaacggagggagcgggctggagtcactccccaaacctctctcggagggattcca   1333
gctccaggtggtggtggtggggtcggtctctacccggctggtgtctgctttggctgttctgcctgcgaacactgtcccggagcgggaccagactactggc   1433

                                                  ValTrpPheGlnAsnArgArgMetLysAspLysArgGlnArgLeuA      243
ctctgagcatcgggccaagtccagctactgaacctgctccgctcctctccccagGTGTGGTTCCAGAACCGGCGCATGAAGGACAAGCGGCAGCGGCTGG   1533

laMetThrTrpProHisProAlaAspProAlaPheTyrThrTyrMetMetSerHisAlaAlaAlaAlaGlyGlyLeuProTyrProPheProSerHisLe    276
CCATGACGTGGCCGCACCCGGCGGACCCCGCCTTCTACACTTACATGATGAGCCATGCGGCGGCCGCGGGCGGCCTGCCCTACCCCTTCCCATCGCACCT   1633

uProLeuProTyrTyrSerProValGlyLeuGlyAlaAlaSerAlaAlaSerAlaAlaAlaSerProPheSerGlySERLeuArgProLeuAspThrPhe    310
GCCCCTGCCCTACTACTCGCCGGTGGGCCTGGGCGCCGCATCCGCCGCCTCCGCCGCCGCCTCGCCCTTCAGCGGCTCGCTGCGCCCGCTCGACACGTTC   1733

ArgValLeuSerGlnProTyrProArgProGluLeuLeuCysAlaPheArgHisProProLeuTyrProGlyProAlaHisGlyLeuGlyAlaSerAlaG    343
CGCGTGCTGTCGCAGCCCTACCCGCGGCCCGAACTGCTGTGCGCCTTCCGCCACCCGCCGCTCTACCCCGGGCCCGCGCACGGACTGGGCGCCTCTGCCG   1833

LYGlyProCysSerCysLeuAlaCysHisSerGlyProAlaAsnGlyLeuAlaProArgAlaAlaAlaAlaAlaSerAspPheThrCysAlaSerThrSerAr    376
GCGGCCCCTGCTCCTGCCTCGCCTGTCACAGCGGCCCGGCCAACGGGCTGGCGCCCGGGCTGCCGCCGCCTCGGACTTCACCTGTGCCTCCACCTCCCG   1933

gSerAspSerPheLeuThrPheAlaProSerValLeuSerLysAlaSerSerValAlaLeuAspGlnArgGluGluValProLeuThrArg *           407
CTCGGACTCCTTCCTCACCTTCGCGCCCTCGGTGCTCAGCAAGGCCTCCTCCGTCGCGCTGGACCAGAGGGAGGAGGTGCCCCTCACTAGATAAGGGGCC   2033

GGGCTGGTGGCTGCCGGCTCCATGACGCCCGTGGGGTCACCCCCCGGCCCGGGACTCAGCCAGCCTCGCCTCCTCGCTCCTCGCTCCTCGCCCCTAGGAC   2133

GCCAAGGGGGAAAGGAGAGGGCGGAAAAGGACCAGCGGGATCC                                                            2176
```
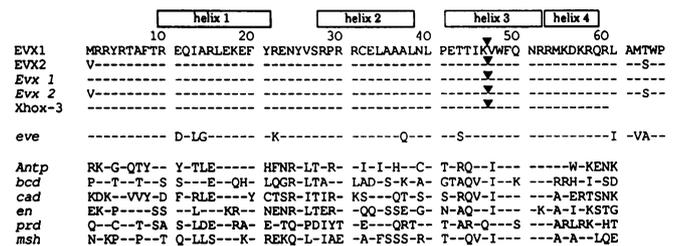
Figure 2. Nucleotide sequence and predicted protein sequence of EVX1, derived from the cDNA clone shown in Fig. 1B up to the Bam HI site downstream from the TAA stop codon. It includes the splice junction regions of the first intron and the entire second intron derived from the analysis of the corresponding genomic region. Small letters indicate intron sequences. The homeodomain is in boldface. Capital amino acid symbols indicate substitutions between EVX1 and the murine homolog, *Evx 1* (21). Three empty arrowheads point to the position where groups of 5, 3 and 1 alanine residues in a row are present in the *Evx 1* protein while missing in EVX1. Not I sites are underlined. The last two identify the 418 nucleotide fragment used in RNase protection experiments.

prepared from human fetal brain. One clone of 2.5 kb was selected for further analysis. This cDNA clone does not contain a poly-A tail and appears to include the entire coding region. Comparison of this cDNA to cloned genomic sequences revealed that the coding region is comprised of three exons (Fig. 1B). A first intron, approximately 1500 bp long, is located 120 bp upstream from the homeobox whereas a second intron of 578 bp occurs in the homeobox itself at the same relative position of EVX2 (23), *Evx 1* and *Evx 2* (21).

The nucleotide sequence of the EVX1 cDNA clone together with its conceptual translation is shown in Fig. 2. The longest open reading frame codes for a protein of 407 amino acid residues, 9 residues shorter than the murine homolog. These 9 residues appear to be all alanine residues present in *Evx 1* in 2 clusters of 5 and 3 residues in a row in addition to an isolated occurrence (21). The 23 amino acid substitutions between the 407 amino acid EVX1 protein and its murine homolog are also shown in Fig. 2.

The EVX1 homeodomain is clearly related to the fly *eve* homeodomain and to cognate sequences like the frog Xhox-3

```
            ┌──── helix 1 ───┐   ┌──── helix 2 ───┐   ┌──── helix 3 ───┐ ┌ helix 4 ┐
            10        20        30        40       ▼ 50        60
EVX1    MRRYRTAFTR EQIARLEKEF YRENYVSRPR RCELAAALNL PETTIKVWFQ NRRMKDKRQRL AMTWP
EVX2    V--------- ---------- ---------- ---------- -----▼---- ----------- --S--
Evx 1   ---------- ---------- ---------- ---------- -----▼---- ----------- -----
Evx 2   V--------- ---------- ---------- ---------- -----▼---- ----------- --S--
Xhox-3  ---------- ---------- ---------- ---------- -----▼---- -----------

eve     ---------- D-LG------ -K-------- ------Q--- --S------- ----------I -VA--

Antp    RK-G-QTY-- Y-TLE----- HFNR-LT-R- -I-I-H--C- T-RQ--I--- -----W-KENK
bcd     P--T--T--S S---E--QH- LQGR-LTA-- LAD-S-K-A- GTAQV-I--K ---RRH-I-SD
cad     KDK--VVY-D F-RLE----Y CTSR-ITIR- KS---QT-S- S-RQV-I--- ---A-ERTSNK
en      EK-P----SS --L---KR-- NENR-LTER- -QQ-SSE-G- N-AQ--I--- -K-A-I-KSTG
prd     Q--C--T-SA S-LDE--RA- E-TQ-PDIYT -E---QRT-- T-AR-Q---S ---ARLRK-HT
msh     N-KP--P--T Q-LLS---K- REKQ-L-IAE -A-FSSS-R- T--QV-I--- ---A-A--LQE
```

Figure 3. Alignment of the EVX1 homeodomain with homeodomains of various classes. The one-letter amino acid code is used. Dashes indicate amino acid identity. Arrowheads point to the splice site. Sequences are from Scott *et al.* (7) but for *Evx 1* and *Evx 2* (21) and EVX1 (23). The μ-helix domains (6) are indicated.

homeodomain (30), the murine *Evx 1* and *Evx 2* (21) and the human EVX2 homeodomains (23) (Fig. 3). EVX1 and *Evx 1* homeodomains are identical, both differing from the *Evx 1* and EVX1 homeodomains at the first position (Fig. 3). All four
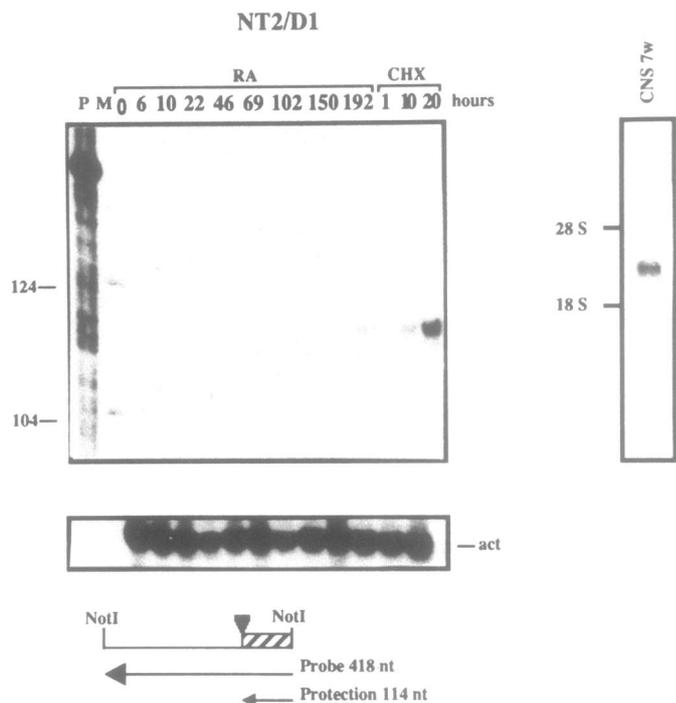
Figure 4. EVX1 expression in embryonal carcinoma cells and embryonic tissue. (Left) RNase protection analysis of the indicated Not I EVX1 probe in total RNA (15 µg) from NT2/D1 cells, untreated (0) and treated with 10µM retinoic acid (RA) for various periods up to 192 hours or treated with 20 µg/ml cycloheximide (CHX) for 1, 10 and 20 hours. P indicates the RNA probe and M the size marker. Band sizes are shown in nucleotides (nt). (Right) Northern blot analysis of polyadenylated RNA (15 µg) extracted form the central nervous system (CNS) of human 7-week embryos.

homeoboxes are interrupted by an intron within the recognition helix of the encoded homeodomain between amino acid positions 46 and 47 as is the case for Xhox-3 (Fig. 3).

Southern blot analysis of human DNA using EVX1 and EVX2 sequences does not seem to suggest the presence of additional related sequences. Hybridisation under relaxed conditions consistently detects only the restriction bands corresponding to these two genes (not shown). Nonetheless, it seems interesting to look for related EVX genes in order to accertain, in particular, whether they map at other HOX loci.

Expression of EVX1 sequences

Evx 1 is expressed in the mouse teratocarcinoma F9 cell system, which is believed to represent a good model system for early embryonic development (21). An *Evx 1* transcript was detected in undifferentiated F9 stem cells and disappears upon differentiation into parietal endoderm cells after administration of retinoic acid. We analysed EVX1 expression in human teratocarcinoma cells of line N-TERA2, clone D1 (NT2/D1)(25,26,31). Northern blot analysis failed to detect EVX1 expression in NT2/D1 cells whether undifferentiated or cultured for 14 days in 10 µM retinoic acid (data not shown). Hence, EVX1 appears not to be regulated in these cells like *Evx 1* in F9 cells. On the other hand, *Evx 1* is not expressed in P19 stem cells (21), suggesting that its expression could be confined to certain lines of teratocarcinoma cells.

We further investigated EVX1 expression in NT2/D1 cells using the RNase protection technique (Fig. 4). EVX1 is weakly

espressed and properly spliced in differentiated NT2/D1 cells after several days of retinoic acid treatment. This is particularly evident in cells induced to differentiate for 192 hours, *i.e.* 8 days. Several different cell populations, including neurons, are present at this period (25) and some cells of this diverse population might express EVX1 as is the case for EVX2 (23). The 418-nucleotide Not I probe we used for RNase protection analysis contains 114 nucleotides of the exon including the 3' half of the homeobox and some intronic sequences laying upstream. Protection signals correspond to the predicted splice site, showing that EVX1 trascripts are properly spliced in these cells. EVX1 expression in NT2/D1 cells is dramatically increased by cycloheximide (CHX) treatment (Fig. 4).

The murine homolog, *Evx 1*, is expressed during early embryogenesis of the mouse in a biphasic manner (21). From day 7 to 9 of development its expression emerges at the posterior end of the embryo within the primitive ectoderm, and later in the mesoderm and neuroectoderm. From day 10 to 12.5, *Evx 1* transcripts are restrcted to specific cells within the neural tube and hindbrain along their entire lenghts, while no expression is detectable in a variety of adult tissues.

We analysed EVX1 expression in human 7-week embryos by Northern blot hybridization of polyadenylated RNA. A 3 kb transcript is detectable in the central nervous system (Fig. 4) whereas we failed to detect expression in other embryonic tissues and organs (not shown). This observation is in agreement with the expression data in the mouse where Northern blot analysis detects extremely weak signals and in situ hybridizations reveal restricted expression domains (21)

CONCLUSIONS

We previously reported that EVX2, a human *eve*-related homeobox gene, is localized at the 5' end of the HOX4 locus on chromosome 2 (23). We now show that EVX1, a second *eve*-related human gene, is localized at the 5' end of the HOX1 locus on chromosome 7. Both genes are transcribed in an opposite orientation as compared to that of linked HOX genes.

Mammalian HOX loci arose as duplications of an ancestral complex locus and have been shown to be true homologs of the *Drosophila* homeotic gene complexes (32). Some of the various paralogy groups were already distinct when lineages leading to insects and vertebrates diverged even if the evolutionary position of 5' genes of the first four groups is still unclear (18). In fact, it is not known whether they were already present in the ancestral locus predating the divergence between insects and vertebrates or, alternatively, they have arisen specifically in the lineage leading to vertebrates. Isolation and mapping of EVX genes poses new questions. *Eve* is not linked to *Drosophila* homeotic genes located at the BX-C or ANT-C, whereas EVX1 maps at the HOX1 locus and EVX2 maps at the HOX4 locus. This observation suggests the occurrence of large-scale duplication of clusters of highly conserved genes before the divergence of insects and vertebrates. An obvious corollary might be that the ancestral cluster included more developmental genes than previously believed.

Mammalian genes located to the 3' end of HOX loci and their fly homologs are expressed in anterior regions whereas genes located upstream are expressed in more and more posterior regions. *Evx 1* is expressed in mouse embryos according to an early pattern (21) compatible with a role in specifying positional information along the embryonic axis similar to that played in

frogs by Xhox-3 (30,33). *Evx 1* expression originates at the posterior end of the gastrulating embryo in the region of the posterior to mid primitive streak in both ectoderm and mesoderm. Over the following stages the domain of expression expands in the mesoderm and to the overlying ectoderm but remains restricted to the posterior portion of the fetus (21). Such an expression pattern is compatible with a localization of *Evx 1* and EVX1 at the 5′ end of a HOX locus. The significance of this conserved genomic organization remains unclear especially because EVX genes and HOX genes appear to be transcribed in opposite orientation. The presence of CpG rich regions between EVX1 and HOX1J (Fig. 1A) and between EVX2 and HOX4I (23) suggests separated, though not necessarily independent, regulatory mechanisms for these transcription units. Further studies in different species are clearly required to assess the evolutionary position of *eve*-related genes with regard to HOX gene complexes.

ACKNOWLEDGMENTS

REFERENCES

1. Akam, M. (1987). Development *101*, 1–22.
2. Ingham, P.W. (1988). Nature *335*, 25–34.
3. Dressler, G. and Gruss, P. (1988). TIG *4*, 214–219.
4. Gehring, W.J. (1987). Science *236*, 1245–1252.
5. Levine, M. and Hoey, T. (1988). Cell *55*, 537–540.
6. Qian, Y.Q., Billeter, M., Otting, G., Mueller, M., Gehring, W.J. and Wuethrich, K. (1989). Cell *59*, 573–580.
7. Scott, M.P., Tamkun, J.W. and Hartzell, G.W. III (1989). BBA Rev. Cancer *989*, 25–48.
8. Robert, W.M., Sassoon, D., Jacq, B., Gehring, W. and Buckingham, M. (1989). EMBO J. *8*, 91–100.
9. Hoey, T. and Levine, M. (1988). Nature *332*, 858–861.
10. Herr, W., Strum, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Binney, M., Ruvkun, G. and Horvitz, R.L. (1988). Genes Dev. *2*, 1512–1515.
11. Frain M., Swart, G., Monaci, P., Nicosia A., Staempfli S., Frank, R. and Cortese R. (1989). Cell *59*, 145–157.
12. Okamoto, K., Okazawa, H., Okuda, K., Sakai, M., Muramatsu, M. and Hamada, H. (1990). Cell *60*, 461–472.
13. Schoeler, H.R., Ruppert S., Suzuki, N., Chowdhury K. & Gruss P. (1990). Nature *344*, 435–439.
14. Gehring, W.J. and Hiromi, Y. (1986). Annu. Rev. Genet. *20*, 147–173.
15. Duboule, D. and Dollè, P. (1989). EMBO J. *8*, 1497–1509.
16. Graham, A., Papalopulu, N. and Krumlauf, R. (1989). Cell *57*, 367–378.
17. Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., Faiella, A. and Simeone, A. (1988). Hum. Reprod. 3, 880–886.
18. Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A. and Boncinelli, E. (1989). Nucl. Acids Res. *17*, 10385–10402.
19. Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F. and Boncinelli, E. (1991). Mechanisms. Develop. *33*, 215–228.
20. Stornaiuolo, A., Acampora, D., Pannese, M., D'Esposito, M., Morelli, F., Migliaccio, E., Rambaldi, M., Faiella, A., Nigro, V., Simeone, A. and Boncinelli, E. (1990). Cell Differentiation and Development *31*, 119–127.
21. Bastian, H. and Gruss, P. (1990). EMBO J. *9*, 1839–1852.
22. Macdonald, P.M., Ingham, P., Struhl, G. (1986). Cell *47*, 721–734.
23. D'Esposito, M., Morelli, F., Acampora, D., Migliaccio, E., Simeone, A. and Boncinelli, E. (1991). Genomics *10*, 43–50.
24. Sanger, R., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA *74*, 5463–5467.
25. Andrews, P.W., Damjanov, I., Simon, D., Banting, G.S., Carlin, C., Dracopoli, N.C. and Fogh, J. (1984). Lab. Invest. *50*, 147–162.
26. Mavilio, F., Simeone, A., Boncinelli, E. and Andrews, P.W. (1988). Differentiation *37*, 73–79.
27. Chirgwin, J.M., Przybyla, A.E., Mac Donald, R.J. and Rutter, W.J. (1979). Biochemistry *18*, 5294–5299.
28. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). Nucl. Acids Res: *12*, 7035–7056.
29. Bird, A.P. (1986). Nature *321*, 209–213.
30. Ruiz i Altaba, A. and Melton, D.A. (1989). Development *106*, 173–183.
31. Simeone, A., Acampora, D., Arcioni, L., Andrews, P.W., Boncinelli E. and Mavilio F. (1990). Nature *346*, 763–766.
32. Akam, M. (1989). Cell *57*, 347–349.
33. Ruiz i Altaba, A. and Melton, D.A. (1989). Nature *341*, 33–38.