Development 99, 603–617 (1987)
Printed in Great Britain © The Company of Biologists Limited 1987

Developmental and spatial patterns of expression of the mouse homeobox gene, *Hox2.1*

ROBB KRUMLAUF*, PETER W. H. HOLLAND, JOHN H. McVEY and BRIGID L. M. HOGAN

Laboratory of Molecular Embryology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

* To whom reprint requests should be sent

Summary

The *Hox2.1* gene forms part of a cluster of homeobox-containing genes on mouse chromosome 11. Analysis of Hox2.1 cDNAs isolated from an 8½-day *p.c.* mouse embryo library predicts that the gene encodes a 269 amino acid protein (M_r, 29432). This deduced protein contains a homeobox 15 amino acids from the carboxy terminus and is very rich in serine and proline. A second partially conserved region present in several other genes containing homeoboxes, the hexapeptide Ile-Phe-Pro-Trp-Met-Arg, is located 12 amino acids upstream of the homeodomain and is encoded by a separate exon. Analysis of *Hox2.1* gene expression reveals a complex and tissue-specific series of RNA transcripts in a broad range of fetal tissues (lung, spinal cord, kidney, gut, spleen, liver and visceral yolk sac). Comparison of the temporal patterns of gene expression during development and in the adult suggests that *Hox2.1* is regulated independently in different tissues. Evidence is also presented that transcripts from other loci have extensive homology to the *Hox2.1* gene in sequences outside of the homeobox. *In situ* hybridization shows that *Hox2.1* transcripts are regionally localized in the spinal cord in an apparent anterior–posterior gradient extending from the hind brain. The distribution of RNA also displays a cell-type specificity in the lung, where mesodermal cells surrounding the branching epithelial cell layer accumulate high levels of *Hox2.1* transcripts.

Key words: mouse embryo, homeobox gene, *Hox2.1*.

Introduction

Current molecular approaches towards understanding of the control of mammalian development are attempting to expand upon classic embryological or genetic studies. However, the cloning and characterization of genes affecting development has proved difficult because of the lack of closely linked markers and the large chromosomal regions involved. In the mouse, insertional mutagenesis *via* retroviral infection of embryos (Jaenisch, Harbers, Schnieke, Lohler, Chumakov, Jahner, Grotkopp & Hoffman, 1983) or microinjection of fertilized eggs (Woychick, Sterwart, Davis, D'Eustachio & Leder, 1985) represent one way of homing in on genes that influence developmental processes. An alternative approach is based on the recent discovery that highly conserved sequences in genes implicated in the control of pattern formation during the embryonic development of *Drosophila* (McGinnis, Levine, Hafen, Kuroiwa & Gehring, 1984*a*; Scott & Weiner, 1984) have a broad phylogenetic distribution (McGinnis, Garber, Wirz, Kuroiwa & Gehring, 1984*b*; McGinnis, 1985; Holland & Hogan, 1986). Genetic analysis in *Drosophila* has characterized two complexes, Antennapedia (ANT-C) (Kaufman, Lewis & Wakimoto, 1980) and Bithorax (BX-C) (Lewis, 1978), which contain genes involved in control of the segmental body plan. Cloning of these complexes revealed eight different genes which contain a common DNA element, termed the 'homeobox' (Garber, Kuroiwa & Gehring, 1984; McGinnis *et al.* 1984*b*; Scott & Weiner, 1984; Regulski, Harding, Kostriken, Karch, Levine & McGinnis, 1985; Laughon *et al.* 1985). These sequences share greater than 80 % homology and are classified as *Antennepedia*-like or Class I homeoboxes. The homeobox itself is a 183bp sequence capable of coding for a highly conserved protein domain (for review see Gehring, 1985) which has sequence similarity to the DNA-binding domains of

yeast and bacterial regulatory proteins (Shepherd, McGinnis, Carrasco, De Robertis & Gehring, 1984; Laughon & Scott, 1984; Johnson & Herskowitz, 1985; Whiteway & Szostak, 1985). The *Drosophila* proteins containing homeoboxes have been localized in the nucleus (White & Wilcox, 1984; Beachy, Halfand & Hogness, 1985; Carroll & Scott, 1985; Carroll, Laymon, McCutcheon, Riley & Scott, 1986) and one gene product, *engrailed*, exhibits a sequence-specific DNA-binding activity (Desplan, Theis & O'Farrell, 1985). The *Drosophila* homeobox probes have provided a means of directly isolating a family of genes that might regulate early stages of mammalian embryogenesis. It is therefore important to determine the properties of the homeobox-containing genes in mammals in order to establish whether these genes are involved in developmental control processes.

Genes containing homeoboxes have been isolated from sea urchins (Dolecki, Wannakrairoj, Lum, Wang, Riley, Carlos, Wang & Humphreys, 1986), frogs (Carrasco, McGinnis, Gehring & De Robertis, 1984; Muller, Carrasco & De Robertis, 1984; Harvey, Tabin & Melton, 1986), mice (Colberg-Poley, Voss, Chowdhury & Gruss, 1985*a*; McGinnis, Hart, Gehring & Ruddle, 1984*c*; Jackson, Schofield & Hogan, 1985; Joyner, Kornberg, Coleman, Cox & Martin, 1985; Hart, Awgulewitsch, Fainsod, McGinnis & Ruddle, 1985; Rabin, Hart, Ferguson-Smith, McGinnis, Levine & Ruddle, 1985; Hauser, Joyner, Klein, Learned, Martin & Tjian, 1985; Breier, Bucan, Francke, Colberg-Poley & Gruss, 1986; Awgulewitsch, Utset, Hart, McGinnis & Ruddle, 1986; Dubuole, Baron, Mahl & Galliot, 1986) and humans (Levine, Rubin & Tjian, 1984; Hauser *et al.* 1985; Boncinelli, Simeone, La Volpe, Faiella, Fidanza, Acampora & Scotto, 1985), with the genes often organized in clusters. Initial experiments on the patterns of gene expression supported a potential role in development. Mammalian homeobox genes are expressed in teratocarcinoma cells (Colberg-Poley *et al.* 1985*a*; Hauser *et al.* 1985), early embryos (Jackson *et al.* 1985; Colberg-Poley *et al.* 1985*a*,*b*; Hart *et al.* 1985; Hauser *et al.* 1985) and some adult stages (Jackson *et al.* 1985; Colberg-Poley *et al.* 1985*b*; Wolgemuth, Engelmyer, Duggal, Gizang-Ginsberg, Mutter, Ponzetto, Viviano & Zakeri, 1986), in a tissue-specific (Jackson *et al.* 1985; Simeone, Mavilio, Bottero, Giampaolo, Russo, Faiella, Boncinelli & Peschle, 1986) and spatially restricted (Awgulewitsch *et al.* 1986) manner. However, as yet there is no direct evidence that establishes the function or regulatory role of these genes in mammals. As a step towards this goal we have isolated genomic and cDNA clones for a mouse homeobox gene, *Hox2.1*, to analyse potential gene products and their patterns of expression. In this study we report the complete protein sequence predicted from the cDNA clones for the *Hox2.1* gene and describe several interesting features of its structure. A second region of homology upstream of the homeobox was found in homeobox-containing genes of several species. We also describe a detailed analysis of the tissue-specific and temporal expression of this gene *via* Northern analysis, which reveals that it is differentially regulated during development and in the adult. The presence of multiple RNA species suggests that there is a complex transcription pattern for the gene which could involve differential processing and multiple gene products. *In situ* hybridization experiments examine the spatial distribution and cell-type specificity of *Hox2.1* in 12½-day mouse embryos.

Materials and methods

Clone isolation and sequencing

A cDNA library, prepared from poly(A)$^+$RNA of C57BL 8½-day *p.c.* embryos, by inserting double-stranded cDNA's into the λgt10 vector using oligo adapters (Farhner, Hogan & Flavell, 1987), was screened using a labelled 600bp *Eco*RI-*Pvu*II fragment (H 24.1 probe A, Jackson *et al.* 1985) which contains the Hox2.1 homeobox and 3′ region of the gene. Nylon filters (Pall Biodyne) of phage were prehybridized and hybridized in 50% formamide, 5 × SSC, 0·1% bovine serum albumin, 0·1% Ficoll 400, 0·1% polyvinylpyrrolidone, 50 mM-sodium phosphate (pH 6·8), 0·1% SDS, and 200 μg ml^{-1} denatured sonicated salmon sperm DNA at 42°C. Filters were washed in 0·1 × SSC, 0·1% SDS at 65°C. Positively hybridizing clones were plaque-purified and the inserts subcloned as *Eco*RI fragments into pGEM1 or pGEM2 (Promega Biotec). Probes from these clones were then used to rescreen the library for additional cDNAs to maximize the area cloned. Both strands of the recombinant clones were then sequenced in the GEMINI vectors by the procedures of Maxam & Gilbert (1980) or using primers for the T7 and SP6 promoters (Promega Biotec), coupled with chain termination sequencing methods (Sanger in full, 1977).

RNA isolation

Tissues were dissected from CBA or C57Bl/10 embryos in Dulbecco's modified Eagle's medium with 10% fetal bovine serum (Hogan, Costantini & Lacy, 1986) or taken directly from adult animals and rapidly frozen in liquid nitrogen. The RNA was isolated from the frozen tissue by a modification of the method of Auffray & Rougenon (1979). Briefly, the tissue was homogenized in 5–10 ml of 3 M-LiCl, 6 M-urea per gram of tissue for 2 min on ice, using an ultraturrax or similar tissue disrupter. The homogenate was stored overnight at 4°C, then centrifuged at ×5000 *g* for 10 min at 0°C and the supernatant poured off. A half volume of cold 3 M-LiCl, 6 M-urea was added, the sample vortexed and recentrifuged, discarding the supernatant.

The pellet was redissolved in 10 mM-Tris–HCl (pH 7·6), 1 mM-EDTA, 0·5 % SDS using 5 ml per gram of original tissue. The sample was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (24:24:1) and the aqueous phase separated by centrifugation at 5000 *g* for 5 min. The RNA was harvested by ethanol precipitation and poly(A)$^+$RNA isolated by oligo dT cellulose chromatography (Aviv & Leder, 1972). The A_{260} and a ^{3}H-poly-U binding assay (Bishop, Rosbash & Evans, 1974; Rosbash & Ford, 1974) were used to quantify the amount of poly(A)$^+$ mRNA.

Northern blots

Poly(A)$^+$RNA, quantified by the poly-U assay, were electrophoresed for 6 h at 60 mV in a 6 % formaldehyde–1 % agarose gel in 1 × MOPS buffer (pH 7·0; 20 mM-morpholine propanesulphonic acid, 5 mM-sodium acetate, 1 mM EDTA) after denaturation at 60 °C for 10 min in 70 % formamide–6 % formaldehyde-1 × MOPS. Following electrophoresis, a marker poly(A)$^-$ lane was cut from the gel and stained to provide size markers (28 S–4·7 kb, 18 S–2·2 kb), the remainder of the gel was soaked successively in 50 mM-NaOH, 100 mM-NaCl (20 min); 100 mM-Tris pH 7·6 (20 min) and 2 × SSC (20 min), then transferred to a Genescreen (Dupont) membrane in 20 × SSC overnight. The filter was rinsed in 6 × SSC, exposed to 600 μwatts cm^{-2} (254 nm u.v. light source) for 5 min to crosslink the RNA to the filter, and then baked for 2 h at 80 °C. Synthesis of anti-sense SP6 and T7 RNA probes using ^{32}P-UTP (600 Ci mmole^{-1}; NEN-Dupont) was performed according to polymerase suppliers (Promega Biotec) and as previously described (Melton, Krieg, Rebagliati, Maniatis, Zinn & Green, 1984; Krumlauf, Hammer, Tilghman & Brinster, 1985). The Hox2.1 probe was constructed by cloning the *Bam*HI-*Hin*dIII fragment of cDNA 2.1A (Fig. A probe 2) into pGEM2 and linearizing the plasmid with *Eco*RI for T7 transcription. Filters were hybridized with 10^7 cts min^{-1} ml^{-1} in 60 % formamide, 5 × SSC, 0·1 % bovine serum albumin, 0·1 % Ficoll 400, 0·1 % polyvinylpyrolodone, 20 mM-sodium phosphate pH 6·8, 1 % SDS, 7 % dextran sulphate, 100 μg ml^{-1} denatured sonicated salmon sperm DNA, 100 μg ml^{-1} tRNA (bakers yeast), 10 μg ml^{-1} poly A at 65 °C for 16–24 h. Filters were washed in 2 × SSC, 1 % SDS at 50 °C for 3×15 min, then in 0·2 × SSC, 1 % SDS at 80 °C for 1–2 h, and exposed to Kodak XAR-5 film with intensifying screens (lightning plus, Dupont) at −70 °C for various times. In some cases the filters were given a further wash with RNase A to remove unmatched hybrids. The membrane was soaked in 2 × SSC plus 20 μg ml^{-1} RNase A (Sigma) at room temperature for 30 min, then washed in 2 × SSC, 0·5 % SDS at 50 °C for 2×30 min. Filters that had been stripped of Hox2.1 probe by washing at 70 °C with 75 % formamide, 0·1 % SDS, or duplicate filters were hybridized with an SP6 anti-sense mouse β-actin RNA probe as a control to test the loading and quality of RNA on the filters. Relative levels of RNA were determined by scanning various exposures of the autoradiographs with a densitometer, using the RNA from 12½-day embryos as the standard.

In situ *hybridization*

Embryos from CBA mice, 12½ days *p.c.*, were fixed with paraformaldehyde, embedded, sectioned using a cryostat and hybridized as previously described (Hogan *et al.* 1986). The Hox2.1 anti-sense (−) and sense (+) strand probes were synthesized from the *Eco*RI fragment of cDNA 2.1A (Fig. 1, probe 1), after subcloning into the vector pGEM1 in both orientations. The vectors were linearized with *Sal*I and labelled with ^{35}S-UTP (NEN-Dupont) to a specific activity of 2×10^9 dissints min^{-1} μg^{-1} using T7 RNA polymerase, as described by Melton *et al.* (1984). The probes were hydrolysed to an average size of 100 nt in a controlled

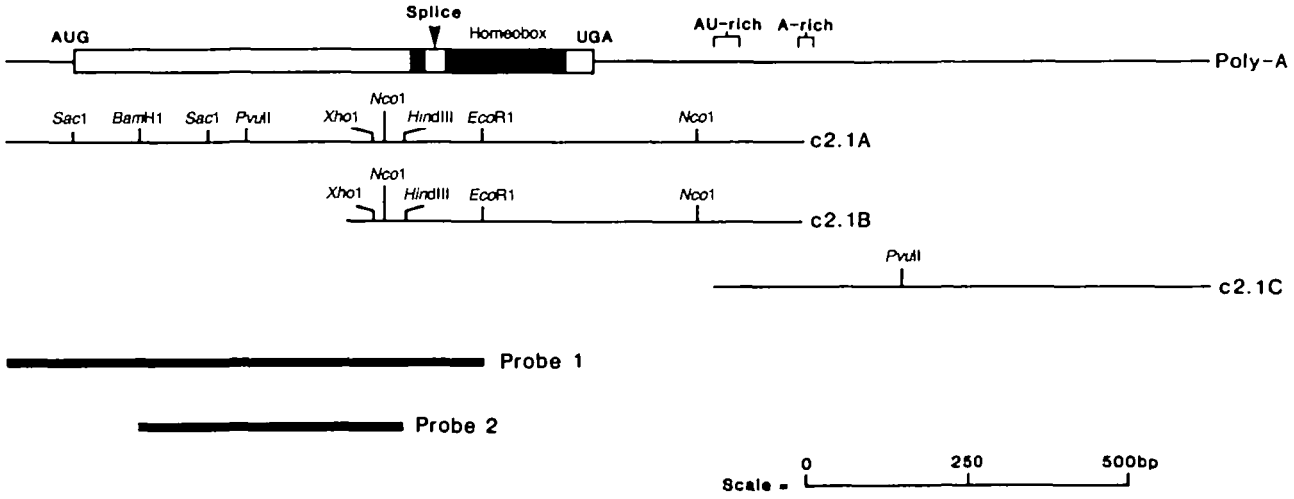


Fig. 1. Structure of Hox2.1 cDNA clones and location of predicted protein coding region. The thin lines labelled c2.1A, c2.1B and c2.1C represent the overlapping cDNA clones isolated from the 8½-day *p.c.* cDNA library and the thick lines denote subcloned regions of the c2.1A cDNA used for *in situ* hybridization, probe 1 (*Eco*RI fragment), or Northern analysis, probe 2 (*Bam*HI-*Hin*dIII fragment). The boxed region on the top line shows the open reading frame in the cDNA which corresponds to the 269 amino acid protein predicted from the sequence in Fig. 2. The large shaded area is the homeodomain and the small shaded area is the conserved hexapeptide region (see text).

reaction and used at a final concentration of 0·035–0·075 ng μl^{-1}. Post-hybridization RNase A washes and autoradiography are according to Hogan *et al.* (1986). Exposure times were 6 to 9 days.

Results

Isolation of cDNA clones

A mouse homeobox genomic clone isolated by Jackson *et al.* (1985), originally termed H24.1, is expressed in mouse embryos and is located on chromosome 11 band 11D (Munke, Cox, Jackson, Hogan & Francke, 1986). Based on restriction site mapping and sequence comparisons it is identical to the *Hox2.1* (Hart *et al.* 1985) or *Mu-1* gene (Hauser *et al.* 1985), which forms part of a complex of at least four and possibly six or more mouse homeobox-containing genes on chromosome 11 (Hart *et al.* 1985; Krumlauf, unpublished data). To obtain cDNA clones of *Hox2.1* for expression and protein analysis, we screened a cDNA library prepared from 8½-day *p.c.* mouse embryo RNA in the vector λgt10 (Farhner *et al.* 1987) with a subcloned Hox2.1 genomic probe (H24.1 probe A, Jackson *et al.* 1985) at high stringency. Sequences from the first positive clones were used to rescreen the library for additional clones. The structure and sequence of three overlapping cDNA clones (2.1A, B and C) spanning the largest area, is shown in Figs 1A, 2. We have confirmed that these cDNA clones are derived from the *Hox2.1* gene, by matching them exactly to sequences within genomic clones isolated from this gene (data not shown).

Northern analysis of RNA from 12½-day mouse embryos showed a major Hox2.1 transcript of 2·2 kb (see below). The three cDNAs (Fig. 2) span 1·9 kb of sequence which suggests that approximately 300 bp of the mRNA are missing from our cDNA sequence. The 3′ end of the mRNA is contained in clone 2.1C, including the polyadenylation signal (AATAAA) and the first few bases of the poly(A) tail. Therefore, the 300 bp, which are predicted to be missing, probably represent some sequences from the 5′ end of the RNA and the length of the poly(A) tail. Some of the cDNA clones terminated in an A-rich region between 1236–1251 (Fig. 2). However, there is no adjacent polyadenylation signal and these clones probably represent internal priming from the run of A's rather than an alternative site of poly(A) addition.

The predicted Hox2.1 protein

The protein predicted by the cDNA sequence is shown in Fig. 2. The longest open reading frame in the sequence is in-frame with the homeodomain, and we have denoted the first AUG, 72 bp after an in-frame stop codon, the methionine-initiation codon. The predicted protein contains 269 amino acids, and has a calculated $M_r = 29\,432$. The 61 amino acids of the homeobox and the following 15 amino acids up to the carboxy terminus are identical to those reported for the mouse (Hauser *et al.* 1985; Jackson *et al.* 1985) and human (Boncinelli *et al.* 1985; Hauser *et al.* 1985; Simeone, Mavilio, Bottero, Grampaolo, Russo, Faiella, Boncinelli & Peschle, 1986) *Hox2.1* gene. The protein-coding information is located in the 5′ portion of the sequence and there is a long 995 bp 3′ untranslated region. The human Hox2.1 cDNA (c10; Boncinelli *et al.* 1985) also has a long 3′ untranslated region (962 bp) and comparison with the mouse sequence shows that these regions are also remarkably conserved (84 %).

Analysis of the predicted Hox2.1 amino acid sequence using the secondary structure program of Chou & Fasman (1981) shows that the homeodomain contains alternating areas of α-helix, similar to the helix-turn-helix motif common to homeobox genes in the *Drosophila* Antennapedia class (Laughon & Scott, 1984). These regions are believed to bind and interact with DNA based on their homology to yeast mating-type genes, such as *mat a2*, which have been shown to encode DNA-binding proteins (Johnson & Herskowitz, 1985). The protein sequences outside of the homeobox are unusual in that they are very rich in serine (22 %) and proline (9 %). Prediction of secondary structure for this region of the protein shows that no α-helices can be formed as a consequence of the widely distributed helix-incompatible residues, such as proline and glycine. It is interesting to note that the *Xenopus* Xhox1-A (Harvey, Tabin & Melton, 1986), *Drosophila fushi tarazu* (Laughon & Scott, 1984) and *Antennapedia* (Schneuwly, Kuroiwa, Baumgartner & Gehring, 1986) proteins also contain approximately 10 % prolines outside of the homeobox. This may therefore represent a general feature of homeobox proteins.

A comparison of the Hox2.1 protein with other homeobox-containing proteins reveals that there is a second region of homology located at a short but variable distance upstream of the homeobox. The hexapeptide (Ile-Phe-Pro-Trp-Met-Arg) starting 18 amino acids upstream of the Hox2.1 homeobox (Fig. 2) is partially conserved in several homeobox proteins from other species (Fig. 3). The mouse Hox1.1 (P. Gruss, personal communication) and Hox1.3 (W. Odenwald, personal communication) proteins have peptides which differ from the 2.1 sequence by only one amino acid, and the *Xenopus* Xhox1-A (Harvey *et al.* 1985), *Drosophila Antp* (Schneuwly *et al.* 1986) and *Dfd* (Laughon *et al.* 1985) proteins have peptides which differ at additional positions. The *caudal* (cad) (Mlodzik, Fjose & Gehring, 1985) protein appears to be the most diverged member of the family, whilst the *fushi tarazu*

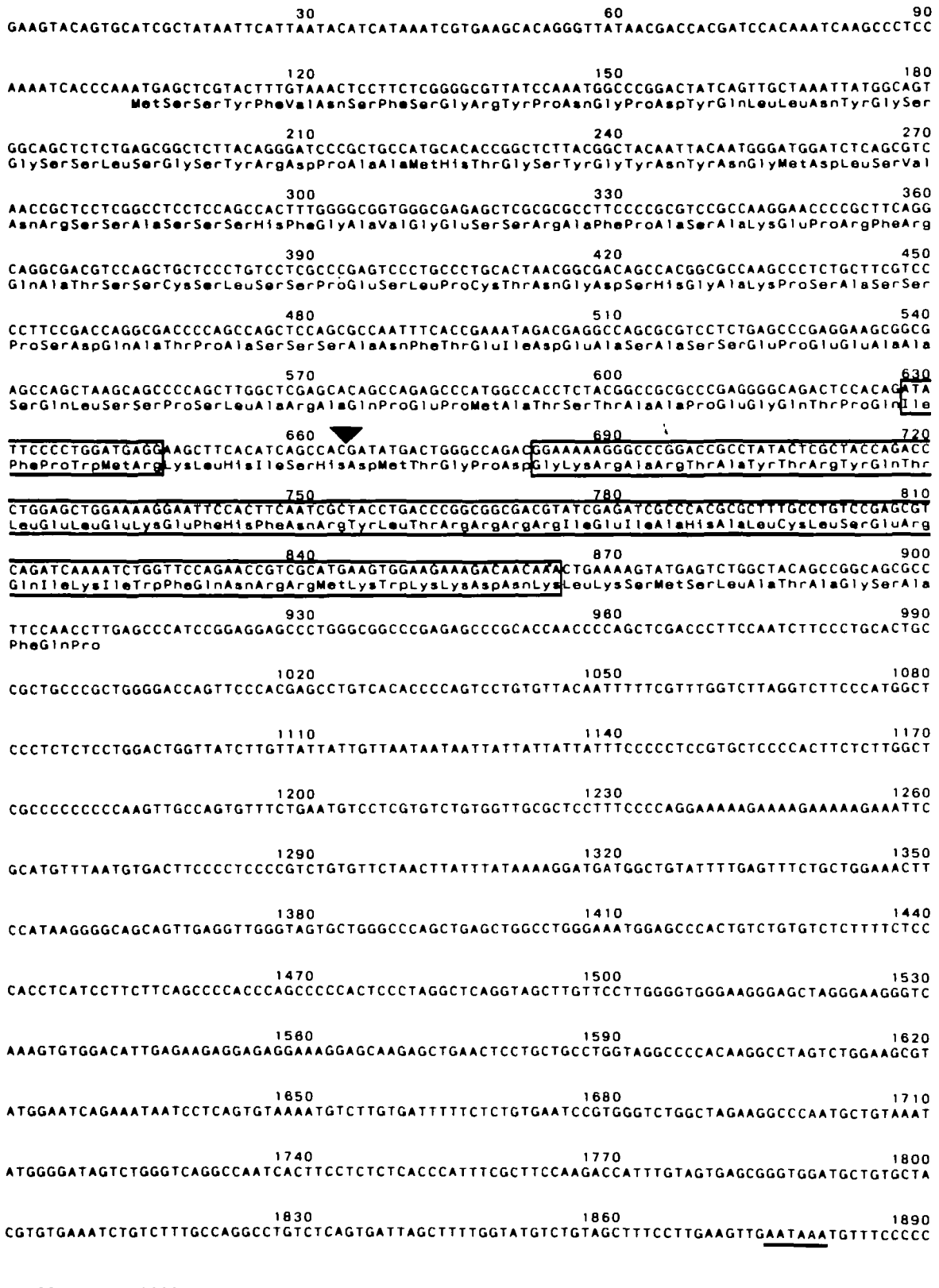


Fig. 2. The nucleotide sequence of Hox2.1 cDNAs and the predicted amino acid sequence of the protein. The nucleotide sequence is a composite produced from the sequence of the three cDNA clones (c2.1A, B and C) in Fig. 1. The numbering starts from the most 5′ base in clone c2.1A. The polyadenylation signal (AATAAA) is underlined. The boxed regions indicated the location of the homeodomain (large box, 682–864 nt) and the conserved peptide (small box, 628–645 nt). The solid triangle denotes the location of an intron and splice site, deduced by sequence comparison of cDNA and genomic clones.

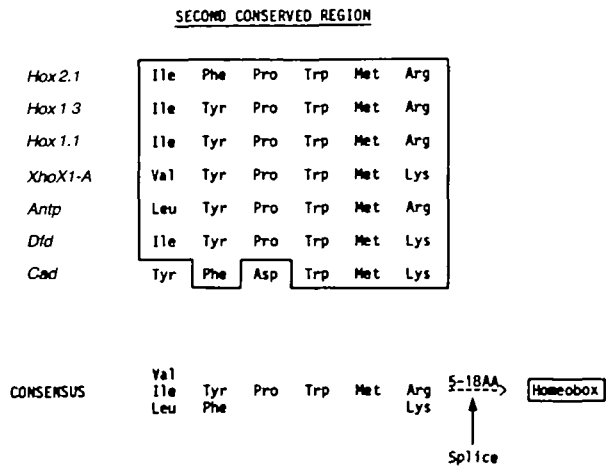


Fig. 3. Conserved hexapeptide region in homeobox genes of different species. The *Hox2.1*, *Hox1.1* and *Hox1.3* genes are from mouse, *Xhox1-A* from *Xenopus*, and *Antp*, *Dfd* and *Cad* from *Drosophila*. A consensus sequence is indicated below the boxed conserved amino acids. The 5–18 amino acids above the dotted arrow indicate the distance upstream from the homeobox and the arrow below indicates that the region is located on a different exon than the homeobox. References for sequences: (*Hox1.1* (Gruss, personal communication), *Hox1.3* (Odenwald, personal communication), *Xhox1-A* (Harvey *et al.* 1986), *Antp* (Schenuwly *et al.* 1986), *Dfd* (Laughon *et al.* 1985), *Cad* (Mlodzik *et al.* 1985).

(Laughon & Scott, 1984) and *engrailed* (Poole, Kauvar, Drees & Kornberg, 1985) proteins do not have a recognizably homologous peptide. These differences are summarized in Fig. 3 and a consensus sequence for this peptide region presented. In all of the examples shown in Fig. 3, the conserved peptide is located between 5 and 16 amino acids upstream of the homeobox domain and is encoded by a different exon to the homeobox domain. The degree of homology and location of this peptide suggest that it represents a second conserved region associated with the homeodomain and it will be important to establish whether this region functionally interacts with the homeobox domain.

Developmental expression

Jackson *et al.* (1985) have shown, using RNase protection experiments, that the *Hox2.1* gene is active in mouse embryos as early as 7½ days *p.c.* and transcripts were enriched in fetal spinal cord and adult kidney. However, these experiments gave no indication of either transcript complexity or temporal patterns of gene expression. Therefore we have examined the pattern and timing of *Hox2.1* expression by Northern blot analysis of RNA extracted from a series of embryonic and adult tissues. For this analysis we subcloned a fragment (*Bam*HI-*Hin*dIII; Fig. 1, probe B) of the 2.1A cDNA clone, which does not contain homeobox sequences, into the vector pGEM1 to provide single-stranded anti-sense RNA probes. Using this probe, a complex pattern of transcripts is observed in both mouse embryo and adult kidney poly(A)$^+$mRNA, as shown in Fig. 4A. A major 2·2 kb transcript and a series of higher molecular weight minor RNA species, including bands at 3·8, 6, 10 and 12 kb are found in 12½- and 13½-day embryonic RNA. Most of these RNA species and an additional 1·9 kb transcript are also detected in kidney RNA, but the relative ratio of each species is very different from that of the embryo. In particular, the 3·8 kb RNA is as abundant as the 2·2 kb RNA in kidney. A similar complex pattern is shown in Fig. 4B, where we have extended this analysis to include a wider range of embryonic and adult tissues. Transcripts were detected in a large majority of the tissues and Fig. 4B illustrates the surprisingly wide range in number, size and intensity of the transcripts observed. It is important to note that these RNA samples were analysed under conditions of high stringency in both the hybridization and washing steps (see Methods). These conditions are usually sufficient to detect only highly homologous or identical sequences and in control experiments using a mouse β-actin probe we only detected a single RNA band. These bands do not represent nonspecific association of the probe with ribosomal RNAs, as no signal is found in lanes containing large amounts (20 μg) of poly(A)$^-$RNA from 12½-day mouse embryos (Fig. 5B).

This array of multiple RNA transcripts could be derived from the *Hox2.1* gene or might represent extensive homology to transcripts from other genes. As a means of examining this problem, the filters in Fig. 4B were treated with ribonuclease (RNase A) to remove any parts of the labelled probe that were not completely matched with other RNAs. This treatment made a dramatic difference in the results (Figs 4C, 5). Some tissues, such as testes, which previously revealed at least seven bands, showed no appreciable signal after this treatment. Patterns in other tissues (spinal cord and kidney for example) did not vary, while the adult lung had two bands that were removed and two that remained unchanged. We have obtained similar results using probes from sequences which are 3′ of the Hox2.1 homeobox. The results, showing that many bands are removed by RNase A treatment, suggest that transcripts from other loci have extensive homology to the *Hox2.1* gene in sequences both 5′ and 3′ of the homeobox. The series of bands resistant to RNase A demonstrate that multiple transcripts are also derived from the gene.

The results of Northern blot analysis, using the RNase A treatment, show that the 2·2 kb transcript is

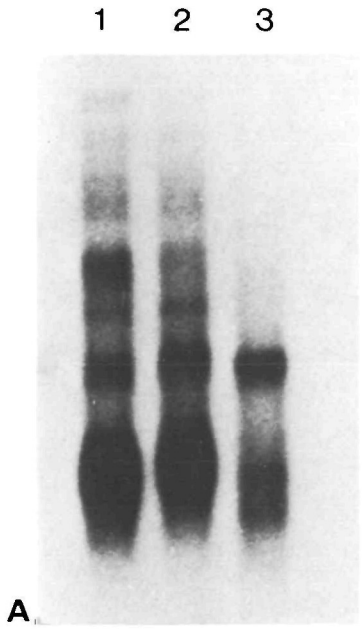
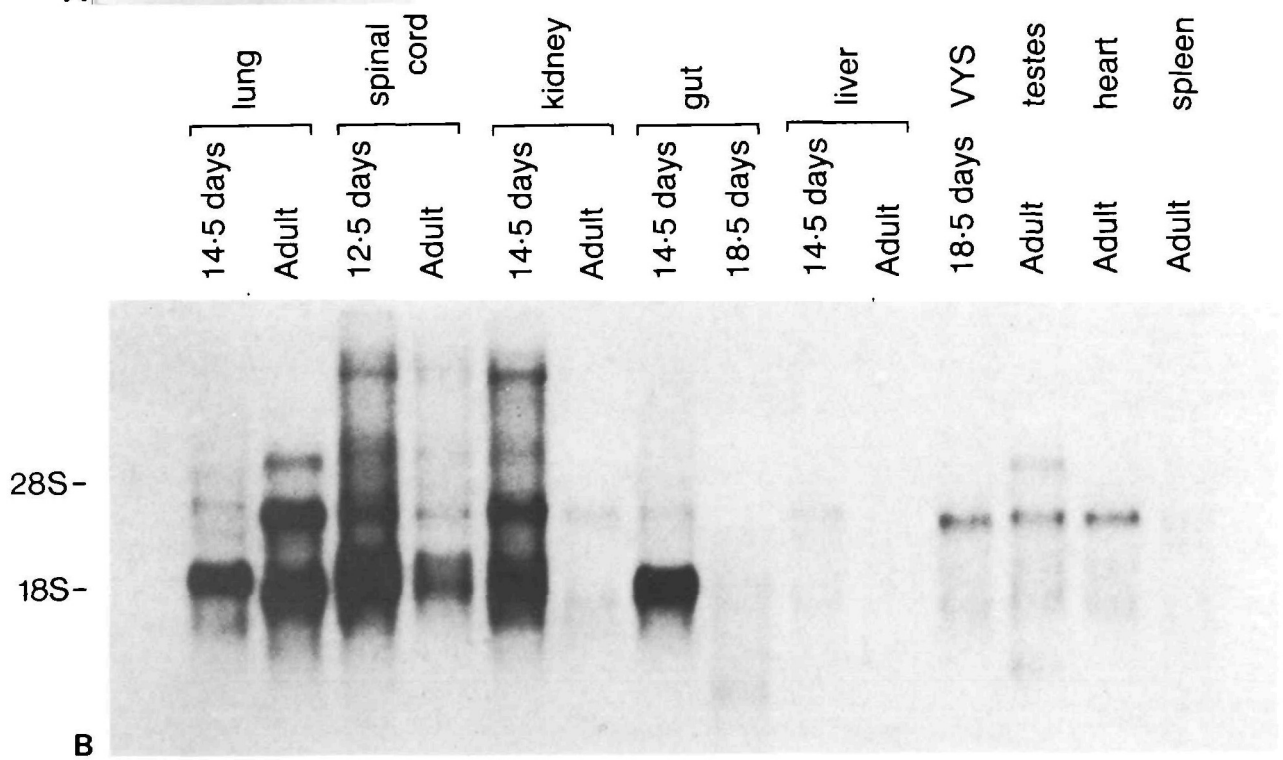
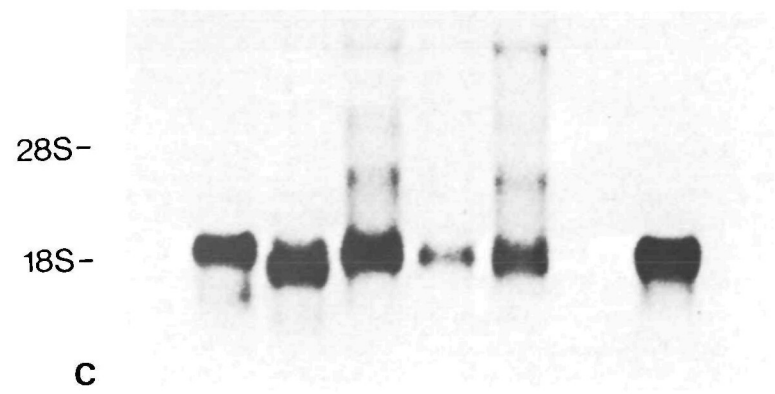


Fig. 4. Northern blot analysis of *Hox 2.1* expression in poly(A)$^+$RNA extracted from fetal and adult mouse tissues. Samples of RNA isolated from each tissue were electrophoresed on denaturing agarose gels, transferred to membranes and hybridized with a Hox 2.1 cDNA probe (probe 2, Fig. 1). (A) Lane 1, 12½-day; Lane 2, 13½-day embryo and Lane 3, adult kidney poly(A)$^+$RNA, 5 μg per lane. (B) The fetal or adult origin of the poly(A)$^+$ is indicated above each lane. In all samples 5 μg of RNA was loaded per lane, except for fetal lung, adult lung, adult spinal cord and visceral yolk sac (VYS) where 1 μg of RNA was applied. (C) The same filter as B treated with RNase A (see Methods). The relative mobilities of the 18 S (2·2 kb) and 28 S (4·7 kb) rRNAs are shown as size markers.





still the major species found in expressing tissues. The relative levels of the 2·2 kb transcript in embryonic and adult tissues analysed by Northern blotting experiments were quantified by densitometry, as described in Fig. 6. The fetal lung, spinal cord, gut, kidney, spleen, liver and visceral yolk sac contain the

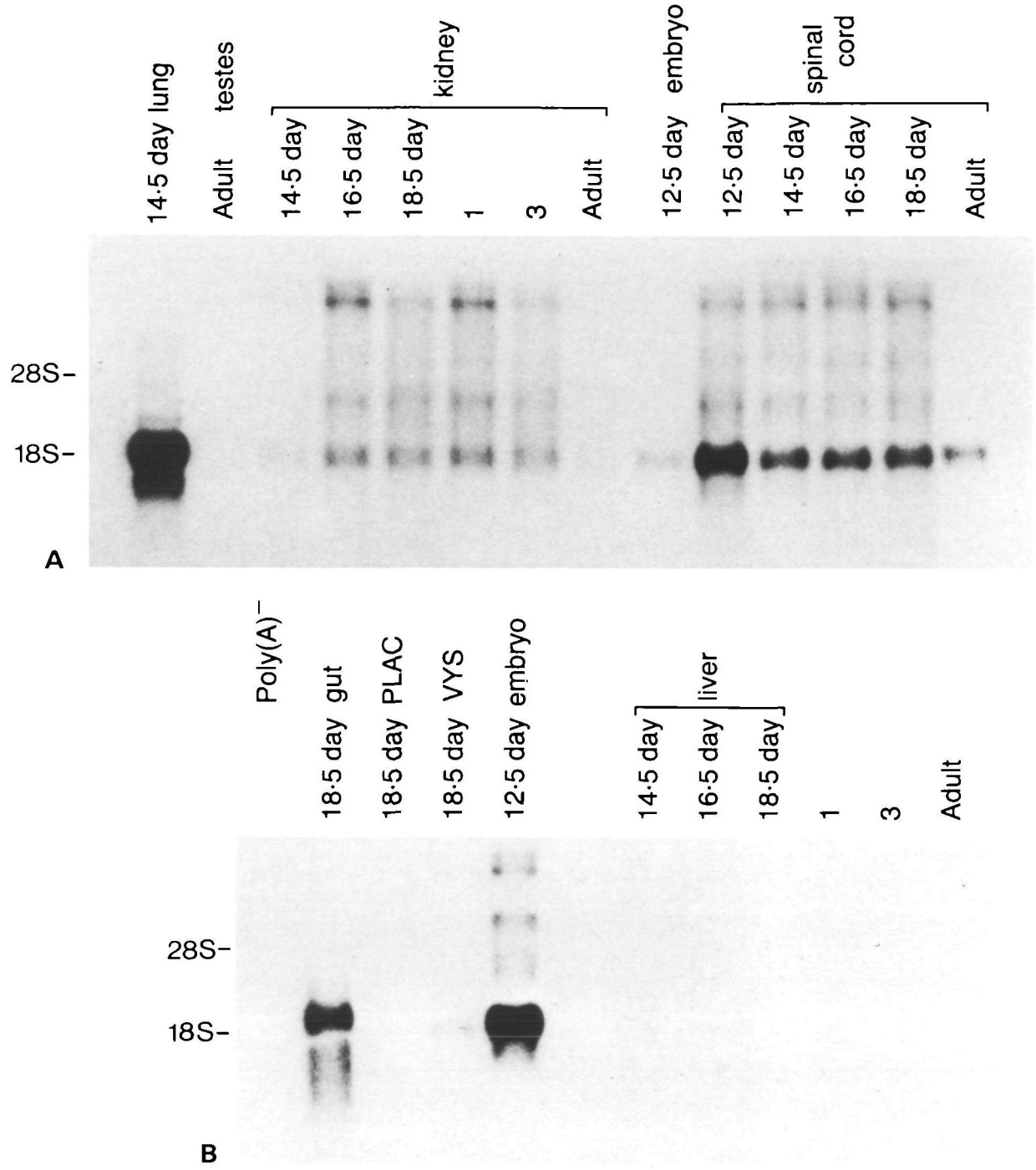


Fig. 5. Tissue-specific and temporal expression of *Hox2.1* in poly(A)$^+$RNA from mouse tissues. Above each lane is marked the tissue and stage in development when RNA was isolated. (A) Time course of *Hox2.1* expression in the kidney and spinal cord. The amount of poly(A)$^+$RNA loaded per lane was 5 μg, except in the adult spinal cord (1·5 μg), the 14½-day fetal kidney (2·5 μg), and the 12½-day total embryo (2·5 μg). (B) *Hox2.1* expression in liver and other weakly expressing tissues. Liver samples have 10 μg of poly(A)$^+$ per lane, the visceral yolk sac (VYS) and placenta (PLAC) 1 μg, the fetal gut and 12½-day embryo 2 μg and a control with 20 μg of poly(A)$^-$RNA from 12½-day mouse embryos. Both A and B were treated with RNase A. The relative mobilities of the 18 S (2·2 kb) and 28 S (4·7 kb) rRNAs are indicated as size markers.

2·2 kb RNA, and the relative levels of expression in each of these tissues varies over a 1000-fold range, as shown in Fig. 6. Negative tissues include fetal brain, heart, muscle, placenta, amnion and 12½-day parietal yolk sac. The temporal expression of the gene during development and in the adult mouse is also shown in Figs 4, 5 and 6. There is a small decrease in spinal cord expression between 12½ and 14½ days in the fetus, and the sizes of the two transcripts in adult lung are slightly smaller than in fetal lung, which may reflect different poly(A) lengths or different transcripts. However, in these two tissues there is essentially no change in the RNA levels between the fetal and adult stages. Levels of RNA decrease fourfold in the adult kidney as compared to fetal and neonatal kidney. In the gut, spleen and liver, RNAs are not detectable in the adult, and the *Hox2.1* gene therefore appears to be turned off after birth in these tissues. The gut RNA levels display the largest decrease between the fetal and adult stages. However, this result is complicated by the fact that the dissected 14½-day gut may also contain some spleen and pancreas. It is difficult to assess the relative contribution of these tissues to the level of expression at 14½ days in the gut and the difference may therefore be smaller than we have observed. However, there remains a large decrease in RNA levels between 18½-day gut, which is free of these contaminants, and adult intestine. The results summarized in Fig. 6 clearly show that the *Hox2.1* gene is modulated during development and that there are tissue-specific differences which control the level and timing of expression.

The complex RNA patterns could represent precursors, multiple transcription starts or differentially processed transcripts from the *Hox2.1* gene. At

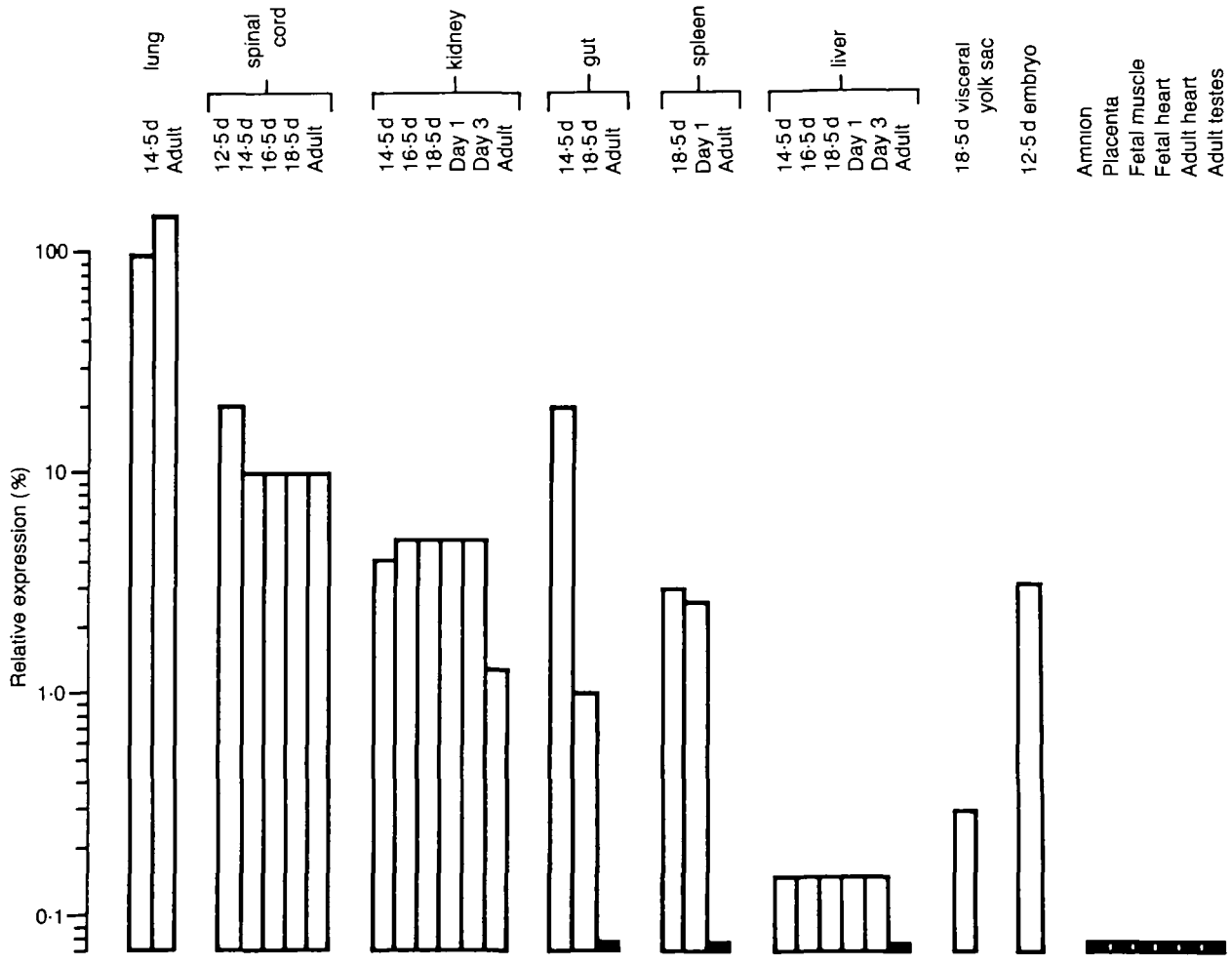


Fig. 6. Relative expression of the *Hox2.1* gene in embryonic and adult mouse tissues. The poly(A)$^+$RNA from the various tissues was analysed by Northern blotting in several experiments, including Figs 4, 5, and all filters were treated with RNase A (see Methods). The relative level of the 2·2 kb transcript for each tissue was determined by densitometer scanning of autoradiographs from multiple-timed exposures of each filter and compensated for by the amount of RNA loaded in each lane. The levels of RNA in the 12½-day embryo samples were used as an internal control in all scans and the level in the fetal lung was arbitrarily set as 100 %.

present we cannot distinguish between these alternatives, but two points must be made. First, the spinal cord and kidney characteristically have four different transcripts larger than the 2·2 kb RNA (3·8, 6, 10 and 12 kb) which hybridize to the Hox 2.1 probe. The lung also has the same species but they are in very low abundance. The level of each of these larger transcripts does not correlate with the level of the major 2·2 kb RNA, in that the ratio of larger species to the 2·2 kb RNA is different in each tissue. Second, the developmental time course for spinal cord and kidney (Figs 4, 5) shows that despite the different ratios of larger transcripts to the 2·2 kb RNA in these tissues, the relative ratio in a given tissue does not change during development. Therefore, regardless of how these RNA species are derived from the *Hox2.1* gene, they are coordinately regulated with the major Hox 2.1 2·2 kb transcript in a tissue-specific manner during development.

In situ *hybridization*

The Northern analysis has allowed us to identify general and temporal patterns of *Hox2.1* expression in the mouse embryo, but the relative levels of RNA from dissected tissue (Fig. 6) does not take into account that these tissues are composed of many cell types. Understanding any functional role for a homeobox-containing gene, however, requires more detailed information on the spatial and cell-type specificity of the gene expression. Toward this goal *in situ* hybridization was used to investigate the distribution of *Hox2.1* transcripts in the 12½-day mouse embryo. A part of the Hox 2.1 DNA was cloned into pGEM (Fig. 1, probe 2) to provide single-stranded RNA anti-sense (−) probes for mRNA and sense (+) control probes.

Results of *in situ* hybridization on frozen sections are shown in Fig. 7. Very strong hybridization, specific to the (−) strand probe, was seen in the embryonic spinal cord and lung. Hybridization in the spinal cord is most intense directly posterior to the hind brain, but it does extend into the hind brain region. There is less hybridization occurring more caudally, suggesting an apparent gradient from the anterior to posterior regions of the spinal cord. No signal above background is detected in the brain. Grain density was higher than background over dense aggregates of red blood cells (e.g. in dorsal aorta). This result, however, was observed with both (+) and (−) strand probes, and was therefore considered to be nonspecific. The hybridization pattern in the embryonic lung is particularly striking in a slightly more lateral section, which passes through three branches of the lung (Fig. 7C,D). Comparison of bright-field and dark-ground photomicrographs reveals that the hybridization grains are most intense in the mesenchymal cells that envelop the branching epithelial cell layer. There appears to be little or no expression in this lung epithelial layer.

Discussion

The striking degree of conservation observed in homeobox sequences from many species, including mammals, has led to the speculation that, by analogy to *Drosophila*, the mammalian homeobox counterparts may play an important role in regulation of development. *Drosophila* homeobox genes are involved in control of the segmental body plan by specifying the number, polarity and identity of segments in the embryo (reviewed by Gehring, 1985; Scott, 1985). However, it is difficult to make direct functional comparisons with mammals, as segmentation in vertebrates appears to have evolved independently of Arthropods, and it is still not known whether lineage restricted compartments are an important feature of mammalian development (for review, Hogan, Holland & Schofield, 1985). Genetic analysis of these developmental strategies in mouse or other mammals is much more difficult than in *Drosophila* and relatively few loci that affect morphogenesis are known. There is no clear evidence for mouse homeotic genes that affect unique somite identity (Hogan *et al.* 1985) and none of the mouse homeobox-containing genes isolated thus far have been shown to be allelic with mapped developmental mutants. Indeed, the only support for a role of homeobox genes in mouse development comes from studies on their expression. Reports have demonstrated that mouse homeobox genes are expressed in different tissues at both embryonic and adult stages (Colberg-Poley *et al.* 1985*a*,*b*; Jackson *et al.* 1985; Ruddle, Hart, Awgulewitsch, Fainsod, Utset, Dalton, Kerk, Rabin, Ferguson-Smith, Fienberg & McGinnis, 1985), and that, in the case of *Hox3*, the localization of the transcripts is spatially restricted in the central nervous system (Awgulewitsch *et al.* 1986). Therefore, to address the possible function of the mouse *Hox2.1* homeobox gene we have isolated cDNA sequences to analyse the structure of the predicted protein product and examined the tissue-specific, temporal and spatial patterns of its expression as a step towards linking gene activity with known developmental processes.

Jackson *et al.* (1985) have demonstrated that the *Hox2.1* gene is expressed in mouse embryos as early as day 7½ *p.c.* during the late primitive-streak-stage and it is enriched in fetal spinal cord and adult kidney. The results from our analysis show that the *Hox2.1* gene is expressed in a broad spectrum of embryonic

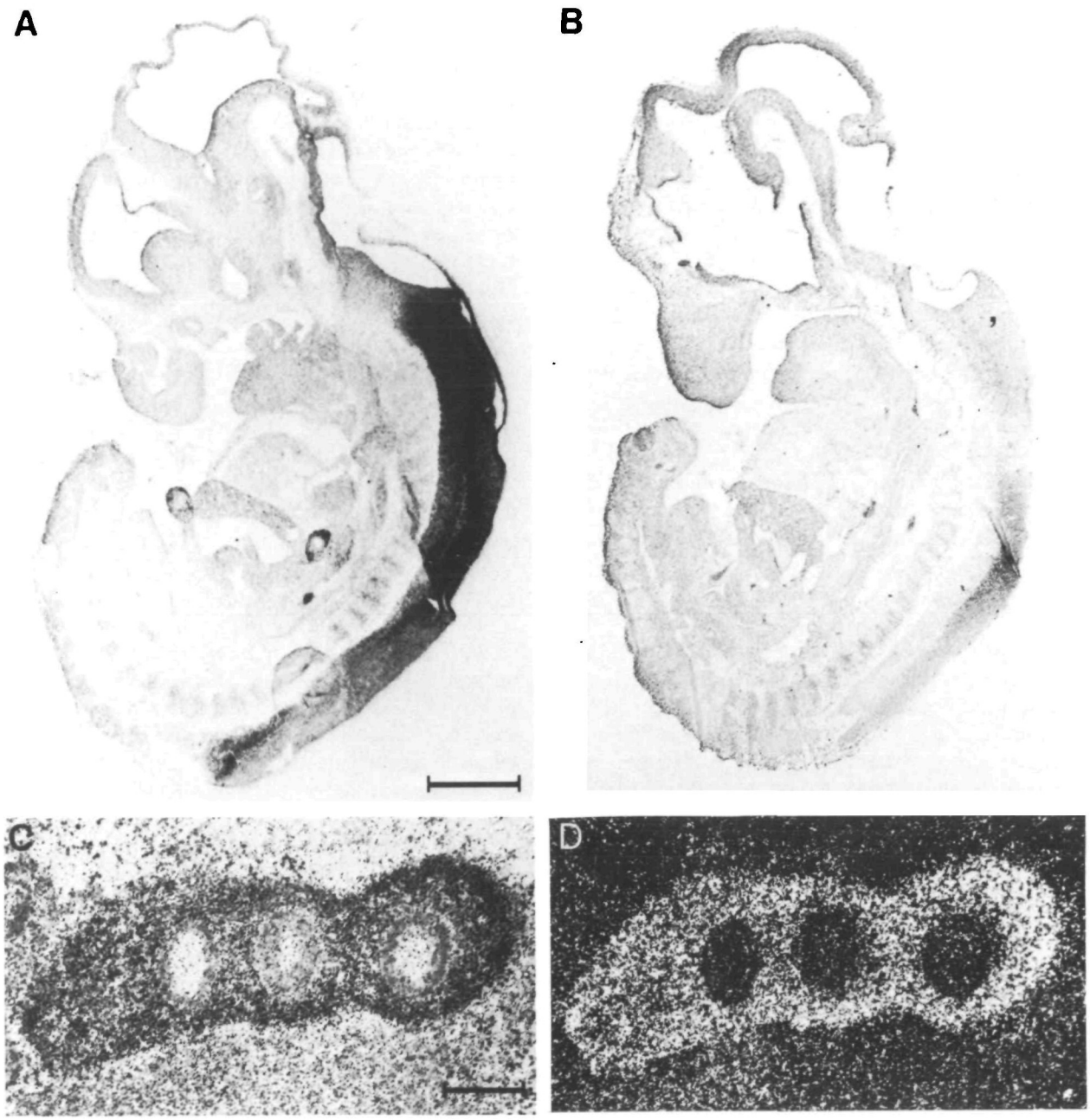


Fig. 7. Expression of *Hox2.1* sequences in sections of 12½-day mouse embryos revealed by *in situ* hybridization. Sagittal cryostat sections from frozen 12½-day embryos were hybridized with anti-sense (−) and sense (+) single-stranded Hox2.1 RNA probes (probe 1, Fig. 1). (A) Whole embryo bright field, low power, (−) probe; (B) whole embryo bright-field, low-power, (+) probe; (C) lung bright-field, high-power, (−) probe; (D) lung dark-field, high power, (−) probe. Bar in A represents 800 μm, and in C 100 μm.

tissues, including fetal lung, spinal cord, gut, kidney, spleen, liver and visceral yolk sac, but the levels of mRNA in each of these tissues varies over a considerable range (see Fig. 6). Negative tissues include fetal brain, heart, muscle, placenta, amnion and 12½-day parietal yolk sac. The fetal lung, spinal cord and gut have the highest levels of transcripts, and the level in fetal liver is the lowest that we are able to detect. The timing of *Hox2.1* expression during development also varies and displays tissue-specific differences. In general, the highest level of expression is observed in the earliest fetal stage dissected for each tissue, with the possible exception of lung. We have never detected transcripts in an adult tissue if not present in its embryonic or fetal counterpart. Due to the difficulty of dissecting sufficient material for Northern RNA hybridization, we have not been able to extend this analysis to time points earlier than 12½–14½ days *p.c.*

Therefore, we do not know whether *Hox2.1* transcription begins during the formation of organ primordia, in the early stages of organogenesis or at later times. Several distinct temporal patterns are illustrated in Fig. 6. Fetal and adult levels of expression are roughly equal in the lung and spinal cord, suggesting that once the gene is activated in these tissues it remains on throughout adult life. This agrees with the finding that *Hox2.1* transcripts are observed in the adult mouse central nervous system (Ruddle *et al.* 1985) and that the human *Hox2.1* homologue is expressed in the human spinal cord in early embryonic stages (Simeone *et al.* 1986). However, in the kidney, adult RNA levels are fourfold lower than fetal levels and the remaining tissues (gut, spleen and liver) do not appear to express the gene at all in adult stages. The decrease in levels of *Hox2.1* expression in some of these tissues (kidney, spleen and liver) occurs after birth, while in the gut this transition is detected between 14½–18½ days of fetal development.

The overall tissue distribution and temporal pattern of *Hox2.1* gene expression is different to that observed for other mouse homeobox genes (Colberg-Poley *et al.* 1985*b*; Awgulewitsch *et al.* 1986; Rubin, Toth, Patel, D'Eustachio & Nuguyen-Huu, 1986; Wolgemuth *et al.* 1986). However, some tissues, such as spinal cord and kidney, contain transcripts from several homeobox genes. These results imply that the individual mouse homeobox genes respond to different tissue-specific and developmental factors which regulate their expression. There is no obvious common origin or feature of these tissues that could account for the large tissue-specific variations in RNA levels and the timing and extent of temporal changes in *Hox2.1* gene expression observed. The pattern of embryonic expression reported in this study may therefore result from the action of multiple control elements or factors active during organogenesis in later stage embryos and adults. By contrast, expression in the earlier embryo may be linked to a different functional role. It will be essential, therefore, to investigate *Hox2.1* expression in 6½- to 8½-day *p.c.* embryos *via in situ* hybridization to observe the patterns of expression in the stages when the germ layers and the body axes are first established.

If the activity of mouse homeobox genes provides a framework for positional information in the embryo it is important to examine the cell-type specificity and spatial localization of transcripts during development. *In situ* hybridization experiments with the *Hox2.1* gene (Fig. 7) in 12½-day *p.c.* embryos clearly demonstrates that transcripts are localized in the spinal cord. The highest levels of RNA appear in the posterior region of the hind brain and extend in an apparent decreasing gradient to posterior regions of the spinal cord. The highest signal appears over the central portion of the cord and at the edges, but cells completely spanning the dorsal–ventral axis of the cord are also expressing *Hox2.1*. Many of these cells have very different fates and are no longer dividing, so there is no strict correlation of levels of expression with rates of cell division or fate in this region of the spinal cord. The *Drosophila* homeobox-containing genes of the ANT-C and BX-C complexes also specify transcripts that appear to accumulate in discrete regions of the mature embryonic central nervous system (Harding, Wedeen, McGinnis & Levine, 1985; Levine, Hafen, Garber & Gehring, 1983). The spatial localization of Hox2.1 in the central nervous system is similar to that observed with the *Hox3* gene (Awgulewitsch *et al.* 1986). However, the *Hox2.1* transcripts appear to be expressed in a more anterior portion of the spinal cord. Therefore, these two genes have similar overlapping but not identical regional localizations which may be a feature of other mammalian homeobox genes as in *Drosophila*.

The *in situ* hybridization results (Fig. 7C) also reveal that *Hox2.1* is not uniformly expressed in all cells of the 12½-day *p.c.* embryonic lung. The mesenchymal cells accumulate high levels of Hox2.1 RNA, in marked contrast to the epithelial cell layer that they surround. Essentially no signal is detected over the epithelial cells. The high level of expression in the lung mesoderm cells is not a general property of all mesoderm cells, as many other mesodermally derived tissues in the 12½-day embryo show no signal with the Hox2.1 probe (Fig. 7A,B). During this stage the lung epithelia have been induced to branch and it is important to examine whether *Hox2.1* expression is involved in, or marks, this process in the lung.

One of the characteristics of *Drosophila* homeobox genes, like *Antennapedia* (Antp), is that they have multiple promoters, which generate a series of different RNA transcripts by alternative RNA splicing, and these transcripts are able to encode similar but different proteins (Scott, Weiner, Polisky, Hazelrigg, Pirotta, Scalengne & Kaufman, 1983; Carroll *et al.* 1986; Schneuwly *et al.* 1986). The transcriptional analysis of the *Hox2.1* gene (Figs 4, 5) also revealed a complex pattern of expression involving multiple transcripts. This raises an important point concerning any *Hox2.1* gene product(s). We have assumed that the Hox2.1 protein is synthesized in parallel with the accumulation of the mRNA, but if multiple proteins are possible a detailed analysis on the distribution of specific products is required to address a functional role. We are presently using the predicted amino acid sequence from our cDNA clones (Fig. 2) to generate Hox2.1-specific antibodies for this purpose.

Analysis of the predicted Hox2.1 amino acid sequence shows that the protein contains high levels

of serine (22%) and proline (9%) in the domains outside of the homeobox, and that a hexapeptide region upstream of the homeodomain is conserved in homeobox genes from several species (Fig. 3). A consensus squence for this peptide shows an aliphatic amino acid (Ile/Leu/Val) in position one, Phe or Tyr in position two, followed by the tripeptide Pro-Trp-Met, ending in the basic residue either Lys or Arg. The conservation of the sequence and its location relative to the homeodomain raises the possibility that it is functionally related to the homeodomain. The *Hox2.1* gene must have additional regions of homology with other mouse loci. The Northern analysis experiments (Fig. 4) carried out under high stringency conditions with and without RNase A, clearly demonstrate that there are transcripts in many tissues with extensive homology to the *Hox2.1* gene probe. If the clusters of mouse homeobox genes (Colberg-Poley, 1985*a*,*b*; Hart *et al.* 1985; Dubuole *et al.* 1986) arose *via* duplication and divergence, then some members of these gene clusters should be very homologous. The degree of *Hox2.1* homology with other family members can only be determined when further sequences are available. Under conditions of reduced stringency we have observed (unpublished) that probe 2 recognizes Hox2.1 homologous sequences in frog, chicken, cow, cat and human genomic DNA but not in *Drosophila*. Therefore, the *Hox2.1* gene appears to be conserved in vertebrate evolution and hence comparisons of the Hox2.1 protein products from several species could provide clues to important functional domains.

The results presented here provide a detailed descriptive picture of the diversity in tissue-specific, temporal, cell type and spatial expression of the *Hox2.1* gene. As similar information on other vertebrate homeobox-containing genes arises, there will be a basis for examining functional relationships between the different genes. However, direct confirmation of an embryonic functional role is still required. The production of transgenic mice *via* introduction of modified genes can test for dominant effects in embryogenesis. Coupled with the use of antibody probes to investigate protein expression in the early embryo, it may be possible to resolve many of the outstanding questions.

We thank Dr Bob Hill for help in screening the cDNA library, Drs Peter Gruss and Ward Odenwald for unpublished sequence information, Raechel Kenney for help in sequencing, Sarah Harper for technical help on *in situ* hybridization, Dr Colin Hetherington for animal assistance and Lydia Pearson for preparing the manuscript. We also thank Drs Mike Snow, David Murphy, Nick Hastie, Robin Lovell-Badge, Jack Price, Dennis Summerbell and Ian Jackson for valuable discussion.

References

Auffray, C. & Rougeon, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* **107**, 303–324.

Aviv, H. & Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412.

Awgulewitsch, A., Utset, M. F., Hart, C. F., McGinnis, W. & Ruddle, F. H. (1986). Spatial restriction in expression of a mouse homeobox locus within the central nervous system. *Nature, Lond.* **320**, 328–335.

Beachy, P. A., Halfand, S. L. & Hogness, D. S. (1985). Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature, Lond.* **313**, 545–551.

Bishop, J. O., Rosbash, M. & Evans, D. (1974). Polynucleotide sequences in eukaryotic DNA and RNA that form ribonuclease-resistant complexes with polyuridylic acid. *J. molec. Biol.* **85**, 75–86.

Boncinelli, E., Simeone, A., La Volpe, A., Faiella, A., Fidanza, V., Acampora, D. & Scotto, L. (1985). Human cDNA clones containing homeobox sequences. *Cold Spring Harbor Symp. Quant. Biol.* **50**, 301–306.

Breir, G., Bucan, M., Francke, U., Colberg-Poley, A. M. & Gruss, P. (1986). Sequential expression of murine homeobox genes during F9 EC cell differentiation. *EMBO J.* **5**, 2209–2215.

Carrasco, A. E., McGinnis, W., Gehring, W. J. & De Robertis, E. M. (1984). Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes. *Cell* **37**, 409–414.

Carroll, S. B., Laymon, R. A., McCutcheon, M. A., Riley, P. D. & Scott, M. P. (1986). The localization and regulation of *Antennapedia* protein expression in *Drosophila* embryos. *Cell* **47**, 113–122.

Carroll, S. B. & Scott, M. P. (1985). Localization of the *fushi tarazu* protein during *Drosophila* embryogenesis. *Cell* **43**, 44–57.

Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzym.* **47**, 415–419.

Colberg-Poley, A. M., Voss, S. D., Chowdhury, K. & Gruss, P. (1985*a*). Structural analysis of murine genes containing homeobox sequences and their expression in embryonal carcinoma cells. *Nature, Lond.* **314**, 713–718.

Colberg-Poley, A. M., Voss, S. D., Chowdhury, K., Stewart, C. L., Wagner, E. F. & Gruss, P. (1985*b*). Clustered homeoboxes are differentially expressed during murine development. *Cell* **43**, 39–45.

Desplan, C., Theis, J. & O'Farrell, P. H. (1985). The *Drosophila* developmental gene, *engrailed*, encodes a sequence-specific DNA binding activity. *Nature, Lond.* **318**, 630–635.

Dolecki, G. J., Wannakrairoj, S., Lum, R., Wang, G., Riley, H. D., Carlos, R., Wang, A. & Humphreys,

T. (1986). Stage-specific expression of a homeobox-containing gene in the non-segmented sea urchin embryo. *EMBO J.* **5**, 925–930.

DUBUOLE, D., BARON, A., MAHL, P. & GALLIOT, B. (1986). A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *EMBO J.* **5**, 1973–1980.

FARNHER, K., HOGAN, B. L. M. & FLAVELL, R. (1987). Transcription of H-2 and Qa genes in embryonic and adult mice. *EMBO J.* (in press).

GARBER, R. L., KUROIWA, K. & GEHRING, W. J. (1983). Genomic and cDNA clones of the homeotic gene *Antennapedia Drosophila*. *EMBO J.* **2**, 2027–2034.

GEHRING, W. J. (1985). The homeobox: a key to the understanding of development? *Cell* **40**, 3–5.

HARDING, K., WEDEEN, C., MCGINNIS, W. & LEVINE, M. (1985). Spatially regulated expression of homeotic genes in *Drosophila*. *Science* **229**, 1236–1242.

HART, C. P., AWGULEWITSCH, A., FAINSOD, A., MCGINNIS, W. & RUDDLE, F. H. (1985). Homeo box gene complex on mouse chromosome 11: molecular cloning, expression in embryogenesis and homology to a human homeo box locus. *Cell* **43**, 9–18.

HARVEY, R. P., TABIN, C. J. & MELTON, D. A. (1986). Embryonic expression and nuclear localization of *Xenopus* homeobox (Xhox) gene products. *EMBO J.* **5**, 1237–1244.

HAUSER, C. A., JOYNER, A. L., KLEIN, R. D., LEARNED, T. K., MARTIN, G. R. & TJIAN, R. (1985). Expression of homologous homeo box containing genes in differentiated human teratocarcinoma cells and mouse embryos. *Cell* **43**, 19–28.

HOGAN, B. L. M., COSTANTINI, F. & LACY, E. (1986). *Manipulating the Mouse Embryo. A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

HOGAN, B. L. M., HOLLAND, P. W. H. & SCHOFIELD, P. (1985). How is the Mouse Segmented? *Trends in Genetics* **1**, 67–74.

HOLLAND, P. W. H. & HOGAN, B. L. M. (1986). Phylogenetic distribution of Antennapedia-like homeoboxes. *Nature, Lond.* **321**, 251–253.

JACKSON, I. J., SCHOFIELD, P. & HOGAN, B. L. M. (1985). A mouse homeobox gene is expressed during embryogenesis and in adult kidney. *Nature, Lond.* **317**, 745–748.

JAENISCH, R., HARBERS, K., SCHNIEKE, A., LOHLER, J., CHUMAKOV, I., JAHNER, D., GROTKOPP, D. & HOFFMAN, E. (1985). Germline integration of Moloney murine leukemia virus at the MOV13 locus leads to recessive lethal mutation and early embryonic death. *Cell* **32**, 209–216.

JOHNSON, A. D. & HERSKOWITZ, L. (1985). A repressor (MAT a2 product) and its operator control expression of a set of cell type-specific genes in yeast. *Cell* **42**, 237–247.

JOYNER, A. L., KORNBERG, T., COLEMAN, K. G., COX, D. R. & MARTIN, G. R. (1985). Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila engrailed* gene. *Cell* **43**, 29–37.

KAUFMAN, T. C., LEWIS, R. & WAKIMOTO, B. (1980). Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosome interval 84A-B. *Genetics* **94**, 115–133.

KRUMLAUF, R., HAMMER, R., TILGHMAN, S. M. & BRINSTER, R. (1985). Developmental regulation of alpha-fetoprotein in transgenic mice. *Mol. cell. Biol.* **5**, 163–168.

KUROIWA, A., KLOTER, U., BAUMGARTNER, P. & GEHRING, W. J. (1985). Cloning of the homeotic *sex combs reduced* gene in *Drosophila*. An *in situ* localization of its transcripts. *EMBO J.* **4**, 3757–3764.

LAUGHON, A., CARROLL, S. B., STORFER, F. A., RILEY, P. D. & SCOTT, M. A. (1985). Common properties of proteins encoded by the Antennapedia complex genes of *Drosophilia melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* **50**, 253–262.

LAUGHON, A. & SCOTT, M. P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature, Lond.* **310**, 25–31.

LEVINE, M., HAFEN, E., GARBER, R. L. & GEHRING, W. J. (1983). Spatial distribution of *Antennapedia* transcripts during *Drosophila* development. *EMBO J.* **2**, 2037–2046.

LEVINE, M., RUBIN, G. N. & TJIAN, R. (1984). Human DNA sequences homologous to a protein coding region conserved between homeotic genes of *Drosophila*. *Cell* **38**, 667–673.

LEWIS, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature, Lond.* **276**, 565–570.

MAXAM, A. M. & GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavage. *Meth. Enzymol.* **65**, 499–560.

MCGINNIS, W. (1985). Homeobox sequences of the *Antennapedia* class are conserved only in higher animal genomes. *Cold Spring Harbor Symp. Quant. Biol.* **50**, 263–270.

MCGINNIS, W., LEVINE, M. L., HAFEN, E., KUROIWA, A. & GEHRING, W. J. (1984*a*). A conserved DNA sequence in homeotic genes of the *Drosophila*, Antennapedia and Bithorax complexes. *Nature, Lond.* **308**, 428–433.

MCGINNIS, W., GARBER, R. L., WIRZ, J., KUROIWA, A. & GEHRING, W. J. (1984*b*). A homologous protein-coding sequence in *Drosophila* genes and its conservation in other metazoans. *Cell* **37**, 403–408.

MCGINNIS, W., HART, C. P., GEHRING, W. J. & RUDDLE, F. H. (1984*c*). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. *Cell* **38**, 675–680.

MELTON, D., KRIEG, P., REBAGLIATI, M., MANIATIS, T., ZINN, K. & GREEN, M. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA probes from plasmids containing a bacteriophage SP6 promoter. *Nucl. Acids Res.* **12**, 7035–7056.

MLODZIK, M., FJOSE, A. & GEHRING, W. J. (1985). Isolation of caudal, a *Drosophila* homeobox-containing gene with maternal expression, whose transcripts form

a concentration gradient at the pre-blastoderm stage. *EMBO J.* **4**, 2961–2969.

MULLER, M. M., CARRASCO, A. E. & DE ROBERTIS, E. M. (1984). A homeo-box-containing gene expressed during oogenesis in *Xenopus*. *Cell* **39**, 157–162.

MUNKE, M., COX, D. R., JACKSON, I. J., HOGAN, B. L. M. & FRANCKE, U. (1986). The murine Hox-2 cluster of homeobox containing genes maps distal on chromosome 11 near the *tail-short* (Ts) locus. *Cytogen. Cell Genets.* **42**, 236–240.

POOLE, P. W. J., KAUVAR, L. M., DREES, B. & KORNBERG, T. (1985). The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell* **40**, 37–43.

RABIN, M., HART, C. P., FERGUSON-SMITH, A., MCGINNIS, W., LEVINE, M. & RUDDLE, F. H: (1985). Two homeo box loci mapped in evolutionarily related mouse and human chromosomes. *Nature, Lond.* **314**, 175–177.

REGULSKI, M., HARDING, K., KOSTRIKEN, R., KARCH, F., LEVINE, M. & MCGINNIS, W. (1985). Homeo box genes of the Antennapedia and Bithorax complexes of *Drosophila*. *Cell* **43**, 71–80.

ROSBASH, M. & FORD, P. J. (1974). Polyadenylic acid-containing RNA in *Xenopus laevis* oocytes. *J. molec. Biol.* **85**, 87–101.

RUBIN, M. R., TOTH, L. E., PATEL, M. D., D'EUSTACHIO, P. & NUGUYEN-HUU, M. C. (1986). A mouse homeobox gene is expressed in spermatocytes and embryos. *Science* **233**, 663–667.

RUDDLE, F. H., HART, C. P., AWGULEWITSCH, A., FAINSOD, A., UTSET, M., DALTON, D., KERK, N., RABIN, M., FERGUSON-SMITH, A., FIENBERG, A. & MCGINNIS, W. (1985). Mammalian homeobox genes. *Cold Spring Harbor Symp. Quant. Biol.* **50**, 277–284.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

SCHNEUWLY, S., KUROIWA, A., BAUMGARTNER, P. & GEHRING, W. J. (1986). Structural organization and sequence of the homeotic gene *Antennapedia* of *Drosophila melanogaster*. *EMBO J.* **5**, 733–739.

SCOTT, M. P. (1985). Molecules and puzzles from the Antennapedia homeotic gene complex of *Drosophila*. *Trends in Genetics* **1**, 74–80.

SCOTT, M. P. & WEINER, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4115–4119.

SCOTT, M. P., WEINER, A. J., POLISKY, B. A., HAZELRIGG, T. I., PIRROTTA, V., SCALENGNE, F. & KAUFMAN, T. C. (1983). The molecular organization of the Antennapedia complex of *Drosophila*. *Cell* **35**, 763–776.

SHEPHERD, J. C. W., MCGINNIS, W., CARRASCO, A. E., DE ROBERTIS, E. M. & GEHRING, W. J. (1984). Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. *Nature, Lond.* **310**, 70–71.

SIMEONE, A., MALVILIO, F., BOTTERO, L., GIAMPAOLO, A., RUSSO, G., FAIELLA, A., BONCINELLI, E. & PESCHLE, C. (1986). A human homeo box gene specifically expressed in spinal cord during embryonic development. *Nature, Lond.* **320**, 763–765.

WHITE, R. A. H. & WILCOX, M. (1984). Protein products of the Bithorax complex in *Drosophila*. *Cell* **39**, 163–171.

WHITEWAY, M. & SZOSTAK, J. W. (1985). The ARDI gene of yeast functions in the switch between the mitotic cell cycle and alternative developmental pathways. *Cell* **43**, 483–492.

WOLGEMUTH, D. J., ENGELMYER, E., DUGGAL, R. N., GIZANG-GINSBERG, E., MUTTER, G. L., PONZETTO, C., VIVIANO, C. & ZAKERI, Z. F. (1986). Isolation of a mouse cDNA coding for a developmentally regulated testis-specific transcript containing homeo box homology. *EMBO J.* **5**, 1229–1235.

WOYCHICK, R. P., STEWART, T. A., DAVIS, R., D'EUSTACHIO, P. & LEDER, P. (1985). An inherited limb deformity created by insertional mutagenesis in transgenic mice. *Nature, Lond.* **318**, 36–40.

(*Accepted 12 January 1987*)