

***Pax2*, a new murine paired-box-containing gene and its expression in the developing excretory system**

GREGORY R. DRESSLER*, URBAN DEUTSCH, KAMAL CHOWDHURY, HOWARD O. NORNEST† and PETER GRUSS‡

Department of Molecular Cell Biology, Max Planck Institute for Biophysical Chemistry, 3400 Goettingen, FRG

*Present address: Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institute of Health, Bethesda, Maryland 20892, USA

† On leave from the Department of Anatomy and Neurobiology, Colorado State University, Fort Collins, Colorado 80523, USA

‡ Corresponding author

Summary

The murine genome contains multiple genes with protein domains homologous to the *Drosophila* paired box, present in certain segmentation genes. At least one of these murine paired box (*Pax*) genes is associated with a developmental mutation. This report, in conjunction with the accompanying paper, describes a second member of this gene family, *Pax2*, that is also expressed during embryogenesis. Two overlapping cDNA clones were isolated and sequenced. At least two forms of the *Pax2* protein can be deduced from the cDNA sequence. In addition to the highly conserved paired domain, an

octapeptide sequence is located downstream. Expression of *Pax2* is primarily restricted to the developing embryo in the excretory and central nervous systems. The transient nature of *Pax2* expression during kidney organogenesis correlates with polarization and induction of epithelial structures and may indicate an important morphogenetic role for this gene.

Key words: paired box, *Pax*, kidney development, mouse embryology.

Introduction

The identification and characterization of *Drosophila* developmental mutations has helped to elucidate the morphogenetic mechanisms of embryonic segmentation and pattern formation (for reviews see Ingham, 1988; Akam, 1987; Scott and Carroll, 1987). As an approach to address genetic aspects of vertebrate development, three different mouse gene families have been identified based on sequence similarities to certain *Drosophila* homeotic and segmentation genes (for review see Dressler and Gruss, 1988). The conservation of protein domains among such highly divergent species suggests a similarity of biochemical function that may control common morphological processes, or alternatively, may be adapted to more unique species-specific events. Many vertebrate genes, with domains homologous to the *Drosophila Antp*-type homeobox, have been identified and much of the available evidence suggests a developmental role for these genes (for reviews see Wright *et al.* 1989; Holland and Hogan, 1988).

A second conserved protein domain, termed the

paired box, was discovered among the *Drosophila* segmentation genes *paired* (*prd*), *gooseberry-proximal* (*gsb-p*), and *gooseberry-distal* (*gsb-d*) (Bopp *et al.* 1986). The mouse genome also contains multiple paired-box-containing genes, called *Pax* genes (Dressler *et al.* 1988). A member of this gene family that is expressed in segmented structures of the developing vertebral column was recently described (Deutsch *et al.* 1988). Furthermore, there is sufficient evidence to suggest that this gene, *Pax1*, is the mouse developmental mutation *undulated* (Balling *et al.* 1988). In *undulated* mice, vertebral malformations occur presumably because sclerotome cells either do not migrate and resegment properly in their respective halves or the differentiation of sclerotome cells into prevertebrae and intervertebral disk anlagen is affected (Grüneberg, 1954). Clearly, the evidence indicates that a mutation in the paired box disrupts a specific morphogenetic event during mouse development.

Since the mouse genome contains multiple copies of the paired box, it is of interest to characterize other members of the *Pax* gene family and determine if they also are involved in developmental processes. This

report, in conjunction with the accompanying paper (Nornes *et al.* 1990), describes the identification, primary structure and developmental expression of *Pax2*.

Materials and methods

Screening of mouse cDNA libraries

After transfer to Hybond-N filters (Amersham), 8×10^5 clones of a 12.5 days p.c. embryonic mouse λ gt10 cDNA library (Joyner and Martin, 1987) were hybridized under low-stringency conditions ($7 \times \text{SSC}$, 60°C) with a 313 bp *HincII-SacI Pax1* paired-box probe, labeled by random priming (Feinberg and Vogelstein, 1983). The filters were washed three times in $2 \times \text{SSC}/0.2\%$ SDS at 42°C . The cDNA clone c31A was isolated by standard techniques.

The cDNA clone cPX was isolated from a 8.5 days p.c. embryonic mouse λ gt10 cDNA library (Fahrner *et al.* 1987) in a low-stringency screen using a mixture of *Pax1* (313 bp *HincII-SacI* fragment), *Pax2* (574 bp *NotI-BamHI* fragment), and *Pax3* (Deutsch, unpublished) paired-box probes as described above.

DNA sequencing

The cDNA inserts were cloned into the plasmid vector Bluescript (Stratagene Inc.) and mapped with various restriction enzymes. Subfragments were cloned into M13 and sequenced with the dideoxy method (Sanger *et al.* 1977). The GC rich *EcoRI-NotI* fragment, spanning nucleotides 1–295, was also sequenced with the method of Maxam and Gilbert (1980).

RNA isolation from mouse tissue and embryos

Female NMRI outbred mice were mated overnight and the day of the vaginal plug was noted as day 0. Embryos were dissected free of extraembryonic tissue and frozen in liquid nitrogen. RNA was isolated by homogenizing tissues and embryos in guanidinium thiocyanate (Chirgwin *et al.* 1979) and centrifugation through a 5.7 M CsCl , 25 mM sodium acetate pH 5.0 cushion in a Beckman SW41 rotor spun at $31\,000 \text{ rev min}^{-1}$ for 22 h. Poly(A)⁺ RNAs were obtained by retention on oligo(dT)-cellulose columns. Approximately $5 \mu\text{g}$ of the respective RNAs were electrophoresed through a 1% agarose, 3.7% formaldehyde gel in 20 mM MOPS (morpholine propane sulfonic acid) buffer. The RNA was blotted onto gene screen plus membranes (New England Nuclear) and hybridized with the 527 bp *BamHI-EcoRI* fragment of c31A. The probe was labeled with the random prime method of Feinberg and Vogelstein (1983). The membranes were washed twice in $2 \times \text{SSC}$, 1% SDS at 65°C and twice in $0.1 \times \text{SSC}$, 1% SDS at 65°C .

In situ hybridization

The 527 bp *BamHI-EcoRI* fragments from clone c31A was subcloned into the Bluescript vector (Stratagene) and the resulting plasmid was linearized with restriction endonuclease. Single-stranded RNA probes were transcribed *in vitro* using $100 \mu\text{Ci } ^{35}\text{S-UTP}$ and T3 or T7 polymerases (Promega Biotech.). After DNase digestion, probes were precipitated with 10% trichloroacetic acid and collected on nitrocellulose filters (Millipore). Probes were eluted from the filters in 20 mM EDTA pH 8.0, 0.1% SDS at 65°C . Following ethanol precipitation, probes were partially degraded with 0.2 N NaOH on ice for 30–60 min and neutralized with 1 M acetic acid. After ethanol precipitation, the probes were resuspended in 50% formamide, 10 mM DTT.

Sections were prepared and hybridized essentially as described by Hogan *et al.* (1986) with modifications by Dony and Gruss (1987). Sections were cut in a cryostat and transferred onto subbed slides. The sections were dried at 55°C , fixed in 4% paraformaldehyde (PFA) and dehydrated in graded ethanol. Slides were kept at -20°C until the day of hybridization. Slides were dipped in phosphate-buffered saline (PBS) and incubated at 70°C in $2 \times \text{SSC}$ (standard saline citrate). After a second PBS rinse, slides were digested with 0.125 mg ml^{-1} pronase for 10 min at room temperature and the digestion was stopped in 0.2% glycine for 30 sec. Slides were rinsed in PBS and refixed in 4% PFA for 20 min and rinsed again in PBS. Slides were acetylated in 0.1 M triethanolamine with $1/400$ volume acetic anhydride, made fresh. Slides were rinsed again in PBS and dehydrated in graded ethanol. After prolonged air drying, the hybridization mixture was added. The probe was diluted to approximately $5 \times 10^4 \text{ cts min}^{-1} \mu\text{l}^{-1}$ in a buffer containing 50% formamide, 0.3 M NaCl , 10 mM Tris , 10 mM NaPO_4 (pH 6.8), 5 mM EDTA , $1 \times \text{Denhardtts}$, 10% dextran sulphate, 10 mM DTT , and 1 mg ml^{-1} tRNA. The hybridization mix was boiled for 2 min, applied directly onto sections and covered with a siliconized cover slip. Hybridization was done overnight in a humid chamber at 50°C . Washing was done for 3–4 h in 50% formamide, $2 \times \text{SSC}$ at 37°C followed by RNase digestion (Ingham *et al.* 1985). A second wash in 50% formamide, $2 \times \text{SSC}$ was done overnight and the slides were then dehydrated in graded ethanol.

Slides were dipped in Kodak NTB-2 emulsion diluted 1:1 with water and allowed to dry in a dark chamber for 2–3 h. Slides were placed in a dark plastic box, wrapped in foil, and allowed to expose for 6–10 days. Development was done at room temperature for 3 min in Kodak D-19, followed by 30 s in 1% acetic acid and 3 min in 30% sodium thiosulphate. After repeated washes in distilled water, the slides were stained with Giemsa and allowed to dry. Photomicrographs were taken with a Leitz Labovert bright-field/dark-field microscope.

Results

Structure of the *Pax2* cDNAs

The two partially overlapping cDNA clones, corresponding to the mouse *Pax2* mRNA, are outlined in Fig. 1. The clone c31A was isolated from a 12 day embryonic cDNA library using the mouse *Pax1* paired box as a probe, under low-stringency conditions. Sub-

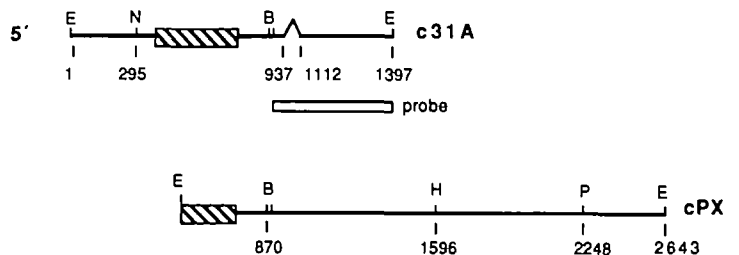


Fig. 1. The structure of *Pax2* cDNAs. The two overlapping *Pax2* cDNAs are outlined. The paired domain is indicated by a hatched box. The probe used for *in situ* hybridization is indicated by an open box. Restriction sites are: B, *BamHI*; E, *EcoRI*; H, *HindIII*; N, *NotI*; P, *PstI*.

sequently, an 8.5 day embryonic cDNA library was screened with a mixture of three murine paired box probes under low-stringency and one additional *Pax2* cDNA clone, cPX, was identified upon further DNA sequencing. Together, the two cDNAs span 2623 bp of the *Pax2* mRNA. The complete nucleotide sequence of the two overlapping cDNA clones is shown in Fig. 2.

The clone c31A does not have a poly(A)⁺ tract at its 3' end and, thus, may be the result of an internal cleavage during the construction of the library. The clone cPX does contain a stretch of 15 adenosine residues at the 3' end; however, no AAUAAA polyadenylation signal (Proudfoot and Brownlee, 1976;

Fitzgerald and Shenk, 1981) is present. Thus, it is unlikely that clone cPX represents the true 3' end of the *Pax2* mRNAs. A comparison of the overlapping region between c31A and cPX revealed a gap of 69 nucleotides in clone c31A between position 937 and 1005. Otherwise, the two cDNA clones are completely colinear. It is likely that this deletion represents an alternative splicing pathway during *Pax2* mRNA maturation. Interestingly, this small deletion does not change the reading frame of the coding sequence.

The Pax2 protein

The overlapping *Pax2* cDNAs encode an open reading

1	GGAAATCCGGCAGGCGCGCCGCGGCGACCGAAGGGCCCTTGGTCCAGCCAGTCGCCCGGGGCTGCGTTGCTGACTGCCAGCCCGG	90
91	TGCCCAACGGTGGCAAGTGCAGCCTTCGGGTCCGAAACCCAGCGGGCCGAGGAGTCCCGCGTGGAGCGCGCGCGTCCCC	180
181	CGCCCTCGCTTGGCCACGATCGGGCGCCGCTACTCTCCCTCCCCACCGTCCCTTCCTTTCTCTCAAGTCTGAAGTTGAGT	270
271	TTGAGAGGCGACACGGCGGGCGGCGCGCTTGTCTCCGCTCCTCTGCCTCCCGATGGATATGCACTGCAAAGCAGACCCCTTCTCCGCG	360
1	MetAspMetHisCysLysAlaAspProPheSerAla	12
361	ATGCACCGGCACGGGGTGTGAACAGCTCGGGGGGGTGTTTGTGAACGGCGGCCCTACCCGACGTGGTGGAGCAGCGCATCGTGGAG	450
13	MetHisArgHisGlyGlyValAsnGlnLeuGlyGlyValPheValAsnGlyArgProLeuProAspValValArgGlnArgIleValGlu	42
451	CTGGCCACAGGGTGTGGCCCTGTGACATCTCCCGCAGCTGCGGGTCCAGCATGGCTGTGTGACAAAATCTGGGCAGGTA	540
43	LeuAlaHisGlnGlyValArgProCysAspIleSerArgGlnLeuArgValSerHisGlyCysValSerLysIleLeuGlyArgTyrTyr	72
541	GAGACTGGCAGCATCAAGCCCGAGTGTGGTGGCTCCAAGCCAAAGTGGCAACGCCAAAGTGGTGGACAAGATTGCCGAATACAAG	630
73	GluThrGlySerIleLysProGlyValIleGlyGlySerLysProLysValAlaThrProLysValValAspLysIleAlaGluTyrLys	102
631	CGACAGAACCCGACTATGTTCCGCTGGGAGATCCGTGCACAGCTGTACCGGAGGGCATCTGCGATAATGACACAGTTCCAGTGTCTCA	720
103	ArgGlnAsnProThrMetPheAlaTrpGluIleArgAlaGlnLeuLeuArgGluGlyIleCysAspAsnAspThrValProSerValSer	132
721	TCCATCAACAGGATCATCCGGACAAAGTTCAGCAGCTTTCACCCAAGCGCGGATGGGGCAGGGACAGGAGTACTGCCCCGGCCAC	810
133	SerIleAsnArgIleIleArgThrLysValGlnGlnProPheHisProThrProAspGlyAlaGlyThrGlyValThrAlaProGlyHis	162
811	ACCATCGTTCCAGCACGGCTCCCTCTGTTCCAGCGCTCTAACGCCAGTGGGATCTACTCCATCAACGGGATCTGGGGATT	900
163	ThrIleValProSerThrAlaSerProProValSerSerAlaSerAsnAspProValGlySerTyrSerIleAsnGlyIleLeuGlyIle	192
901	CCTCGTCCAACGGTGGAGAGGAAACGCGAGGAAGTCGAGGTATACACTGATCTGCCCACATTAGAGGAGGTGGAGGTTTACATCTG	990
193	ProArgSerAsnGlyGluLysArgLysArgGluGluValGluValTyrThrAspProAlaHisIleArgGlyGlyGlyLeuHisLeu	222
991	GTCTGGACTTTAAGAGATGTGTCTGAGGGCTCTGTCCCTAATGGAGACTCCAGAGTGGTGTGGACAGTTTTCGGGAAGCACCTGCGAGCC	1080
223	ValTrpThrLeuArgAspValSerGluGlySerValProAsnGlyAspSerGlnSerGlyValAspSerLeuArgLysHisLeuArgAla	252
1081	GACACCTTACCCAGCAGCTGGAAGCTCTGGATCGAGTCTTTGAGCGTCTTCCATCCCGATGCTTCCAGGCATCAGAGCACATC	1170
253	AspThrPheThrGlnGlnLeuGluAlaLeuAspArgValPheGluArgProSerTyrProAspValPheGlnAlaSerGluHisIle	282
1171	AAATCAGAACAGGGAAATGAATACTCTCTCCAGCCCTGACCCTGGGCTTGATGAAGTCAAGTCCAGTCTATCTGCATCGGCCAACCT	1260
283	LysSerGluGlnGlyAsnGluTyrSerLeuProAlaLeuThrProGlyLeuAspGluValLysSerSerLeuSerAlaSerAlaAsnPro	312
1261	GAGCTGGCAGCAATGTGTGAGGCACACAGACGTACCCCGTGTGACCGGTCTGATATGACGAGCACCCTACTCTGGTTACCCCGG	1350
313	GluLeuGlySerAsnValSerGlyThrGlnThrTyrProValValThrGlyArgAspMetThrSerThrThrLeuProGlyTyrProPro	342
1351	CATTGCCCCCACTGGCCAGGGAAGCTACCCACCTCCACCTGGCAGGAATGGTGCCTGGGAGCGAGTTCTCAGGCAACCCATACAGC	1440
343	HisLeuProProThrGlyGlnGlySerTyrProThrSerThrLeuAlaGlyMetValProGlySerGluPheSerGlyAsnProTyrSer	372
1441	CATCCCCAGTACCCGCTACAATGAGGCTTGGAGATTGAGCAACCCCGCTTACTAAGTTCCCTTATTATTATAGTGCGCCCCCGG	1530
373	HisProGlnTyrThrAlaTyrAsnGluAlaTrpArgPheSerAsnProAlaLeuLeuSerSerProTyrTyrTyrSerAlaAlaProArg	402
1531	TCCGCCCTGCGCTCGTGCCTGCTATGACCGCCACTAGTTACCGCGGGGACCACATCAAGCTTCAGGCAGACAGCTTCGGCTCCA	1620
403	SerAlaProAlaAlaArgAlaAlaAlaTyrAspArgHis	415
1621	CATCGTCCCGTCTGACCGCCACCCAGAGGGAGGGAGGACCGAAGCGACGCCGTGCTCTCGGCCACCGTCCCTGTCACCCCCACC	1710
1711	TCGAGACATCTCCCCCAAGCCCGCAGCCCTTCTCATCACCTGTGATGGCCAGACAGGAGGAGCCGCGGGTGGGACCCCTCAG	1800
1801	GCCCCAGCCCCCGCTCGCCCCCTCCCGCCTGCTGGACCGCGCGCGCTCCGGAAGCCCGCCCGCTCTGGCCCGCGCCGTTCTG	1890
1891	TGACACACAATCAGCGTGGCCACCCAGGGCAGCCACTTCGGAGCCCGGGAGTCCGAGGCTTCGCTGGAGAAGCTGAGCAAGGGAAG	1980
1981	CTAAGGAAAGGACTTTGTGCTGAAGGAGGAGCCGCGGGGAGACCTAGGGGGGAAATCTGCCCAATCCCCCTGTGCTGCCAAGCTGA	2070
2071	CCCCCTTCTGTCTGCGCCCCCGCCAGACGGTGGCACTGGGACGCAAGGTCGGACTCGAGGGACAGCGTTGGGGCGACAGCGTTTT	2160
2161	CCAAGGCTGGTACCAAGGCGAGCAGAGCCAACTGAGTCTATCCACTGAAGTGCCTAGTTCCCTTGGGCTGCGATCTCCTTCTGCAG	2250
2251	TGACTCTCTCTCGGCCCGCTTGCCTTACGTCAGCGCTTTCACCTGCTGGCTCTCACTTCCCTCTGTGCTTCTCTCGGCTCA	2340
2341	TCCCACGACACTTGTGTTTGGTCCCAAGCAGCAGCCACGTCGGGGCCACCGGGGAACCGTGTCTCAGGCTGCGGCTGCGAGCATTG	2420
2421	ATCTTTTCCCACTAGTAATCGGCTGCCACCCCGCCCTCCCATCCCAACTCCCAACTGCGGGCGCAAAAAGCCTGGAAGGAAGTGC	2510
2511	ACCGAGCCCGCCAGATGAAAATCCGGTCTCCCGCCCCATTCCCCCAAGCACCCCTCCCGAGCTTCCCTTGGCAGGAGTGAAC	2600
2601	AGAGCTCAAAAAAAAAAAAAA	

Fig. 2. The DNA sequence and conceptual amino acid sequence of the *Pax2* gene. The paired domain is indicated with a solid underline. The paired octapeptide is indicated with a dashed underline. The 5' and 3' borders of the 69 nuc. deletion in clone c31A are indicated with triangles. The predicted amino acid sequence starts at nucleotide 325. Two additional potential translational start sites are located at nucleotides 331 and 361.

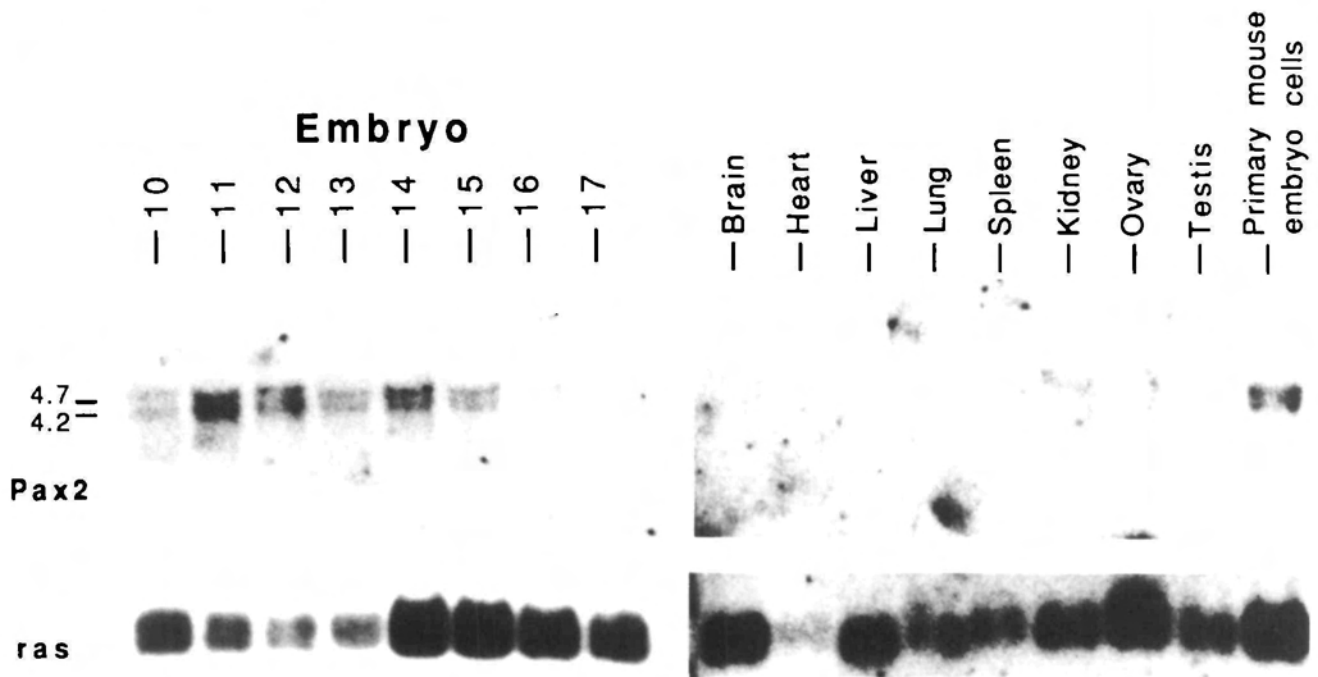


Fig. 3. Northern blot analysis of *Pax2* transcripts in embryos and adult tissues. Approximately 5 μ g of A+ RNA was hybridized with the *Pax2*-specific probe (indicated in Fig. 1) and a *c-ras* probe. For the *Pax2* probe, exposure times were 1 day for the embryonic RNAs and 7 days for the adult tissue RNAs.

frame of at least 415 amino acids that includes the paired box domain. The paired domain, as underlined in Fig. 2, begins with a histidine at amino acid position 14 and ends with a glutamine at position 144. This open reading frame continues until a stop codon is encountered at amino acid position 416. The most 5' methionine is encoded at nuc. 325, with a second methionine codon at nuc. 331 and a third methionine codon at nuc. 361. At present, it has not been determined which of these potential AUGs serves as a translational start site. In addition, because the extremely GC-rich 5' leader sequence does not contain any upstream stop codons, it cannot be ruled out that the *N*-terminus of the *Pax2* protein lies further upstream in a region not represented on the cDNAs.

Comparison of the *Pax2* paired domain to that of other paired domain proteins shows their high degree of homology. As evident in previous analyses of these sequences (Bopp *et al.* 1986; Deutsch *et al.* 1988; Burri *et al.* 1989), the paired domain consists of a more highly conserved *N*-terminal domain and less stringently conserved *C*-terminal domain. The first 74 amino acids of *Pax2* are most homologous to the mouse gene *Pax1* (Deutsch *et al.* 1988), the human gene HuP48 (Burri *et al.* 1989), and the *Drosophila* gene P29 (Burri *et al.* 1989). Overall, *Pax2* is slightly more homologous to *Pax1*, P29, and the human genes, with 69–72% amino acid identity, compared to the *Drosophila* *prd* and *gsb* genes, with 63–67% amino acid identity. A second homologous region among paired-domain proteins, termed the octapeptide, was recently noted by Burri *et al.* (1989). In *Pax2*, an octapeptide is found between amino acid positions 184–191. When compared to two

human and two *Drosophila* sequences, six out of the eight amino acids match a consensus sequence and the only two substitutions also appear in one of the *Drosophila* *gsb* genes. There may be at least two variant forms of the *Pax2* protein. The clone c31A has spliced out a small 69 nuc. exon, spanning nuc. 937–1005 and present in the clone cPX. Thus, the corresponding mRNA would encode a *Pax2* protein with a 23 amino acid deletion, as indicated in Fig. 2.

In a search for similarities to other known proteins, the NBRF protein sequence databank (Genetics Computer Group; Devereux *et al.* 1984) was screened with the complete *Pax2* sequence and the *Pax2* sequence minus the paired domain. The complete *Pax2* protein is 22% identical, 41% similar, to human *c-fos* (van Straaten *et al.* 1983) and 22% identical, 40% similar, to mouse *c-myc* (Stanton *et al.* 1984), as determined by the BestFit sequence alignment program (Genetics Computer Group). Interestingly, the regions of similarity between *Pax2* and *c-fos* and *Pax2* and *c-myc* are different. These major sequence similarities are downstream of the paired domain.

Transcripts of Pax2 are primarily restricted to embryogenesis

In order to determine the spatial and temporal expression analysis of *Pax2* in embryonic and adult tissues, Northern blotting and *in situ* hybridization was done with a *Pax2* specific probe downstream from the paired box sequence (see Fig. 1), to eliminate potential cross hybridization to other Pax genes. Initially, poly(A)⁺ selected RNA was isolated from embryos from day 10 to day 17 of gestation. Concurrently,

poly(A)⁺ RNAs from various adult tissues were prepared. As demonstrated in Fig. 3, transcripts corresponding to the *Pax2* gene can be detected during embryogenesis. Two transcripts are detected in embryonic poly(A)⁺ RNAs, approximately 4.2 kb and a 4.7 kb in size, that hybridize with a fragment from the *Pax2* cDNA clone. Using *c-ras* transcripts (Leon *et al.* 1987) for standardization, the relative peak of *Pax2* transcript accumulation occurs at approximately 11 days gestation and decreases thereafter. The two *Pax2* transcripts are expressed at near equimolar amounts in all positive samples tested. In adult tissues, *Pax2* transcripts could not be detected in the brain, heart, liver, lung, spleen, ovary or testis, even after prolonged exposure of Northern blots (Fig. 3). In kidney RNA, a faint smear is visible after a prolonged 7 day exposure, although this appears somewhat larger than the embryonic transcripts. In addition, *Pax2* mRNA could not be detected by *in situ* hybridization to adult kidneys or adult spinal chord (data not shown). Thus, *Pax2* gene expression is primarily restricted to the developing embryo.

The tissue-specific developmental expression pattern of *Pax2* was examined by *in situ* hybridization to sections from mouse embryos, isolated at various developmental stages. The sections were hybridized with ³⁵S-UTP-labeled RNA probes corresponding to the sense and antisense strands of a unique *Pax2* cDNA fragment. In addition, a mouse *Hox 3.1* probe was used as a positive control (Breier *et al.* 1988). In mesoderm-derived tissues, *Pax2* is expressed in the pronephric tubules and extending nephric duct beginning at day 9. There is no detectable expression in mesoderm cells prior to extension of the Wolffian duct and formation of the pronephros. Representative micrographs are shown in Fig. 4 beginning at 10 days gestation. In the developing excretory system, *Pax2* transcripts can be detected at 10 days p.c. in the nephric cord and Wolffian duct (Fig. 4A,B). At this time individual pronephric tubules express *Pax2* transcripts, as seen in Fig. 4C,D. These segmented embryonic excretory organs are derived from the intermediate mesoderm, lateral to the somites (for a description and review of pro-, meso- and metanephric organogenesis see Saxen, 1987). However, no *Pax2*-specific hybridization is seen in mesoderm prior to formation of the nephric cord and Wolffian duct (data not shown).

The developing mesonephros also expresses *Pax2* transcripts at 11 days gestation, as evident in Fig. 4E,F. At this time and more posterior to the mesonephros, the ureter has budded out from the Wolffian duct and has reached the metanephric mesenchyme to begin formation of the adult kidney. Part of the ureter is shown in Fig. 4G,H and it also hybridizes with the *Pax2* probe, whereas the surrounding mesenchyme cells do not. Fig. 4I,J show a section through the first branching of the ureter and the condensation of metanephric mesenchyme cells around the branches. Both the branched ureter and the mesenchymal condensations express high levels of *Pax2* transcripts.

At 12 days gestation, the ureter has branched repeat-

edly and sections through the metanephros reveal many early mesenchymal condensations around the developing collecting ducts. *Pax2* is expressed in the collecting ducts as well as in the condensed mesenchyme cells (Fig. 4K,L). It is of interest to note the absence of *Pax2* expression in non-condensed mesenchyme cells. This is demonstrated more clearly in Fig. 5. Cells of the metanephric mesenchyme that have not condensed around the branching collecting ducts do not appear to express *Pax2* (Fig. 5A–D). At later developmental stages, *Pax2* expression in the kidney declines (Fig. 5E–H). Although still actively expressed in the perimeter of the growing kidney at 14 and 17 days gestation, more mature renal tubules in the interior are expressing less *Pax2*-specific RNA. Thus, consistent with the Northern blotting data, expression of *Pax2* declines and can no longer be detected by *in situ* hybridization in adult kidneys (data not shown).

Discussion

This report describes the isolation and characterization of a novel paired-box-containing gene, *Pax2*, that is expressed during development. Using a mouse *Pax1* paired box probe (Deutsch *et al.* 1988), cDNA clones containing a related paired box and spanning 2623 bp of *Pax2* mRNA sequence were isolated. These two overlapping cDNAs do not represent the complete *Pax2* transcripts since Northern analysis reveals two mRNAs of approximately 4.2 and 4.7 kb in length. Whether the two *Pax2* transcripts are generated by different transcription initiation sites, polyadenylation sites, or by an alternative splicing mechanism remains to be determined. In addition, it is unlikely that these two transcripts represent a kidney-specific and a neural tube specific mRNA because the relative molarity of the two transcripts remains constant during gestation while maximum levels of kidney expression and neural tube expression (Nornes *et al.* 1990) do not coincide precisely. An alternatively spliced mRNA was found among the cDNAs described; however, the extra 69 nuc. cannot account for the size difference seen between the two *Pax2* transcripts. It is likely that multiple forms of the *Pax2* mRNAs exist that can encode different forms of the paired domain protein.

At least two proteins, of 415 and 392 amino acids, could be encoded if translation starts at the most 5' AUG of clone c31A. Because there are no stop codons 5' to this AUG, the possibility that translation starts at an upstream AUG not present in the clone c31A cannot be ruled out. However, several observations suggest either the first, second or third AUG, all located within a 36 nucleotide stretch, is the translation start site. The optimal sequence context for a eukaryotic translational start site, as determined by Kozak (1987a), is (GCC)GCC(A/G)CCAUGG. If a one nucleotide space is allowed for, 9/13 nucleotides flanking the first methionine are identical to this consensus sequence. However, the sequence surrounding this first methionine has a C at position -3, where normally A and G

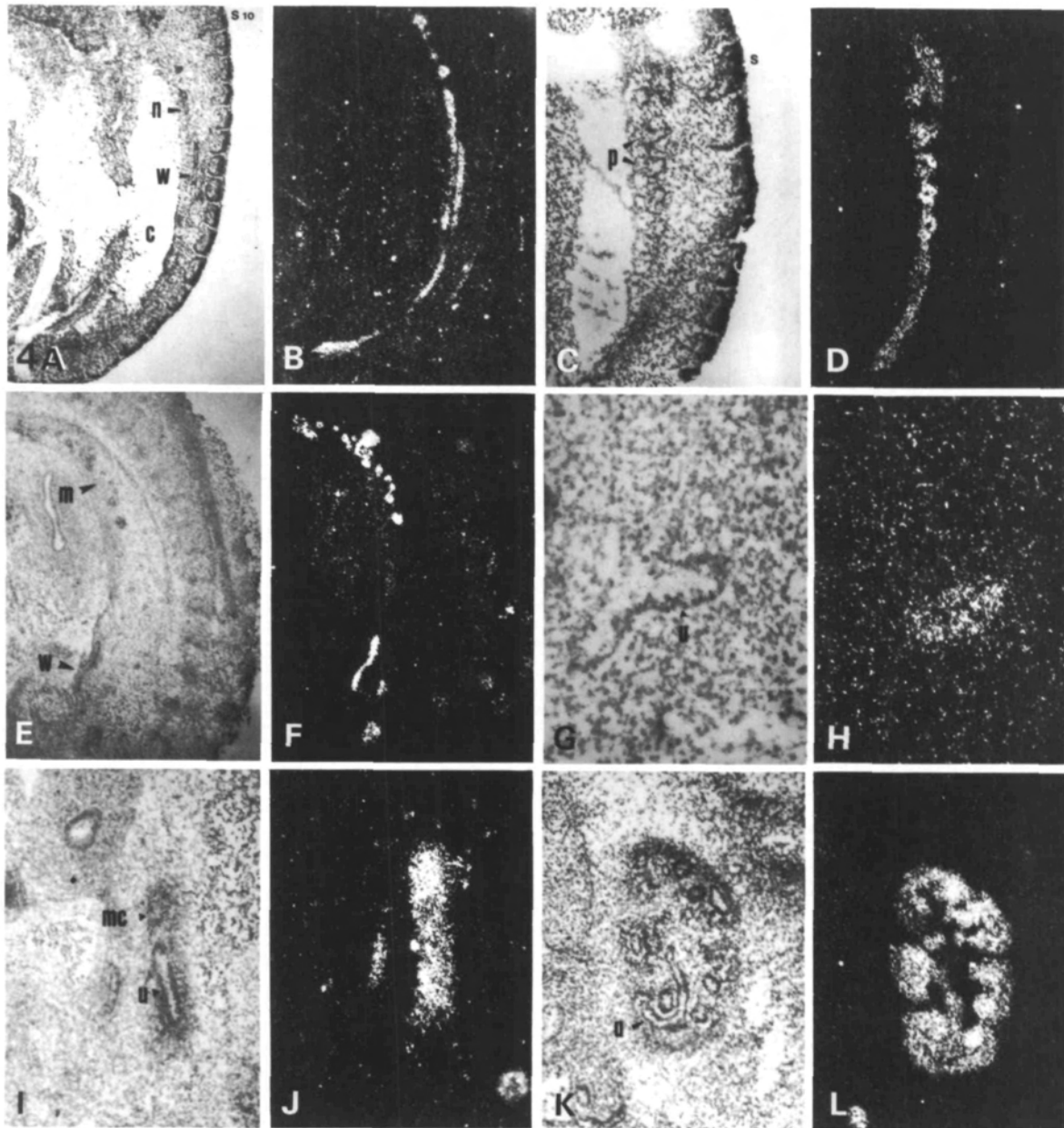


Fig. 4. Early expression of *Pax2* in the developing excretory system. (A) Parasagittal section of 10 days p.c. embryo from approximately the 10th to the 27th somite. Indicated are the coelom (c), the nephric cord (n) and the Wolffian duct (w), magnification 40 \times . (B) Dark-field image of A. Note the continuous expression of *Pax2* along the anterior-posterior axis. (C) Pronephric tubules (p) at 10 days p.c., magnification 100 \times . (D) Dark-field image of C. (E) Mesonephros (m) of 11 days p.c. embryo, the Wolffian duct (w) is also indicated, magnification 100 \times . (F) Dark-field image of E. (G) The ureter (u) budding from the Wolffian duct, magnification $\times 200$. (H) Dark-field image of G. (I) Section through the branched ureter from an 11.5 days p.c. embryo showing mesenchymal condensations (mc) around the ureter. (J) Dark-field image of I showing *Pax2* expression in the branched ureter and in mesenchymal condensations. (K) Section through the 12 days p.c. metanephros showing the multiply branched ureter (u), which forms collecting ducts, and surrounding mesenchymal condensations. (L) Dark-field image of K; again, note the high levels of *Pax2* expression in the branched ureter and in mesenchymal aggregates surrounding the ureter.

predominate. This can be a critical substitution (Kozak, 1986) and may inhibit translational initiation. The second and third methionine codons, located 2 and 13 amino acids downstream, respectively, both contain a G at position -3, but the remaining flanking sequences deviate more from the proposed consensus. Furthermore, the 5' leader sequence is extremely GC-rich,

approximately 73%; this has been noted for many naturally occurring long 5' leader sequences (Kozak, 1987b). In general, the paired domain is near the N-terminus, located 18-30 amino acids from the terminal methionine for the three *Drosophila* genes (Frigerio *et al.* 1986; Baumgartner *et al.* 1987) and three mouse genes (Deutsch and Gruss; Goulding and Gruss;

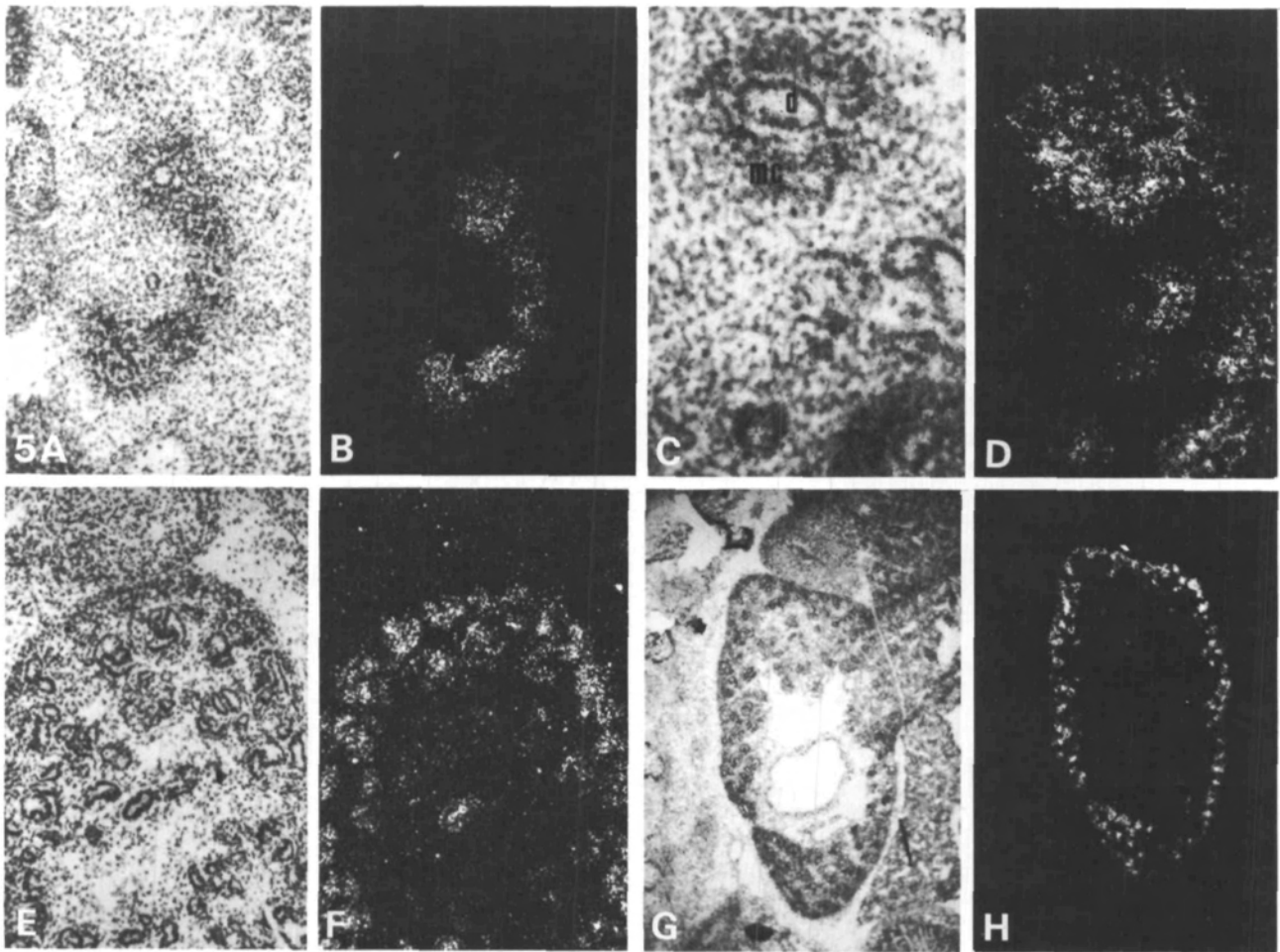


Fig. 5. Late expression of *Pax2* in the developing metanephros. (A) Metanephros at 12 days p.c., magnification $\times 100$. (B) Dark-field image of A, note the lack of expression in loose mesenchyme cells. (C) High magnification of 12 days p.c. metanephros showing a transverse section through a collecting duct (c) and surrounding mesenchymal condensations (mc), magnification $\times 200$. (D) Dark-field image of C. *Pax2* expression is restricted to collecting ducts and condensed mesenchyme. (E) Metanephros at 14 days p.c., magnification $\times 100$. (F) Dark-field image of E showing *Pax2* expression on the perimeter of the growing kidney. (G) Metanephros at 17 days p.c., magnification $\times 40$. (H) Dark-field image of G; again, note the absence of *Pax2* expression in the interior of the cortex.

Walther and Gruss, unpublished) sequenced to date. In addition, a new cDNA clone has recently been identified that extends more than 1000 bp downstream of the 3' end of clone cPX and contains an internal polyadenosine stretch (Dressler, unpublished observation). This would suggest that the *Pax2* mRNA contains a 3' untranslated sequence of at least 2 kb.

The conservation of the paired domain across divergent species suggests a similarity of biochemical function. In *Drosophila*, the paired-box-containing genes *prd*, *gsb-p* and *gsb-d* also contain a paired type homeobox (Frigerio *et al.* 1986; Baumgartner *et al.* 1987). However, the homology of *Pax2* to other paired domain proteins is restricted to the paired domain and the paired octapeptide, as *Pax2* does not contain a paired-type homeobox. Presently, the function of the paired domain is not clear. Computer analysis reveals a potential helix-turn-helix motif in the C-terminal end of the paired-box domain. However, there is no direct evidence that this domain has DNA-binding activity or

that the Pax genes encode nuclear proteins. A computer search also found some similarity to *c-fos* and *c-myc*, approximately 21% identical and 40% similar. Presently, it is not clear if this similarity is significant. Previously, the *Pax2* locus was mapped to mouse chromosome 7 (Dressler *et al.* 1988); however, fine mapping of the locus is required before any known mutations near *Pax2* can be identified.

The expression of Pax2 correlates with distinct morphological changes

Although the development of the metanephros is a complex process, individual aspects of induction and cellular interactions have been studied in great detail. At 11 days gestation, development of the kidney begins when the ureteric bud, an outgrowth from the posterior Wolffian duct, meets the metanephric blastema, a loosely organized mesenchymal tissue in the posterior nephric cord. Shortly thereafter, a reciprocal induction event occurs. The ureteric bud is induced to branch and

the metanephric mesenchyme is induced to form epithelial condensations around the branching ureter. These mesenchymal condensates will eventually form the S-shaped body and subsequently join the collecting duct system. It is not surprising that the ureter expresses the *Pax2* gene, since the ureter is an outgrowth of the Wolffian duct. However, it is of interest to note that high levels of *Pax2* transcripts can be detected in the early condensations around the budding ureter, whereas the more loosely organized mesenchyme cells do not appear to express *Pax2*. Thus, the expression of *Pax2* correlates with induction and formation of early metanephric epithelial condensations. At later developmental stages (day 17), *Pax2* expression declines in the more developed glomerular crevices but continues to be expressed near the perimeter of the growing kidney. Because the kidney continues to grow after birth, the faint *Pax2* hybridization seen with adult kidney RNA may reflect a very low amount of residual transcripts.

The induction of the mesenchymal blastema to form early epithelial condensations requires cell contact with cytoplasmic processes of the inducing tissues (Grobstein, 1957; Saxen and Lehtonen, 1978). Using an *in vitro* organ culture model, it has been shown that antibodies against the gangliosides G_{D3} (Sariola *et al.* 1988) or the laminin A chain (Klein *et al.* 1988) disrupt metanephric epithelium formation, indicating that these molecules are involved in the interaction between inducing and responding cells. Changes in gene expression after formation of the early epithelium include the induction of the cell adhesion molecule uvomorulin (Vestweber *et al.* 1985) as well as the basement membrane glycoprotein laminin (Leivo *et al.* 1980; Ekblom *et al.* 1980). Interstitial proteins expressed in the mesenchyme that are not found in metanephric epithelium after induction include collagen type I and III (Ekblom *et al.* 1981). However, the transduction of signals received at the cell surface and the subsequent regulation of gene expression in responding cells are mechanisms that remain to be characterized. The localization of *Pax2* transcripts in inducing ureteric epithelial cells, the subsequent early activation of *Pax2* transcripts in responding mesenchyme cells and the transient nature of *Pax2* transcript accumulation may indicate that this gene functions in the transduction pathway during cell polarization and/or cell condensation.

In contrast to the *Pax1* and *Pax2* genes, many murine homeobox (*Hox*) genes are expressed in a region-specific, rather than tissue-specific, manner along the rostrocaudal axis (for review see Holland and Hogan, 1988), supporting the concept of positional specification through the combinatorial effect of *Hox* gene expression. The *Pax* genes may be involved in more specific aspects of morphogenesis, such as the migration and differentiation of sclerotome subsets during vertebral column development (*Pax1*) or metanephric mesenchyme condensation and cell polarization (*Pax2*). As discussed in detail in the accompanying report (Nornes *et al.* 1990), *Pax2* is also expressed during neurogenesis in the hindbrain and spinal cord, the otic vesicle and the optic cup. This biphasic expression

pattern in mesoderm- and ectoderm-derived tissues may indicate multiple functions of the *Pax2* proteins. In fact, many *Drosophila* segmentation genes have neurogenic functions that are evident later during embryogenesis. Expression of the mouse gene *Pax1*, however, is limited to segmented mesodermal structures and cannot be detected in nervous tissue. Thus, the neurogenic aspect of *Pax* gene expression does not appear to be a general principle. In any event, the two *Pax* genes characterized to date are members of an entirely new gene family whose expression pattern is distinctly different from that of murine homeobox genes and whose function may involve the regulation of specific morphogenetic events.

We thank Sabine Geisendorf for expert technical assistance, D. Plachov for rescreening, H. Jäckle and S. Côté for the *Drosophila* *gooseberry* clones, A. Joyner and G. Martin for the cDNA library and R. Altschäffel for excellent photography. G. R. D. is a recipient of an Alexander von Humboldt fellowship.

References

- AKAM, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**, 1–22.
- BALLING, R., DEUTSCH, U. AND GRUSS, P. (1988). *undulated*, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of *Pax1*. *Cell* **55**, 531–535.
- BAUMGARTNER, S., BOPP, D., BURRI, M. AND NOLL, M. (1987). Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during embryogenesis. *Genes and Dev.* **1**, 1247–1267.
- BOPP, D., BURRI, M., BAUMGARTNER, S., FRIGERIO, G. AND NOLL, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*. *Cell* **47**, 1033–1040.
- BREIER, G., DRESSLER, G. R. AND GRUSS, P. (1988). Primary structure and developmental expression pattern of *Hox 3.1*, a member of the murine *Hox 3* homeobox gene cluster. *EMBO J.* **7**, 1329–1336.
- BURRI, M., TROMVOUKIS, Y., BOPP, D., FRIGERIO, G. AND NOLL, M. (1989). Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO J.* **8**, 1183–1190.
- CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J. AND RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonucleases. *Biochemistry* **18**, 5294–5299.
- DEUTSCH, U., DRESSLER, G. R. AND GRUSS, P. (1988). *Pax1*, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* **53**, 617–625.
- DEVEREUX, J., HAEBERLI, P. AND SMITHIES, O. (1984). A comparison set of sequence analysis programs for the VAX. *Nucl. Acid Res.* **12**, 387–395.
- DONY, C. AND GRUSS, P. (1987). Specific expression of the *Hox 1.3* homeo box gene in murine structures originating from or induced by the mesoderm. *EMBO J.* **6**, 2965–2975.
- DRESSLER, G. AND GRUSS, P. (1988). Do multigene families regulate vertebrate development? *Trends Genet.* **4**, 214–219.
- DRESSLER, G. R., DEUTSCH, U., BALLING, R., SIMON, D., GUENET, J. L. AND GRUSS, P. (1988). Murine genes with homology to *Drosophila* segmentation genes. *Development* **104** suppl. 181–186.
- EKBLUM, P., ALITALO, K., VAHERI, A., TIMPL, R. AND SAXEN, L. (1980). Induction of a basement membrane glycoprotein in embryonic kidney: possible role of laminin in morphogenesis. *Proc. natn. Acad. Sci. U.S.A.* **77**, 485–489.
- EKBLUM, P., LEHTONEN, E., SAXEN, L. AND TIMPL, R. (1981). Shift

- in collagen types as an early response to induction of the metanephric mesenchyme. *J. Cell Biol.* **89**, 276–283.
- FAHRNER, K., HOGAN, B. L. M. AND FLAVELL, R. A. (1987). Transcription of H-2 and Qa genes in embryonic and adult mice. *EMBO J.* **6**, 1265–1271.
- FEINBERG, A. P. AND VOGELSTEIN, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.
- FITZGERALD, M. AND SHENK, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**, 251–260.
- FRIGERIO, G., BURRI, M., BOPP, D., BAUMGARTNER, S. AND NOLL, M. (1986). Structure of the segmentation gene *paired* and the *Drosophila* PRD gene set as part of a gene network. *Cell* **47**, 735–746.
- GROBSTEIN, C. (1957). Some transmission characteristics of the tubule-inducing influence on mouse metanephrogenic mesenchyme. *Exptl Cell Res.* **13**, 575–587.
- GRÜNEBERG, H. (1954). Genetical studies on the skeleton of the mouse. XII. The development of *undulated*. *J. Genet.* **52**, 441–455.
- HOGAN, B., CONSTANTINI, F. AND LACY, E. (1986). *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Press, pp228–235.
- HOLLAND, P. W. H. AND HOGAN, B. L. M. (1988). Expression of homeo box genes during mouse development: a review. *Genes & Dev.* **2**, 773–782.
- INGHAM, P., HOWARD, K. AND ISH-HOROWITZ, D. (1985). Transcription pattern of the *Drosophila* segmentation gene *hairy*. *Nature* **318**, 439–445.
- INGHAM, P. W. (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* **335**, 25–34.
- JOYNER, A. L. AND MARTIN, G. R. (1987). *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes & Dev.* **1**, 29–38.
- KLEIN, G., LANGEGGER, M., TIMPL, R. AND EKBLUM, P. (1988). Role of laminin A chain in the development of epithelial cell polarity. *Cell* **55**, 331–341.
- KOZAK, M. (1986). Point mutations define a sequence flanking AUG initiator codons that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.
- KOZAK, M. (1987a). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. molec. Biol.* **196**, 947–950.
- KOZAK, M. (1987b). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nuc. Acids Res.* **15**, 8125–8146.
- LEIVO, I., VAHERI, A., TIMPL, R. AND WARTIOVAARA, J. (1980). Appearance and distribution of collagens and laminin in the early mouse embryo. *Devl Biol.* **76**, 100–114.
- LEON, J., GUERRERO, I. AND PELLICER, A. (1987). Differential expression of the ras gene family in mice. *Mol. Cell. Biol.* **7**, 1535–1540.
- MAXAM, A. M. AND GILBERT, W. (1980). Sequencing end labeled DNA with base specific chemical cleavages. *Meth. Enzymol.* **65**, 499–560.
- NORNES, H. O., DRESSLER, G. R., KNAPIK, E. W., DEUTSCH, U. AND GRUSS, P. (1990). Spatially and temporally restricted expression of *Pax2* during murine neurogenesis. *Development* **109**, 797–809.
- PROUDFOOT, N. AND BROWNLEE, G. G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211–214.
- SANGER, F., NICKLEN, S. AND COULSON, A. R. (1977). DNA sequencing with chain termination inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.
- SARIOLA, H., AUFDERHEIDE, E., BERNHARD, H., HENKE-FAHLE, S., DIPPOLD, W. AND EKBLUM, P. (1988). Antibodies to cell surface ganglioside G_{D3} perturb inductive epithelial-mesenchymal interactions. *Cell* **54**, 235–245.
- SAXEN, L. (1987). Organogenesis of the Kidney, *Developmental and Cell Biology Series* 19 (ed. P. W. Barlow, P. B. Green and C. C. White) Cambridge Univ. Press.
- SAXEN, L. AND LEHTONEN, E. (1978). Transfilter induction of kidney tubules as a function of the extent and duration of intercellular contacts. *J. Embryol. exp. Morph.* **47**, 97–109.
- SCOTT, M. P. AND CARROLL, S. B. (1987). The segmentation and homeotic gene network in early *Drosophila* development. *Cell* **51**, 689–698.
- STANTON, L. W., FAHRLANDER, P. D., TESSER, P. M. AND MARCU, K. B. (1984). Nucleotide sequence comparison of normal and translocated murine *c-myc* genes. *Nature* **310**, 423–425.
- VAN STRAATEN, F., MÜLLER, R., CURRAN, T., VAN BEVEREN, C. AND VERMA, I. M. (1983). Complete nucleotide sequence of a human *c-onc* gene: Deduced amino acid sequence of the human *c-fos* protein. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183–3187.
- VESTWEBER, D., KEMMLER, R. AND EKBLUM, P. (1985). Cell adhesion molecule uvomorulin during kidney development. *Devl Biol.* **112**, 213–221.
- WRIGHT, C. V. E., CHO, K. W. Y., OLIVER, G. AND DE ROBERTIS, E. M. (1989). Vertebrate homeodomain proteins: families of region-specific transcription factors. *TIBS* **14**, 52–56.

(Accepted 20 April 1990)