Volume 15 Number 17 1987 Nucleic Acids Research

A mouse homeobox containing gene on chromosome 11: sequence and tissue-specific expression

Frits Meijlink, Rozalia de Laaf, Peter Verrijzer, Olivier Destrée, Vera Kroezen[1], John Hilkens[1] and Jacqueline Deschamps

Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht and [1]Department of Tumor Biology, The Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received July 23, 1987; Accepted August 13, 1987 Accession no. Y00436

ABSTRACT

We have molecularly cloned a mouse homeobox containing gene by isolating cDNA and genomic clones. The gene is located in a previously described cluster on chromosome 11 (Hart et al. (1985) Cell 43, 9–18) and was identified as the Hox2.3 gene. We present the complete mRNA sequence of this gene and describe similarities to other homeobox containing genes, among which its human homologue, the c1 gene. High expression of the Hox2.3 gene was found in kidney, testis, and spinal cord of adult mice, in the spinal cord of 12.5–17.5 day embryos and in differentiating EC cells depending on their treatment. Three different treatments of the pluripotent EC cell line P19, each leading to the induction of a specific differentiation pathway, resulted in all cases in induction of Hox2.3; however, major quantitative differences in this response were observed.

INTRODUCTION

In Drosophila melanogaster, a number of genes have been linked to specific functions fundamental and essential in pattern formation. A subset of these genes was shown to contain a highly conserved, 180-bp stretch of DNA, which was termed homoeobox(1). The subsequent discovery that the homeobox is also present in certain previously unknown genes of a broad range of invertebrate and vertebrate animals (see for instance Ref. 2) has generated a great deal of interest in these genes as candidates to play an important role in development. Several research groups have molecularly cloned homeobox-containing genes, in particular from the mouse (3–12), the usual model system for mammalian development. On the basis of sequence comparison with the Drosophila genes, two classes of mouse homeoboxes can be distinghuished: the Antennapedia (Antp)-type (3–11) and the Engrailed -type (12). The Antp- type genes have been found to be present in clusters of up to six genes in at least three positions in the mouse genome (3,6,8,10,13,14). The molecular function of the homeo domain is most likely to involve site-specific binding to double stranded DNA, as is evident from a combination of theoretical considerations (15–17) and experimental data (18,19). Apparently, the homeo domain containing proteins are members of a large family of such DNA binding proteins, characterized by the presence of a so-called helix-turn-helix structure directly involved in the binding to DNA (17).

In view of their high conservation at the protein level, mammalian homeodomain containing proteins are very likely to have at the molecular level the same function as their Drosophila cognates. It remains, however, unclear to what extent the implications of this function for the pattern formation of the organism are similar.

Eventually, the linkage of a natural mutation in a homeobox gene with a particular phenotype or the production of cell lines or transgenic mice in which the expression of a homeobox gene is altered (see Ref. 20) should demonstrate the biological function of these genes. Prerequisite to such studies is the detailed characte-

rization of the genes. This involves their genomic organization and structure and their sequence. Such studies, as well as descriptive studies on the expression patterns of the homeobox genes are also hoped to yield clues to the nature of their function and the way it is exerted. By comparison of these mammalian genes with their Drosophila analogues, the involvement in development of which is often well-documented, the hypothesis that mammalian and insect homeobox genes would be equally important for pattern formation , can be tested.

Several investigators have described the activity of homeobox containing genes in mouse and human embryos. High levels of spatially restricted expression were found in the central nervous system of embryos in the second half of gestation (3,5,6,8,10,21-26). A drawback of the mouse system, the technical difficulty to obtain sufficient amounts of material from (very) early stages of development is in part compensated by the availability of embryonic stem (ES) cells and teratocarcinoma or embryonal carcinoma (EC) cells. It has been shown that homeobox containing genes which are usually silent in EC cells, can be activated when the cells are induced to differentiate (3,8,27). We showed recently (28) that this activation is highly dependent on the way by which differentiation is induced: the use of chemical inducers such as retinoic acid is responsible for induction of very high levels of homeobox transcripts in the differentiating EC cells.

We now report the molecular cloning of a mouse homeobox containing gene, its complete protein coding sequence, and chromosomal localization. We identified the gene as the previously decribed Hox2.3 gene. The gene is strongly expressed in a tissue-specific fashion in adults, in a restricted region of the central nervous system of embryos and can be activated in EC cells. We confirm and extend our previous reports on homeobox expression in EC cells by showing that the level of Hox2.3 expression is strongly dependent on the presence of a chemical inducer like RA. However, we now report a dramatic difference in the magnitude of Hox2.3 activation in the same cell line induced to differentiate into neural or endodermal cell types respectively, according to methods both involving RA.

MATERIALS AND METHODS.

Nucleic acid procedures.

Cloning and analysis of DNA was according to standard procedures (29). The pGem Blue plasmid vector DNA was purchased from Promega Biotec (Promega Holland, Leiden, The Netherlands). Sequence analysis was done by the dideoxy chain termination method of Sanger (30), using double stranded supercoiled plasmid DNA (31). Computer analysis of sequences was performed using the 'MicroGenie' software (Scisoft Inc.) on a IBM PC-XT.

RNA isolation and analysis. Total cellular RNA was purified according to Chirgwin et al (32). Poly(A)$^+$ RNA was obtained by oligo (dT)-cellulose affinity chromatography. Total RNA from tissue samples was obtained by homogeneization of the tissue in guanidinium thiocyanate, low speed centrifugation to remove the cell debris and ultracentrifugation through a cushion of 5.7 M cesium chloride. Poly(A)$^+$ RNA (10 μg/lane) or total RNA (30 μg/lane) was fractionated by electrophoresis in 1.2 % agarose slab gels containing formaldehyde (6.6% v/v). RNA was transferred to nitrocellulose filters (Schleicher & Schuell) essentially as described by Thomas (33). Filters were hybridized with DNA probes labeled to a specific activity of at least 10^9 dpm/μg using the Amersham random priming kit (Amersham) (34). The Mubl probe normally used in Northern blot experiments was the purified insert from the cDNA clone pR1.2. Hybridization was as described (28). Final washing was in 0.2 x SSC (1x SSC, standard saline citrate = 150 mM NaCl, 15 mM Sodium Citrate, pH 7.0). 0.1% Sodium Dodecyl Sulphate at 65°C. After exposure of the filters and before hybridization with another probe, the probe was stripped by boiling the filters for 10 minutes in 0.2 x SSC .

Chromosome localization.

The panel of mouse/chinese hamster cells that we used has been described in Refs. 35-37. The presence of mouse chromosomal DNA containing the Mubl gene was established by Southern Blot analysis in each hybrid. The results were then compared to the data previously obtained with the following 39 markers: peptidase-1,2,3,4 and 7 (PEP-

1,2,3,4 and 7), sorbitol dehydrogenase-1 (SDH-1), adenylate kinase-1 (AK-1), 6-phosphogluconate dehydrogenase (PGD), phosphoglucomutase-2 (PGM-2), enolase-1 (ENO-1), triosephosphate isomerase-1 (TPI), glucose-6-phosphate isomerase (GPI), lactate dehydrogenase (LDH-1), glutathione reductase-1 (GR-1), adenine phosphoribosyltransferase (APRT), malate dehydrogenase (MOD-1), mannose phosphate isomerase (MPI), pyruvate kinase-3 (PK-3), hexokinase-1 (HK-1), galactokinase (GALK), transcobalamin (TCN-2), acid phosphatase-1 (ACP-1), arylsulfatase-A and BB (ARS-A and B), superoxide dismutase-1,2 (SOD1,2), glyoxylase-1 (GLO-1), complement factor 3 (C3), glutamate transaminase-1 (GOT-1), purine-nucleoside phosphorylase (NP1), esterase-10 (ES10), interferon-A (IF-A), the protooncogenes N-ras (RASN), c-myc (MYC), int-1 (INT1), Pim 1 and 2, and two Pim-1 related loci (LOC2 and 3; see ref. 36).

Tissue culture procedures.

PCC7 EC cells (38) and PSA-1 EC cells (39) were grown and induced to differentiate as described previously (28). P19 EC cells (40) were differentiated into a population of neural and glial cells or into a population of endodermal cells as described in Ref 41. Differentiation into mesoderm derivatives among which cardiac and skelettal muscle cells was according to Ref 42. Retinoic acid (RA) (Sigma) was used at a concentration of 1 μM in the cell culture medium. Stock solutions (10 mM) were prepared in dimethylsulfoxide and stored at -70^{o}C.

Experimental animals.

Mouse embryos were from a closed non-inbred Swiss mouse colony. Gestation was assumed to have begun at midnight before the day on which the vaginal plug was detected. 12.5-day embryos were dissected into brain, spinal cord, the cranio-facial part of the head, the limb buds, the viscera (gut, heart, lung, liver, kidney and rudiments of genital organs), and the carcass (defined as what remained after the above-mentioned tissue fractions had been removed). Tissue samples were immediately frozen on dry ice before being homogenized in guanidinium thiocyanate.

RESULTS AND DISCUSSION

Isolation of genomic clones.

A genomic DNA library prepared from a mouse plasmacytoma cell line was screened with a probe specific for the Drosophila melanogaster segmentation gene fushi tarazu (16), containing a homeobox. Screening of approximately 500,000 plaques resulted in the isolation of one phage that upon rescreening remained positive. DNA from this phage (named L13) was analyzed according to standard procedures and was found to contain only a portion of a homeobox containing gene, tentatively named Mubl. Not more than 44 nucleotides from a homeo box are present in this clone, apparently sufficient for detection by the ftz probe. A 531-bp Sau3A I fragment containing this homeobox stretch and 487 bp of sequences downstream from the homeo box was subcloned in pGem-4 and subjected to sequence analysis. This subclone as well as other Mubl clones were used for screening of a mouse liver genomic library (T. van Agthoven and G.C. Grosveld, Erasmus University, Rotterdam, unpublished). This resulted in the isolation of a series of clones containing one or more homeoboxes. At least two of these clones, L2 and L23, were shown to be derived from the same gene as clone L13 and to completely overlap the Mubl gene. The other clones contain portions of other homeo box genes and were apparently picked up due to the presence of homeo box sequences in the probes used. Two of the genes present in these clones were tentatively identified as the previously described Hox1-1 (3) and Hox2-1 genes (5), using probes for these genes supplied by their isolators.

CDNA cloning.

Complementary DNA clones were isolated from two different lambda gt11-based libraries. Several clones were isolated from a mouse testis cDNA library (D. Meyer and G.C. Grosveld, unpublished), the largest of which, LR12, was used for further analysis. One Mubl clone, named LR0.8 was a gift of A. van de Kamp and B. Hooft van Iddekinge. It had been isolated by them in our laboratory from a library (S. van Genesen, B. Hooft van Iddekinge, unpublished) prepared with RNA from the fibroblast-like cell

Table 1. Correlation between the presence of the Mubl gene and chromosome markers in the mouse/chinese hamster hybrid cells.

Chromosome marker	Chromosome	Presence of Mubl/presence of chromosome marker				
		+/+	-/-	+/-	-/+	% concordancy
PEP3	1	1	7	0	6	57.1
SDH1	2	1	3	0	10	28.5
AK1	2	1	3	0	10	28.5
RASN	3	0	3	0	6	33.3
PGD	4	1	6	0	7	50.0
PGM2	4	1	6	0	7	50.0
ENO1	4	1	6	0	7	50.0
IF-A	4	0	4	0	8	50.0
PEP7	5	1	5	0	8	42.8
TPI1	6	1	6	0	7	50.0
LOC2	6	0	2	0	4	33.3
GPI1	7	1	3	0	10	28.5
LDH1	7	1	3	0	10	28.5
PEP4	7	1	3	0	8	33.3
GR1	8	1	9	0	4	71.4
APRT	8	1	7	0	4	66.6
MOD1	9	1	6	0	7	50.0
MPI1	9	1	6	0	7	50.0
PK3	9	1	6	0	7	50.0
PEP2	10	1	5	0	8	42.8
HK1	10	1	5	0	8	42.8
GALK	11	1	7	0	1	88.8
TCN2	11	0	7	0	1	87.5
ACP1	12	1	1	0	11	15.3
ARS B	13	1	4	0	8	38.4
NP1	14	1	9	0	4	71.4
ES10	14	1	9	0	4	71.4
ARS A	15	1	1	0	12	14.2
MYC	15	1	2	0	5	37.5
INT1	15	1	2	0	2	60.0
SOD1	16	1	6	0	7	50.0
LOC3	16	0	2	0	4	33.3
GLO1	17	1	1	0	8	20.0
SOD2	17	1	3	0	10	28.5
C3	17	0	4	0	6	40.0
PIM-1	17	0	4	0	8	33.3
PIM-2	17	0	4	0	8	33.3
PEP1	18	0	3	0	8	27.2
GOT1	19	1	4	0	9	35.7

line Fib9, a cloned derivative from the C1003 EC line (A. Piersma and A. Willemse, unpublished). Several additional clones from the mouse testis library hybridizing to homeobox probes were isolated and identified by sequence analysis as Hox1–4 (11) clones. This gene is known to be highly expressed in testis (11,22,43). The inserts from the Mubl clones LR8 and LR12, EcoRI fragments of appr. 0.8 and 1.2 kb respectively, were subcloned in pGem–4 or pGem Blue respectively (clones pR0.8 and pR1.2).

Chromosomal localization.

To determine the chromosomal localization of the Mubl gene we used a panel of mouse x chinese hamster somatic cell hybrids(35–37). These cells have retained the complete set of chinese hamster chromosomes and segregated the mouse chromosomes. High molecular weight DNA from these cells was restricted with TaqI and analyzed by Southern analysis using the entire L2 insert as a probe. Upon stringent washing of the blot, three bands of ca. 1, 1.1 and 5 kilo base pair (kbp) (not shown) that are present in mouse control DNA of various sources and in a few of the hybrids, but never in chinese hamster DNA, were assigned 'diagnostic mouse bands'. Paired comparison of the presence or absence of known chromosomal markers and the Mubl gene resulted in the concordancies shown in TableI. The highest correlation was found with the two chromosome 11 markers GALK and TCN2 making it very likely that Mubl is located on this chromosome.

Restriction mapping.

Using the genomic clones as well as the cDNA clones, we obtained the restriction map displayed in Fig. 1. The region on the map containing sequences hybridizing to the Mubl cDNA spans a 3.5–kb area of genomic DNA. The cDNA sequence is, as a combination of restriction and sequence data (not shown) shows, interrupted by a single intron, with an approximate length of 2.2 kb. We have not yet identified the site at which

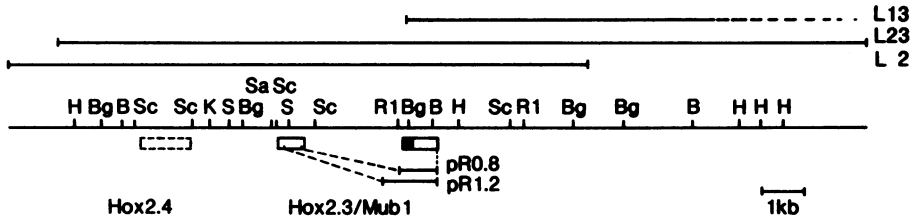


Fig. 1. Genomic restriction map of the Mub1 region. Open boxes, exon sequences from the Mub1/Hox2.3 gene. Black box, Hox2.3/Mub1 homeobox. Stippled box, approximate location of the Hox2.4 homeobox. The position of the inserts of the genomic clones L2, L13, L23, is indicated above the restriction map; the position of the cDNA clones, pR0.8 and pR1.2 below. H, HindIII; Bg, BglII; B, BamHI; Sc, SacI; K, KpnI; S, SmaI; Sa, SalI; RI, EcoRI restriction sites.

transcription of the gene initiates, and we can therefore not exclude that there are introns located in the 5' untranslated region of the gene. As indicated in the map, upstream from the gene we observe a second region that shows hybridization with different homeobox probes. This gene is currently being analyzed in more detail.

Identification of the gene.

Comparison of the restriction map (Fig.1) of the genomic region of the Mub1 gene with previously published genomic maps of mouse homeobox genes reveals that it is completely compatible with a part of the map of the Hox2 cluster as determined by Hart et al.(6). This indicates that Mub1 is identical to the Hox2.3 gene and is in agreement with our localization of Mub1 on chromosome 11, since the Hox2 cluster was previously located on mouse chromosome 11, band D (7,44,45). Unfortunately, no sequence data for Hox2.3 have been published. However, the human c1 gene which was mapped to a position corresponding to Hox2.3 was recently sequenced (46). As we will show below, the sequence of Hox2.3 is extremely similar to that of the c1 gene. Finally, the length

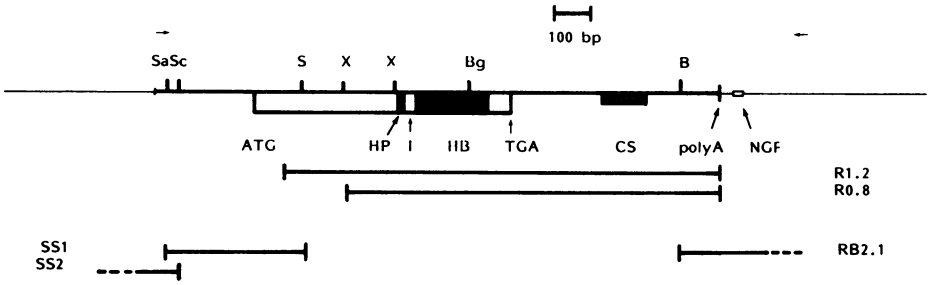


Fig. 2. Restriction map of the sequenced portion of the Mub1/Hox2.3 gene. Restriction sites are abbreviated as in Fig. 1; X= XhoII. Bold line, mRNA sequence, thin line, genomic DNA. The protein-coding region is indicated by the large box, the closed portions of it representing the homeobox (HB) and the conserved hexapeptide (HP), respectively. I, location of intron; CS (shaded box), location of highly conserved region (as compared to the human c1 gene) in the 3' untranslated region; NGF, location of sequence similarity to the human β-NGF gene. R1.2 and R0.8 are cDNA clones used in the sequence analysis; SS1, SS2, and RB2.1 are subcloned restriction fragments from the genomic clones used in sequence analysis. The small horizontal arrows indicate the portion of the DNA which has been sequenced.

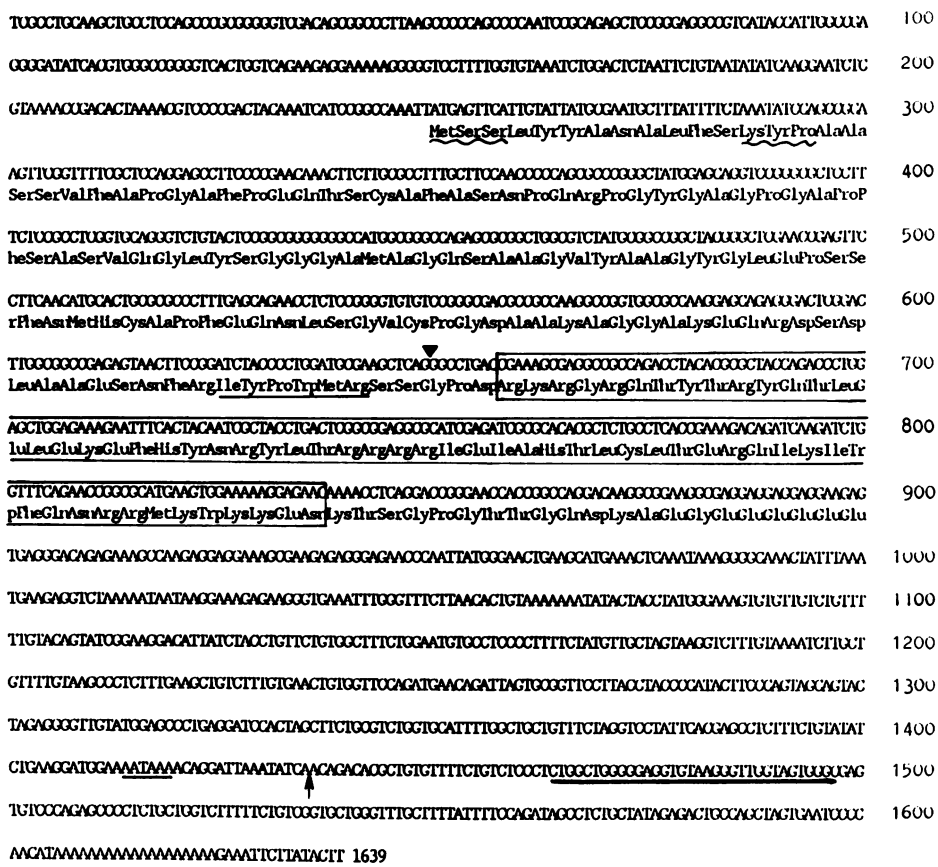


Fig. 3. Sequence of the Mubl/Hox2.3 mRNA and its immediate 3′ flanking region. These sequence data are derived from the cDNA clones pR0.8 and pR1.2, and from the genomic clones L13 and L2 (See Fig. 2). The longest open reading frame is indicated by its conceptual translation. The homeobox is indicated by a box, the conserved hexapeptide is underlined. Wavy lines: protein similarity to a number of other homeodomain containing proteinsamino terminal region of . The polyadenylylation signal at nucleotide 1414 is underlined. An arrow indicates the polyadenylylation site (position 1435). Sequences downstream from nucleotide 1435 were found in the genomic clone L13; sequences upstream from nucleotide 325 were found in clone L2. It is not yet known at which site the transcription starts. The underlined sequence at positions 1464–1497 displays similarity with the human β-NGF gene. A black triangel indicates the position of the intron. No intron sequences are shown.

of the mRNA encoded by the Mubl gene is of the same size class as the Hox2.3 transcript (6; and see below).

We conclude, that the Mubl gene is identical to the Hox 2.3 gene, and that the region upstream showing homeobox hybridization belongs to the Hox2.4 gene (6).

Sequence analysis.

The mRNA-encoding sequence of the Mubl gene was determined as follows (see Fig. 2):

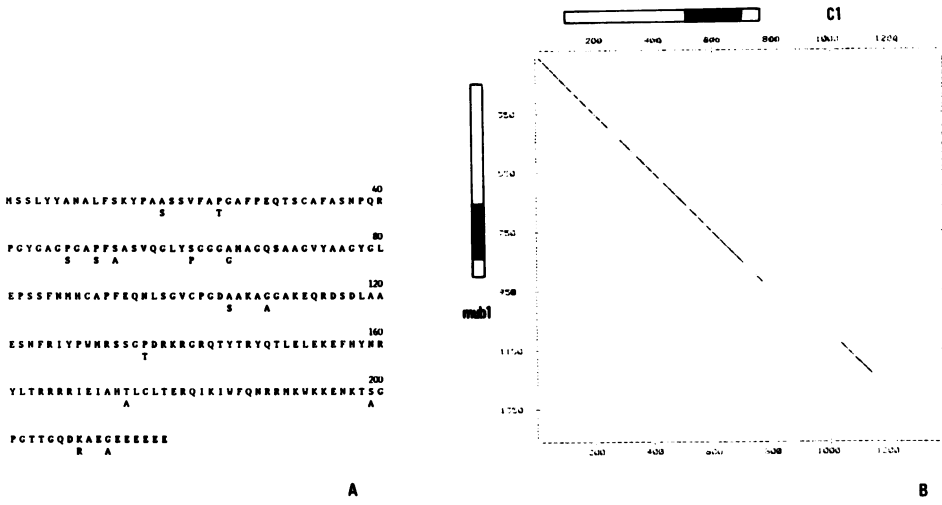


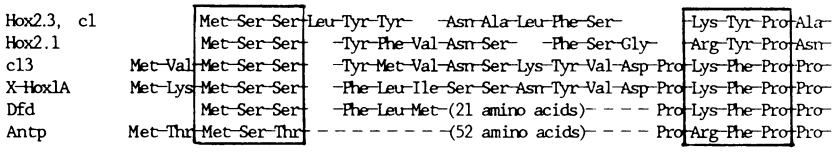
Fig. 4. Comparison of the Mubl/Hox2.3 gene with its human homologue cl. A. comparison of protein sequences. The Hox2.3 protein is displayed. Where a change in the human protein is found, this is indicated underneath. B. Dot-matrix comparison of the Mubl/Hox2.3 gene (nucleotides 150-1415 from the sequence as displayed in Fig. 3) and cl (Ref.46). The protein coding regions are indicated by boxes (black box = homeobox). A dot indicates identity of 19 or 20 out of 20 in the nucleotide segments that are compared.

The subcloned insert from clone LR0.8 was sequenced by using clone pR0.8 and a number of smaller subclones derived from pR0.8. Then, from clone pR1.2 the sequence extending more 5' was determined as well as the 3' terminal fragment to confirm that the 3' end was identical in both cDNAs. Since it appeared that the sequence obtained did not contain the entire amino acid coding sequence we then sequenced subcloned fragments SS1 and SS2 (see Fig. 2) from the genomic clone L2 overlapping the most 5' sequence of the cDNA. We also sequenced a subcloned fragment (RB1.2) from clone L13 overlapping the 3' end of the cDNA sequence, to confirm the location of the poly(A) site and to determine the sequence immediately downstream from the gene. In both cases overlapping stretches of identical sequences from cDNA and genomic DNA were found, confirming the correctness of the mapping data. The sequence, displayed in Fig. 3, consists of 1639 nucleotides including 200 nucleotides from the genomic DNA immediately downstream from the site of polyadenylylation. We observe between the ATG triplet at nucleotide 250 and the TGA triplet at nucleotide 901 an open reading frame of 651 nucleotides which could encode a 217-amino acid protein. We believe that it is reasonable to assume that this is the complete protein coding region of the gene, since

1] It is the longest open reading frame we observe;
2] Two stop codons are present within the 50 bp preceding the ATG codon at nucleotide 250 in the same reading frame;
3] In the putative 217-amino acids protein we recognize a homeo domain and, at a short distance amino proximal from this homeo domain the hexapeptide Ile-Tyr-Pro-Trp-Met-Arg (see below).

The protein has a calculated molecular weight of 23,915 and is like previously described homeodomain proteins rich in proline(6%), serine(8.8%) and basic amino acids (12.4%). The length of the 3' untranslated region is 532 nucleotides and contains the signal AATAAA preceding the site of polyadenylylation (derived from comparison of

a



b

1464 CTGGCTGGGGGAGGTGTAAGGGTTGGTAGTGGGC 1497 Hox2.3

CTGGCTGGGGGTGGGGTAAGGGTTACAGGTGGGC Human β NGF

Fig. 5. A. Similarities between the amino-terminal portions of several homeodomain containing proteins. The highest homologies were found in the boxed portions of the sequence. The sequences are listed in such way to demonstrate several additional homologies; gaps that result from this arrangement therefore do not represent amino acids missing from the original sequence. Only from the Dfd and Antp sequence a stretch of amino acids is deleted as indicated. Since Hox2.3 and c1 are identical in the portion of the protein shown, both are listed in the top line. References: Hox2.3 and c1: this paper and 45; Hox2.1 Ref.21, c13 Ref. 25; X-hox1A Ref. 47; Dfd (Drosophila melanogaster Deformed gene), Ref. 48; Antp (D. melanogaster Antennapedia gene), Ref.49 . B. Sequence similarity between a stretch downstream from the Mubl/Hox2.3 gene and the human β-Nerve Growth Factor gene (Ref.52).

genomic and cDNA sequences) by 21 nucleotides. The site of polyadenylylation is identical in both sequenced cDNAs from two different sources.
Comparative analysis. The above mentioned sequence similarity of Hox2.3 and human c1 is illustrated in Figs. 4a and b. The conservation at the protein level is 93% and at the nucleotide level 92% within the protein coding region, 97% in the 100 bp preceding the ATG codon and 67% in the 3' untranslated portion of the gene which is in the case of hox2.3/Mubl 73 bp shorter than for c1. This is in part a consequence of a deletion from the Hox2.3 gene compared to the c1 gene, of a 40-bp A-rich stretch located 110 bp downstream from the protein-coding region. The nucleotide sequences are compared in Fig. 4b which is a dot matrix sequence comparison using very stringent parameters. Segments of 20 nucleotides are compared and a dot indicates identity of at least 19. A 60-nucleotide stretch located 200 nucleotides downstream from the open reading frame displays a very high degree of conservation suggesting a biological function for this sequence. We did not compare any intron sequences, but the estimated lengths are very close: 2.4 kb for c1 (E. Boncinelli personal communication) and 2.2 kb for Hox2.3). The location of the intron, in the codon for amino acid #67 is identical. The Hox2.3 homeo domain shares with the c1 gene a high degree of similarity to the Drosophila Antennapedia homeodomain. Simeone et al. (46) report a second open reading frame of 156 triplets in the c1 gene. An open reading frame of the same length is also present in the Hox2.3 gene, starting at the corresponding position (an ATG triplet at nucleotide 266). We do not know whether this open reading frame is of biological significance.

When non-homologous homeodomain containing proteins are compared, the conservation in the homeodomain is often extremely high, but usually only a few small regions of similarity are found outside the homeodomain. Comparison of the Hox2.3 protein sequence to other published protein sequences of homeodomain containing proteins yields, in addition to the homeo domain itself, three more regions of similarity:

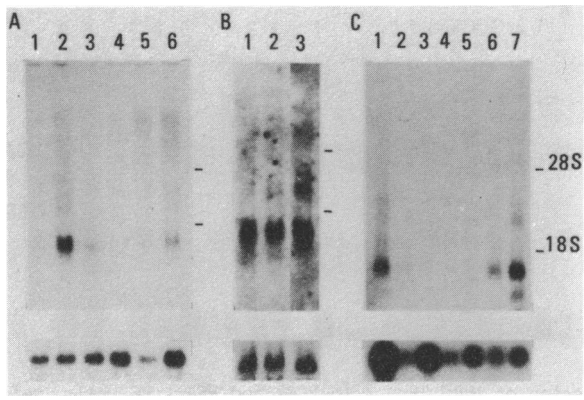


Fig. 6. Expression of Mubl/Hox2.3 in mice. A. Expression in 12.5-day embryos. Total RNA (30 μg/lane) from the following tissue samples were loaded: 1, limb buds; 2, spinal cord; 3, brain; 4, cranio-facial tissue; 5, carcass; 6, viscera. B. Expression of Hox2.3 in spinal cord at different time points. In each lane 30 μg of total RNA is loaded: 1, 12.5-day embryos; 2, 17.5-day embryos; 3, adult mice. C- Expression of Hox2.3 in adult tissues (Lanes 2-7) compared to expression in 13.5 day embryo (total body; lane 1). Approximately 5 μg of poly(A)$^+$ RNA from the following tissue samples were loaded: 1, 13.5-day whole embryos; 2, ovary; 3, intestine; 4, liver; 5, brain; 6, testis; 7, kidney. The three filters were hybridized with a Hox2.3/Mubl probe (A, B, C, upper panels) and subsequently with a β-actin probe (lower panels).

1) a conserved peptide sequence , originally described by Simeone et al. (25) the consensus sequence of which was described by Krumlauf et al. (21) as Ile/Val/Leu - Tyr/Phe - Pro - Trp - Met - Arg. Until now, the Drosophila ftz gene is the only sequenced Antp type homeobox gene that does not contain such a sequence. It preceeds the homeo domain at less than 20 amino acids and its coding sequence is always separated by an intron from the homeobox.
2) The amino-terminal peptide (See Fig. 5a). Mavileo et al. (47) noted that the human c13 gene and the Xenopus laevis XHox-1A (48) gene share a very similar 23 amino acid amino terminal peptide. Based on this and other sequence similarities it appears that both genes are related to the Drosophila Deformed (Dfd) gene (49). We observe in the Mubl/Hox2.3 amino-terminal protein sequence some homology with this sequence and also to that of a number of other homeodomain proteins: the Drosophila Antp (50) protein and the mouse Hox2.1 protein (21). The similarity consists of a Met-Ser-Ser- tripeptide at the start of the protein and a Arg/Lys-Phe/Tyr-Pro- tripeptde located at a distance of about 10 amino acids in the vertebrate proteins, but much farther in the two Drosophila proteins. This motif is absent from the ftz (16) and the human c8 sequence (46). Of relatively few homeodomain proteins the aminoterminal sequence is available. In view of its location in the protein, an obvious guess for its function would be involvement in nuclear localization but we did not find these sequences in other nuclear proteins like histones and nuclear onc genes. Amino terminal amino acids have also been reported to influence the stability of proteins (51).
3) The carboxy-terminal peptide sequence which in a number of cases consists of a row of Glutamic acid residues: six in Hox2.3 (this paper), c1(46) and the X. laevis gene MM3 (52), and 12 in the Hoxl.1 (3) gene. Binding to basic proteins like histones has been suggested as a possible function for this peptide (52). Experimental data addressing the biological function of the sequence similarities mentioned here are lacking . The limited available sequence data suggest that MM3 may be the X. laevis homologue of c1 and Hox2.3/Mubl.

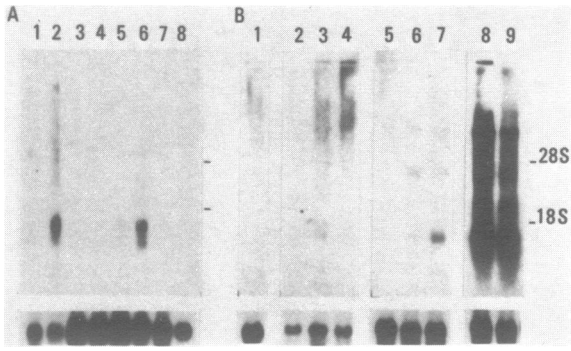


Fig. 7. A. Expression of Mub1/Hox2.3 in PCC7 and PSA-1 EC cells, before and during their differentiation. 10 μg of poly(A)$^+$ from the following samples were loaded: 1, PCC7 EC; 2, PCC7 grown for 3 days as a monolayer in a RA (10^{-6}M) containing medium; 3, PCC7 grown as aggregates for four days and replated for four days in the absence of RA. 4, PSA-1 EC cells; 5, PSA-1 cells grown as aggregates for three days and replated in RA-containing medium for two days; 6, as 5 but cells have grown for five more days in the presence of RA; 7 and 8, same as 5 and 6, respectively,but cells grown in the absence of RA. B- Expression of Hox2.3 in P19 cells before and during their differentiation. 5μg (lanes 2-4) or 10 μg (lanes 1, 5-9) of poly(A)$^+$ RNA were analysed: 1, P19 EC, 2-4, P19 grown as a monolayer in the presence of RA (10^{-6} M) for one (lane 2), three (lane 3) and five (lane 4) days; lane 5-7, P19 cells grown as aggregates in the presence of DMSO (1%) for three days (5) and subsequently plated in DMSO-containing medium for one (lane 6) or seven (lane 7) days; lanes 8 and 9, P19 grown as aggregates in the presence of RA (10^{-6} M) for three days and subsequently plated in the presence of RA for 1 (lane 8) or 2 (lane 9) more days. A and B filters were hybridized with a Hox2.3/Mub1 probe (upper panels) and subsequently with a β-actin probe (lower panels).

No further similarities of the Mub1 protein with other proteins could be detected by searching against a protein sequence data bank. A rather unexpected nucleotide sequence similarity was found when comparing our complete sequence data with those present in a data bank. A 24-nucleotide stretch located at a short distance downstream from the polyadenylylation site shows a remarkable similarity with a sequence which is present in the opposite orientation in the 3' untranslated region of the human β-Nerve Growth Factor (NGF) gene (53). See Fig.5b. Although it is presently unclear to us what the function of this sequence could be, we feel that the similarity is probably more than a coincidence. In view of the location the element in both genes it could only have a coding function if it is part of a different transcription unit. Alternatively, it is possible that it has a regulatory function; in this respect it might be relevant that the Hox 2.3 gene is expressed at relatively high levels in the central nervous system of embryos.

Expression of Mub1/Hox2.3.

Expression in embryos. The results of analysis of RNA from dissected 12.5 day embryos are shown in Fig. 6A. Only one major band, corresponding to a 1.6-kb RNA is detected. The specificity of the Mub1 probe, usually the insert of the cDNA clone pR1.2 is demonstrated by the fact that

1] no other known homeobox RNAs are detected;
2] upon stripping of the filters and rehybridization with other homeobox probes the 1.6 kb band is not observed;
3] in several cases experiments were repeated using the XhoII fragment upstream from the homeobox box band (see Fig. 2), giving identical results.

By far the highest expression is detected in the spinal cord fraction, while we

observe a weak band in RNA from brain and from the viscera. The latter is a mixture of many different cell types; we can therefore not exclude that if the low expression detected were confined to a single organ or tissue type it might still result from a very high local activity of the gene. Cranio facial tissues, and remaining tissues ('carcass') were negative, as were placenta and extraembryonal membranes. A very low level of expression was detected in the limb buds fraction (only visible in longer exposures of the autoradiograms shown here. A more detailed analysis (J.D. et al. submitted elsewhere) of embryos at day 13.5 shows that the weak expression of the brain fraction originates from the posterior part, while the spinal cord signal originates from all three fractions along the longitudinal axis that were obtained by dissection.

To investigate a possible time-dependence of this expression pattern, we then compared Mubl/Hox2.3 expression in spinal cord from embryos at 12.5 days, 17.5 days and adults. As shown in Fig. 6B, no major quantitative difference is detected. The boundaries of the region in the spinal cord where expression is found between 12.5 and 17.5 day, do not appear to change either . We conclude from these experiments that the expression pattern of the Mubl/Hox2.3 gene is highly region specific but relatively stable in the spinal cord in the time span we have now investigated. Similar findings for the Hox3.1 and Hox1.5 gene were reported by Utset et al. (26) and by Krumlauf et al. (21) for Hox2.1.

Expression in adults. We analyzed RNA from various tissues of adult mice by Northern blot analysis. Results of these experiments are shown in Fig.6C. Mubl/Hox2.3 is expressed at high levelsin adult mice in the kidney, testis and spinal cord . We detect a major transcript of about 1.6 kb, while in the kidney a second RNA band corresponding to a fragment length of about 1.4 kb is detected (Fig. 6C). The uterus of a pregnant mouse also contained relatively high amounts of Hox2.3 mRNA (not shown).

In several tissues no expression is detected: liver, gut, heart, brain, ovary (Fig. 6C), skeletal muscle, spleen and lung (not shown). Upon overexposure of the autoradiogram shown in Fig. 6 faint bands become visible in all lanes, more prominently in the lanes containing RNA from brain and ovary (not shown).

We conclude that Hox2.3 is no exception to the rule that expression of homeobox containing genes is highly tissue-specific. Expression in testis was previously reported for the Hox1.1 (3), 2.1 (5,21) and 1.4 (11,42,43) genes. The former two genes are expressed as well in kidney and also in several tissues in which Mubl/Hox2.3 is not expressed (3,5,21).

Expression in embryonal carcinoma cells.
Colberg-Poley et al. (27) have reported that induction of differentiation of F9 EC cells by RA treatment is accompanied by induction of homeobox containing genes. In other early work on homeobox gene expression in differentiating EC cells no or only limited activation was reported (5,12) or the results were variable depending on the cell system used (8). We recently showed that homeobox mRNAs can accumulate to high levels in differentiating EC cells, but that alternative ways of treating EC cells giving rise to very similar or identical cell types, can lead to very different levels of transcripts. It appeared that the induction of a specific differentiation pathway is not sufficient for strong induction, but should be combined with the presence of a chemical inducer like RA.

Expression in PCC7 and PSA-1 cells.The results we show in Fig. 7a confirm and extend these conclusions for Mubl/Hox2.3: we followed expression of Mubl/Hox2.3 in PCC7 and in PSA-1 cells during differentiation. PCC7 cells are induced to differentiate into cholinergic neurons by either one of two treatments: 1) growing of monolayers in the presence of RA or 2) letting the cells grow as aggregates (without adding RA) and subsequently plating them in tissue culture dishes. Only in RA-treated cells high levels of Mubl/Hox 2.3 RNA are observed Fig 7A lane 2). PSA-1 cells can be induced to differentiate into a large variety of cell types resembling those that appear during early embryogenesis, by aggregation of the cells and subsequent plating of the aggregates. The same treatment in the presence of RA results in a similar population of cells (as concluded on morphological grounds). A major difference however, is ob-

served in the levels of Mubl/Hox2.3 RNA: see Fig.7A, lanes 4–8: without RA expression of Hox2.3/Mubl is not detectable, in its presence mRNA accumulates to a high level. Expression in P19 cells. P19 EC cells are a multipotential cell line that can be induced to differentiate along three different pathways, all involving the use of RA. Growing the cells in aggregates in the presence of RA results in differentiation into mainly neural and glial cells. Alternatively, treatment of the aggregates with DMSO results in generation of mesoderm type cells among which myoblasts and no neural cells. RA treatment, on the other hand, of P19 cells growing in monolayers induces a differentation pathway leading after three to five days to flat endoderm-like cells. Fig. 6B shows the results of RNA analysis of these cells.

While expression in the neural cells resulting from RA treatment of aggregates is extremely high (Fig. 6B lanes 8 and 9), RA treatment of the monolayer giving rise to endoderm cells results in much lower levels of expression (Fig. 6B, lanes 2–4). Expression in the mesoderm derivatives that are induced by DMSO treatment is also much lower than in the cell population arising from RA treatment of P19 aggregates. We do not know the origin of the additional high bands that appear in lanes where very high expression of the major band is observed.

We conclude that not only the use of RA is of importance but also the nature of the induced pathway of differentiation. In our experiments with a large variety of EC and EK cells and several homeobox containing genes, (28, J.D. et al. submitted and unpublished, this paper) high levels of homeobox RNA were only detected after treatment with RA, but only if the cells investigated were differentiating, often into neurons and otherwise into a mixed population containing a neural cell type. Apparently the induction of Mubl/ Hox2.3 to high levels (and of other homeobox genes as well, J.D., unpublished), depends on a combination of several factors. The cells should be differentiating, and RA or an other chemical inducer (J.D. et all, submitted) should be present. Generally, the levels of expression observed in EC cells, are higher if neural cells are generated. This is reminiscent of the very high expression of homeobox genes in the central nervous system of embryos. However, there is only a partial correlation of Mubl/Hox2.3 expression with the neural cell type, given the very high expression in for example adult kidney and testis. Furthermore, in the absence of in situ hybridization data we cannot be sure whether expression in spinal cord fractions originates from the nerve cells themselves. Another caveat to our conclusions is: for all our expression experiments, data are lacking that show the amount of protein which is translated from the mRNA, while it is the protein that probably executes the function of homebox genes. Furthermore, it is not certain whether the high expression we detect under certain conditions is biologically relevant: it is conceivable for regulatory genes like we think Mubl is, that very low levels of expression are sufficient and that high expression levels like shown here are in fact inherent to the cell culture systems employed in our studies. Clearly, complete understanding of the biological role of homeobox containing genes can only be achieved by experiments in which their function is more directly tested, for instance by their introduction and subsequent expression in EC cells, or by the generation of transgenic mice in which the expression of the gene is altered.

ACKNOWLEDGEMENTS.
We would like to thank L. Joosen for technical assistance, H. Beks and W. de Graaff for performing some of the experiments and M. Hagenaar and H. Schilthuis for advice in sequence technology. We thank Drs. W. Gehring, M. Scott, P. Gruss and B. Hogan for giving homeobox probes and Dr. G. Grosveld for sending the mouse libraries. We thank B. Hooft van Iddekinge, S. van Genesen and A.van de Kamp for giving the LR0.8 phage. We also thank Dr. E. Boncinelli for communicating unpublished results to us.

REFERENCES
1. Gehring, W.J. (1987) Science 236,1245–1252.
2. Holland. P.W.H. and Hogan, B.L.M. (1986) Nature 321, 251–253.
3. Colberg–Poley, A.M., Voss, S.D., Chowdury, K., Stewart, C.L., Wagner, E.F. and Gruss, P. (1985) Cell 43,39–45.
4. McGinnis, W., Hart, C.P., Gehring, W.J. and Ruddle, F.H. (1984) Cell 38,675–680.
5. Jackson, I.J., Schofield, P. and Hogan, B. (1985) Nature 317, 745–748.
6. Hart, C.P., Awgulewitch, A., Fainsod, A., McGinnis, W. & Ruddle, F. (1985) Cell 43,9–18.
7. Rabin, M., Hart, C.P., Ferguson–Smith, A., McGinnis, W., Levine, M. and Ruddle, F.H.(1985) Nature 314,175–177.
8. Hauser, C.A., Joyner, A.L., Klein, R.D., Learned, T.K., Martin, G.R. and Tjian, R. (1985) Cell 43,19–28.
9. Breier, G., Bucan, M., Francke, U., Colberg–Poley, A. and Gruss, P. (1986) EMBO J. 5,2209–2215.
10. Awgulewitch, A., Utset, M.F., Hart, C.P., McGinnis, W and Ruddle, F.H. (1986)Nature 320,328–335.
11. Duboule, D., Baron, A., Maehl, P. and Galliot, B. (1986) EMBO J. 5,1973–1980.
12. Joyner, A.L., Kornberg, T., Coleman, K.G., Cox, D.R. and Martin, G.R. (1985) Cell 43,29–37.
13. Rabin, M., Ferguson–Smith, A., Hart, C.P. and Ruddle, F.H. (1986) Proc.Natl. Acad.Sci.USA 83,9104–9108.
14. Bucan, M., Yang–Feng, T., Colberg–Poley, A.M., Wolgemuth, D.J., Guenet, J.–L., Francke, U., and Lehrach, H. (1986) EMBO J. 5, 2899–2905.
15. Sheperd, J.C.W., McGinnis, W., Carrasco, A.E., De Robertis, E.M. and Gehring, W.J. (1984) Nature 310 70–71.
16. Laughon, A. and Scott, M.P. (1984) Nature 310, 25–31.
17. Pabo, C.O. and Sauer, R.T. (1984) Ann. Rev. Biochem. 53,293–321.
18. Desplan, C., Theis, J. and O'Farrel, P.H. (1985) Nature 318,630–635.
19. Fainsod, A., Bogarad, L.D., Ruusala, T., Lubin, M., Crothers, D.M. and Ruddle, F.H. (1986) Proc. Natl. Acad. Sci.USA 83,9532–9536.
20. Schneuwly, S., Klemenz, R. and Gehring, W. (1987) Nature 325,816–818.
21. Krumlauf, R., Holland, P.W.H., McVey, J.H. and Hogan, B.L.M. (1987) Development 99,603–617.
22. Wolgemuth, D.J., Engelmyer, E., Duggal, R.N., Gizang–Ginsberg, E., Mutter, G.L., Ponzetto, C., Viviano, C. and Zakeri, Z.F. (1986) EMBO J. 5,1229–1235.
23. Gaunt, S.J., Ross Miller, J., Powell, D.J. and Duboule, D. (1986) 324,662–664.
24. Joyner, A.L. and Martin, G.R. (1987) Genes Devel. 1,29–38.
25. Simeone, A., Mavilio, F., Bottero, L., Giampaolo, A., Russo, G., Faiella, A., Boncinelli, E. and Peschle, C. (1986) Nature 320, 763–765.
26. Utset, M.F., Awgulewitsch, A., Ruddle, F.H. and McGinnis, W. (1987) Science 235,1379–1382.
27. Colberg–Poley, A.M., Voss, S.D., Chowdury, K. and Gruss, P. (1985) Nature 314,713–718.
28. Deschamps, J., De Laaf, R., Joosen, L., Meijlink, F. and Destree, O. (1987) Proc.Natl. Acad.Sci. USA 84,1304–1308.
29. Maniatis, T., Fritsch, E.F. and Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor laboratory.
30. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl. Acad.Sci.USA 74,5463–5467.
31. Chen, E.Y. and Seeburg, P.H. (1985) DNA 4, 165–170.
32. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18,5294–5299.
33. Thomas, P.S. (1980) Proc.Natl.Acad. Sci. USA 77,5201–5205.
34. Feinberg A.P. and Vogelstein B. (1983) Anal. Biochem. 132,6–13.

35. Hilkens, J., Colombatti, A., Strand, M., Nichols, E., Ruddle, F.H. and Hilgers, J. (1979) Somat. Cell Genet. 5, 39–49.
36. Hilkens, J., Cuypers, H.Th., Selten, G., Kroezen, V. and Hilgers, J. (1986) Somat. Cell Genet. 12, 81–88.
37. Hilkens, J., Van der Zeijst, B., Buijs, F., Kroezen, V., Bleumink, N. and Hilgers, J. (1983) J. Virol. 45, 140–147.
38. Fellous, M., Gunther, M., Kemler, R., Wiels, J., Berger, R., Guenet, J.L., Jakob, H. and Jacob, F. (1978) J.Exp.Med. 148,58–70.
39. Martin, G.R., Wiley, L.M. and Damjanov, I. (1977) Dev.Biol. 61, 230–244.
40. Mc. Burney, M.W., and Rogers, B.J. (1982) Dev.Biol. 89,503–508.
41. Jones–Villeneuve, E.M., McBurney, M.W., Rogers, K.A. and Kalnins, V.I. (1982) J. Cell.Biol. 94, 253–262.
42. Edwards, M.K., Harris, J.F. and McBurney, M.W. (1983) Mol.Cell.Biol. 3, 2280–2286.
43. Rubin, M.R., Toth, L.E., Patel, M.D., D'Eustachio, P. and Nguyen–Huu M.C. (1986) Science 233,663–667.
44. Joyner, A.L., Lebo, R.V., Kan, Y.W., Tjian, Cox, D.R. and Martin, G.R. (1985) Nature 314, 173–175.
45. Muenke, M., Cox, D.R., Jackson, I.J., Hogan, B.L.M. and Francke, U. (1986) Cytogenet. Cell Genet. 42,236–240.
46. Simeone, A., Mavilio, F., Acampora, D., Giampaolo, A., Faiella, A., Zappavigna, V., D'Esposito, M., Pannese, M., Russo, G., Boncinelli, E. and Peschle, C. (1987) Proc. Natl. Acad. Sci. USA 84, in press.
47. Mavilio, F., Simeone, A., Giampaolo, A., Faiella, A., Zappavigna, V., Acampora, D., Poiana, G., Russo, G., Peschle, C. and Boncinelli, E. (1986) Nature 324, 664–668.
48. Harvey, R.P., Tabin, C.J. and Melton, D.A. (1986) EMBO J. 5,1237–1244.
49. Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987) EMBO J. 6, 767–777.
50. Laughon, A., Boulet, A.M., Bermingham, J.R., Laymon, R.A. and Scott, M.P. (1986) Mol.Cell.Biol. 6,4676–4689.
51. Bachmair, A., Finley, D. and Varshavsky, A. (1986) Science 234,179–186.
52. Mueller, M.M., Carrasco, A.E. and De Robertis, E.M. (1984) Cell 39,157–162.
53. Ullrich, A., Gray, A., Berman, C. and Dull, T.J. (1983) Nature 303, 821–825.