Nucleic Acids Research, 1993, Vol. 21, No. 1 127–134

Cloning, chromosomal localization and expression pattern of the POU domain gene *Oct-11*

Andrew S.Goldsborough+, Lyn E.Healy, Neal G.Copeland[1], Debra J.Gilbert[1], Nancy A.Jenkins[1], Keith R.Willison and Alan Ashworth*

Chester Beatty Laboratories, Institute of Cancer Research, Fulham Road, London SW3 6JB, UK and [1]Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA

Received September 4, 1992; Revised and Accepted November 23, 1992 EMBL accession no. Z18537

ABSTRACT

POU domain genes encode a family of highly conserved transacting factors that influence the transcriptional activity of several cell type-specific and ubiquitous genes. We have cloned and sequenced cDNAs encoding a novel mouse POU domain protein, Oct-11, that is closely related within the POU domain to the POU class II proteins, Oct-1 and Oct-2. Recombinant Oct-11 protein binds specifically to an octamer sequence *in vitro*. The *Oct-11* gene is expressed during mouse embryogenesis and in the adult thymus and testis. In addition, it is abundant in the myeloma cell line P3/NS-1/1-Ag4.1. We describe the structure of *Oct-11* and its chromosomal localization, and discuss the evidence that the POU class II gene family has evolved by duplication and divergence of a common ancestral gene.

INTRODUCTION

The POU domain proteins are a family of structurally related transacting factors which have been isolated from a variety of organisms (1, 2). The POU domain has a characteristic bipartite structure; a 75–80 amino acid POU-specific domain separated by a short linker from a 60 amino acid carboxy-terminal region related to homeodomain proteins. POU-homeodomains are more closely related to one another than to other homeodomain proteins (1) and can interact weakly with DNA. The POU-specific domain appears to be required for stabilization and high affinity protein/DNA binding (3–7). The linker that separates the POU-specific domain from the POU-homeodomain is poorly conserved, of variable length, and may not be critical for functions such as DNA binding (8).

POU domain genes have been isolated from a variety of organisms and show diverse patterns of expression (9–15). Some of these have been shown to bind to a specific octamer DNA motif (2, 16–18). Several octamer binding proteins called Oct-1 to Oct-10, identified by distinct band shifts, have been reported to be present in various mouse tissues at different stages of development (17). To date four genes that encode these octamer binding proteins, Oct-1 (19), Oct-2 (3, 5, 20–22), Oct-3/4 (15, 23, 24) and Oct-6 (25–28), have been analysed in detail and shown to encode sequence-specific DNA binding proteins. Assigning the remaining band shift activities (17) to specific genes is complicated by the existence of splicing variants of the genes; for example, at least six distinct coding variants of *Oct-2* exist (29). Here we describe a novel murine POU domain gene, Oct-11, that is expressed during embryogenesis as well as in specific adult tissues. We compare the predicted amino acid sequence of *Oct-11* with other POU domain proteins and describe the structure of the *Oct-11* gene and its chromosomal localization.

MATERIALS AND METHODS

Isolation and sequencing of *Oct-11* cDNA and genomic clones

In order to isolate full length cDNA clones, a λZAP thymus cDNA library (Stratagene) was screened with an *Oct-11* POU domain probe (30) (nucleotides 646–1055 in Figure 1) generated by PCR. Hybridization was overnight at 42°C in 50% formamide, 6×SSC, 10×Denhardt's solution [100×Denhardt's solution is 2% bovine serum albumin; 2% polyvinylpyrollidone; 2% Ficoll], 50 μg/ml denatured salmon sperm DNA and 0.1% w/v SDS. Filters were washed twice for 15 minutes at 65°C in 2×SSC/0.1% SDS and once for 30 minutes in 0.1×SSC/0.1% SDS. From approximately 1×10^6 plaques, 5 hybridizing clones were isolated, representing two-independent cloning events. Genomic clones of both the *Oct-11a* and *Oct-11b* loci were isolated from a BALB/c library in the vector λ Charon 4A. Five hybridizing clones were isolated, one of which represented the intronless *Oct-11b* locus and the other four the structural gene *Oct-11a*.

Double stranded plasmid sequencing reactions were carried out according to the Sequenase Version 2.0 kit (USB). PCR products were directly sequenced as follows. PCR products (0.2–1 μg) were excised from an agarose gel, purified by the Geneclean

* To whom correspondence should be addressed

+Present Address: University of California, School of Medicine, Department of Biochemistry, San Francisco, CA 94143-0554, USA

(BIO101) procedure and eluted into 7 μl of water. Following the addition of sequencing primer (100 ng), the DNA was denatured by boiling for 3 minutes and then snap frozen on dry ice. Immediately after thawing, the labelling reaction mix containing enzyme was added and sequencing reactions performed as for plasmid DNA templates.

Reverse transcriptase PCR conditions

A Fast-Track kit (Invitrogen, San Diego, CA) was used to prepare poly(A)$^+$ RNA. Reverse transcription of RNA was performed as follows. Poly(A)$^+$ RNA (1 μg) was mixed with oligo $(dT)_{12-18}$ (0.5 μg) and water added to a total volume of 11 μl. The mixture was heated at 70°C for 10 minutes, quick chilled on ice and 4 μl of 5×reaction buffer (250 mM Tris−HCl pH 8.3, 375 mM KCl and 15 mM $MgCl_2$), 2 μl DTT (100 mM), 1 μl 10 mM dNTP stock and 1 μl M-MLV RNAse H$^-$ reverse transcriptase (BRL, 200 units) were added. After incubation at 37°C for 60 minutes the reactions were diluted to 100 μl with water. PCRs contained 2.5 μl of 10×PCR buffer (150 mM Tris−HCl, pH 8.8; 600 mM KCl; 25 mM $MgCl_2$), 0.625 μl 10 mM dNTPs, 1 μl of each primer (30 ng/μl), 0.2 μl Taq DNA polymerase (Cetus, 5 units/μl), 1 μl template cDNA and water to a total of 25 μl. PCR conditions were 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 1 minute for 30 cycles using an automatic heating block apparatus (Techne, Programmable Dri-Block PHC-1). The following primer pairs were used to analyse the expression of *Oct-1* 5′ TTCAAGCAGAGACGCATTAAGCTAGGC 3′ and 5′ CTCTGCATCATTTAGCCACTTCTCTAA 3′; *Oct-2* 5′ TTCACACAGGGTGATGTGGG 3′ and 5′ CCAGCTGAGCAGCCCAG 3′, and *Oct-11* 5′ AGTCCTCCCCGTCAGACC 3′ and 5′ ATCGATGGAGAAGGAGGT 3′.

RACE PCR

To isolate the 3′ end of *Oct-11* mRNA, RACE PCR was used (31). 1 μg of cell line P3/NS-1/1-Ag.4 poly (A)$^+$ RNA was reverse transcribed with a primer oligo T_{57} (AAGGATCCGTCGACATCGATAATACGACTCACTATAAGGGATTTTTTTTTTTTTTTTT) as described under cDNA synthesis. Aliquots of the cDNA reaction were used as templates for PCR with primers OL1918 (5′ CTCAAGCCACTGCTGGAG 3′) and Ro (5′ AAGGATCCGTCGACATC 3′ — complementary to T_{57}). Another round of amplification was performed with the third internal nested primer, OB10 (5′ AGAAGAGAATCAACTGCC 3′) with Ri (5′ GACATCGATAATACGAC 3′ — complementary to T_{57}). PCR products were electrophoresed on agarose gels and purified by the Geneclean procedure followed by direct sequencing.

Electrophoretic mobility shift assays

Oct-11 and Oct-2 proteins were expressed in *E.coli* as protein A-fusions. *Oct-11* (clone Thoc 3 in Figure 1) and *Oct-2* (full-length cDNA clone, a gift of D.Meijer) coding sequences were introduced into a vector (S.Swift and A.A., unpublished) that contains the protein A gene (32) under the control of a T7 RNA polymerase promoter (33). These constructs were introduced into *E.coli* BL21/DE3(pLysS) (33) and fusion protein synthesis induced with IPTG. Protein A-fusion proteins were purified by IgG−Sepharose (Pharmacia) chromatography according to the manufacturers instructions. Electrophoretic mobility shift assays were performed essentially as previously described (34). Briefly, approximately 10 ng of purified protein A-fusion protein was preincubated in 20 μl of binding buffer (10 mM Hepes (pH 7.9), 60 mM KCl, 1 mM DTT, 1 mM EDTA, 0.25 mg/ml bovine serum albumin and 12% glycerol) with appropriate competitor DNA for 5 min at room temperature, then 10 fmol (approx. 20,000 cpm) of ^{32}P-labelled, annealed octamer oligonucleotides were added and incubated for a further 20 min at room temperature. The reactions were electrophoresed on a non-denaturing 4% polyacrylamide gel in 0.5×TBE buffer at room temperature at 10 V/cm for 90 min. Gels were dried prior to autoradiography. The double stranded octamer oligonucleotide (a gift of G.May) was (top strand) 5′ GGGCTGATTGATTTGCATGTCCAG 3′. The double-stranded non-specific competitor oligonucleotide containing a Sox-5 binding site (34) was (top strand) 5′TCGAGCACTAAAACAATTCAAGCCCGGGG 3′.

Northern blot analysis

RNA electrophoresed in formaldehyde denaturing gels was transferred to Hybond N$^+$ membranes (Amersham, UK) and alkali treated by submerging the membrane in 0.05 M NaOH for 5 minutes. Filters were hybridized at 65°C overnight in 50% formamide, 6×SSC; 10×Denhardt's solution, 50 μg/ml denatured salmon sperm DNA, 0.1% w/v SDS to probes radiolabelled with ^{32}P by hexamer priming. The *Oct-11* probe was a 430 bp fragment (nucleotides 646−1055 in Figure 1), the *Oct-2* probe was derived from a POU domain cDNA (30). Following hybridisation, the membrane was washed twice in 2×SSC/0.1% w/v SDS at 65°C for 15 minutes, once in 2×SSC/0.1% w/v SDS at 65°C for 20 minutes and then twice in 0.1×SSC/0.1% w/v SDS at 65°C for 30 minutes.

Interspecific backcross mapping

Interspecific backcross progeny were generated by mating (C57BL/6J×*M.spretus*) F_1 females and C57BL/6J males as described (35). A total of 205 N_2 progeny were obtained; a random subset of these N_2 mice were used to map the *Oct-11* loci. DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described (36). The *Oct-11* probe was the 430 bp fragment described above. Washing was done to a final stringency of 0.5×SSC, 0.1% SDS, 65°C. Fragments of 5.8, 5.2, 5.0 and 3.9 kb were detected in PstI digested C57BL/6J DNA; fragments of 10.0, 8.7, 5.2, 4.8 and 4.0 kb were detected in PstI digested *M.spretus* DNA. The 10.0 and 4.8 kb *M.spretus*-specific fragments cosegregated in this analysis and defined the *Oct-11a* locus. The 8.7 kb *M.spretus*-specific PstI fragment segregated independently and defined the *Oct-11b* locus. A description of the probes and RFLPs for the loci linked to *Oct-11a* including the *Ets-1* and *Thy-1* genes has been reported (37). The probe for the D2 dopamine receptor (*Drd2*) locus was a 2.4 kb rat cDNA (38) that was kindly provided by O.Civelli (Beaverton, OR). The probe detected fragments of 9.0 and 3.9 kb in SphI digested C57BL/6J DNA and fragments of 9.0 and 5.1 kb in *M.spretus* DNA. The probes and RFLPs for loci linked to *Oct-11b* including the *Acrg*, *Bcl-2* and *En-1* genes have been described (39). Recombination distances were calculated as described (40) using the computer program SPRETUS MADNESS developed by D.Dave (Data Management Services Inc., Frederick, MD) and A.M.Buchberg (NCI-FCRDC, ABL-BRP, Frederick, MD). Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

RESULTS

Isolation of *Oct-11* cDNA clones

The POU domain of *Oct-11* was originally isolated in a screen for novel POU domain genes expressed during spermatogenesis (30). In order to isolate cDNA clones representing the entirety of the *Oct-11* mRNA, various mouse cDNA libraries were screened with a probe derived from the *Oct-11* POU domain. Although the original *Oct-11* clone had been isolated from testis cDNA using PCR (30), screening of three different testis cDNA libraries, including one that was unamplified, failed to reveal any clones corresponding to *Oct-11*. Clones were found in a mouse thymus cDNA library, however. Sequence analysis of two independent cloning events (called Thoc1 and Thoc3) showed that both have 3′ ends that start close to the carboxy-terminus of the POU domain (Figure 1). Neither had a polyA stretch despite the fact that the library was oligo-dT primed. In order to extend the open reading frame 3′ to the POU domain, first strand cDNA was prepared from P3/NS-1/1-Ag-4.1 cells and *Oct-11* cDNA sequences amplified according to the RACE procedure (31). The 3′ RACE resulted in two distinct PCR products that extended approximately 380 bp and 1.5 kb beyond the 3′ end of Thoc3. Characterization of these cDNAs demonstrated that both encoded the same protein sequence but the shorter RACE product was the result of oligo-dT primer annealing to a short A rich sequence within the 3′ untranslated region that does not have an identifiable polyadenylation signal associated with it (data not shown). The coding regions of clones Thoc1, Thoc3 and the 3′ RACE were sequenced on both strands. The combined sequence from Thoc1, Tboc3 and the 3′ RACE is shown in Figure 1. Both Thoc1 and Thoc3 have a single open reading frame and it is possible that one of the two methionine residues at nucleotides 76 and 94 (Figure 1) is used for translational initiation *in vivo*. Both have reasonable matches to a consensus start codon (41). Translation of the longest open reading frame of the combined cDNA sequence indicates that there is a potential open reading frame of 428 amino acids, encoding a protein with a molecular weight of 44 kD that has some homology to the amino terminal region of *Oct-2* but not *Oct-1* (Figure 2). This region of *Oct-11* is rich in the amino acids

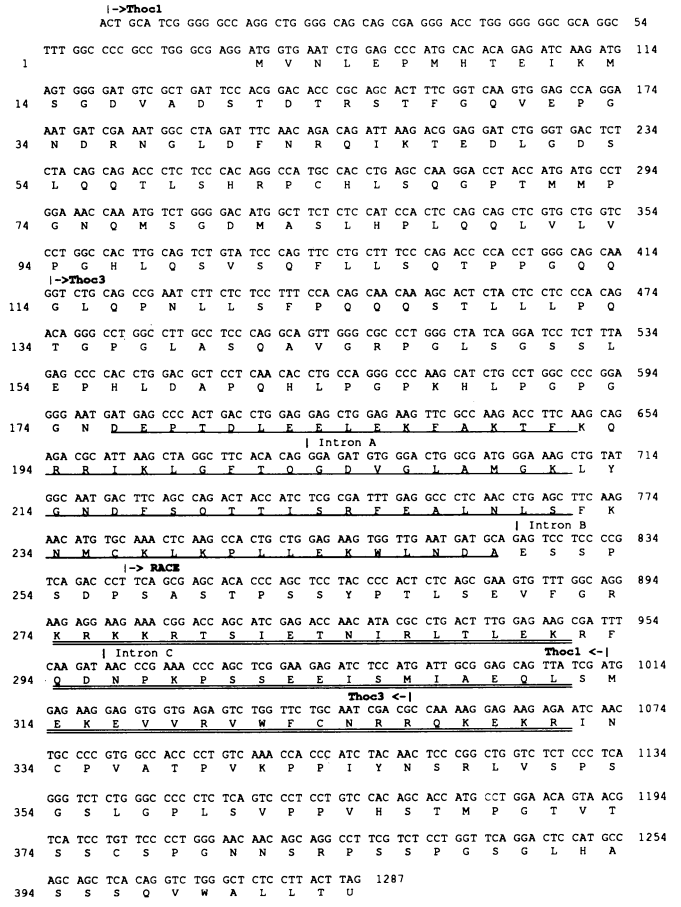

Figure 1. Nucleotide and predicted amino acid sequence of the *Oct-11* cDNA. Amino acids are shown in single letter code. The POU-specific domain is underlined and the POU-homeodomain is double underlined. Numbers to the right of each row refer to the nucleotide position, numbers to the left refer to amino acids. The positions of the 5′ ends of the *Oct-11* cDNA clones Thoc1 and Thoc3 and the 3′ RACE are indicated by the arrows. The end of the coding region is indicated by a U. Translation is shown initiating at the first methionine at nucleotide 76. The positions of introns within the POU domain (shown in Figure 7) are also indicated.

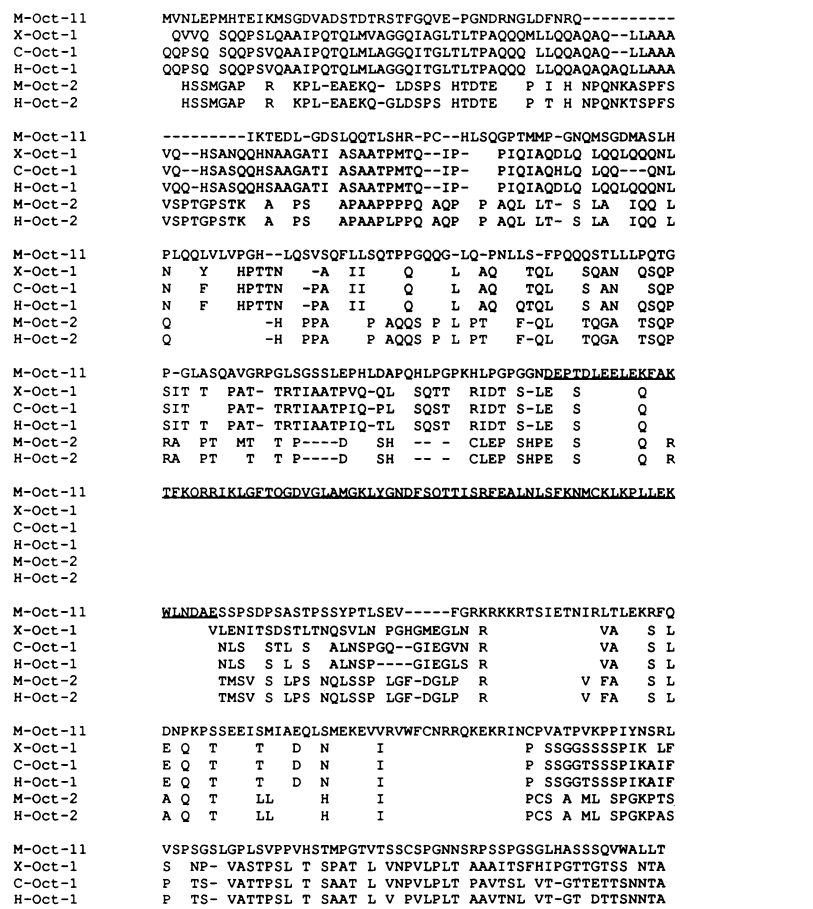

Figure 2. Comparison of the amino acid sequences of the POU class-II proteins. Proteins sequences were compared with the Oct-11 protein from Figure 1. The POU domain is underlined. Spaces indicate identity with Oct-11. Dashes indicate deletions to optimize the alignment. Abbreviations as follows: X-Oct-1, *Xenopus laevis* Oct-1 (42); C-Oct-11, chicken Oct-1 (43); H-Oct-1, human Oct-1 (19); M-Oct-2, mouse Oct-2a (29, 30, 46); H-Oct-2, human Oct-2a (3, 20, 21); M-Oct-11, mouse Oct-11.

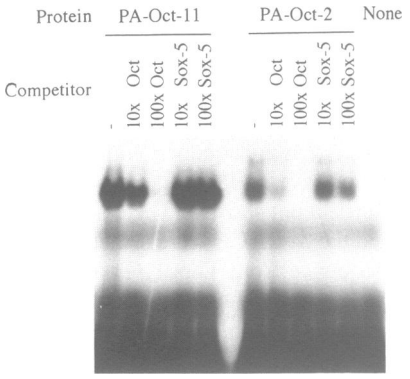

Figure 3. Oct-11 can bind to an octamer motif sequence specifically. Recombinant protein A-Oct-11 or protein A-Oct-2 was incubated with a ^{32}P-labelled oligonucleotide containing an octamer sequence in the absence or presence of the indicated molar excess of specific (Oct) or non-specific (Sox-5) double stranded ofigonucleotides. Protein-DNA complexes were resolved on a polyacrylamide gel and the gel dried and autoradiographed.

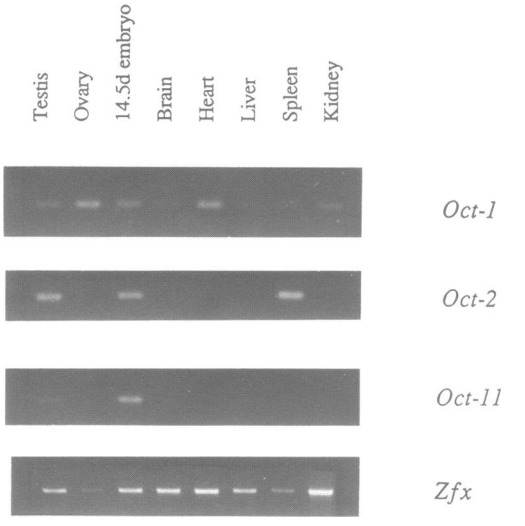

Figure 4. PCR analysis of (a) *Oct-1*, (b) *Oct-2*, (c) *Oct-11*, (d) *Zfx* expression in various murine tissues. Specific PCR primers (see Materials and Methods) were used to amplify cDNA sequences from reverse transcribed mRNA. The PCR products were electrophoresed on agarose gels.

serine, glutamine and proline which are frequently associated with the activation domains of transcription factors. The Oct-11 amino terminus has 43% amino acid identity with the corresponding region of Oct-2 compared to 24% with this region of Oct-1. However, the much higher phylogenetic conservation of this region within the Oct-1 proteins (88% sequence identity between *Xenopus* (42) chicken (43) and human (19) Oct-1) suggests this region plays a specific rather than a general function. The carboxy terminal region of Oct-11 is significantly shorter than either Oct-1 or Oct-2. The carboxy-terminus has previously been demonstrated to be essential for Oct-2 to activate transcription (20, 44, 45). It is worth noting that Oct-1 and Oct-2 have extensive sequence homology within this region, and both are multiply spliced within their coding regions (18, 20, 21, 29, 46).

Oct-11 binds sequence specifically to an octamer motif *in vitro*

The high sequence homology of Oct-11 to Oct-1 and Oct-2 in the POU domain strongly suggested that Oct-11 would be able to bind to an octamer motif. In order to address this, we expressed part of the Oct-11 cDNA containing the POU domain in *E.coli* as a protein A (PA) fusion protein. This protein, and a PA-Oct-2 fusion protein, were purified from bacterial extracts by affinity chromatography and tested for sequence-specific DNA binding in an electrophoretic mobility shift assay (Figure 3). This demonstrated that both PA-Oct-11 and PA-Oct-2 were able to bind to an octamer motif *in vitro* and that this binding was competed by an excess of unlabelled probe DNA. A similar excess of an irrelevant competitor DNA had no effect on DNA binding. Thus Oct-11 is a *bona fide* octamer binding protein and may be able to interact *in vivo* with similar target genes as Oct-1 and Oct-2.

Expression of the POU domain genes *Oct-1, Oct-2* and *Oct-11*

The expression of the POU class II genes was analysed using specific sets of PCR primers and first strand cDNA prepared from a variety of mouse tissues (Figure 4). The expression of *Oct-1* is generally regarded as ubiquitous and our results agree with this, although the relative amount of product appears to vary slightly between RNA samples (Figure 4a). These variations may be a result of slight variations in the amount of template cDNA. The distribution of *Oct-2* transcripts is more restricted than *Oct-1*; expression is detected only in the 14.5 day *post-coitum* embryo, and adult testis and spleen (Figure 4b). Expression in the spleen is in good agreement with previous data indicating that *Oct-2* expression is found within B cell lineages in the adult (22). *Oct-2*-like protein complexes have been detected in sperm and embryos (17) and transcripts in the testis and embryo (46). Our results are in agreement with this data. *Oct-11* transcripts are also restricted in their tissue distribution, expression is detectable only in testis, thymus (data not shown) and 14.5 day p.c. embryos (Figure 4c). The zinc finger gene *Zfx*, was used as a control for cDNA integrity. By Northern blot and PCR analysis *Zfx* appears to be non-abundant and expressed in all tissues (47).

Oct-11 expression in haemopoietic cell lines

The isolation of *Oct-11* cDNAs from a thymus library suggested that *Oct-11* might be expressed in a lymphoid cell type. PCR analysis of a variety of murine haemopoietic cell lines with *Oct-11*-specific primers detected abundant expression of *Oct-11* in the myeloma cell line P3/NS-1/1-Ag4.1 (48–50). These results are summarised in Table 1. Northern blot analysis of RNA extracted from a variety of cell, organ and other tissue types confirmed that expression of *Oct-11* was restricted to the myeloma cell line P3/NS-1/1-Ag4.1 (Figure 5). This cell line expresses *two Oct-11* transcripts, estimated to be 2.8 kb and 5 kb in size (Figure 5a, lane 4). Expression of *Oct-2* was also studied by Northern analysis in the myeloma cell lines using an *Oct-2* POU domain probe (Figure 5b). The P3U1 myeloma cell line expresses two distinct *Oct-2* transcripts of 5 kb and 2 kb (Figure 5b, lane 7), whilst the other 3 myeloma cell lines all express the 5 kb but not the 2 kb transcript (Figure 5b, lanes 4, 5 and 6). Integrity and equality of loading of RNAs was confirmed by reprobing the blot with an actin probe (Figure 5c). This confirms that *Oct-11* can be expressed within the same cell type as *Oct-2* and that *Oct-11* is abundantly expressed in the P3/NS1/1-Ag4.1 cell line and its two hybridoma derivatives MR26/8cC11/E6/C12 and 2MR/47/A/D4/H1.

Table 1. Expression of Oct-2 and Oct-11 in various haemopoietic cells studied by PCR

Cell Line	*Oct-2*	*Oct-11*
PB1 (Pre-B)	+	−
2e8 (Pre-B)	+	−
Lyd9 (Lymphomyeloid progenitor)	nd	−
A4 (Lymphomyelold progenitor)	nd	−
416B (myeloid progenitor)	nd	−
WEHI 3B (myelomonocytic)	nd	+
P3X63 Ag8 (myeloma)	+	−
PU31 (myeloma)	+	−
P3/X63.653 (myeloma)	+	−
P3/NS-1/1-Ag4.1 (myeloma)	+	+++
MR26/8cC11/E6/C12 (hybridoma)	+	+++
2MRa/47a/D4/III (hybridoma)	+	+++

− indicates no detectable expression, + low to moderate expression and +++ high expression.
nd indicates not done.

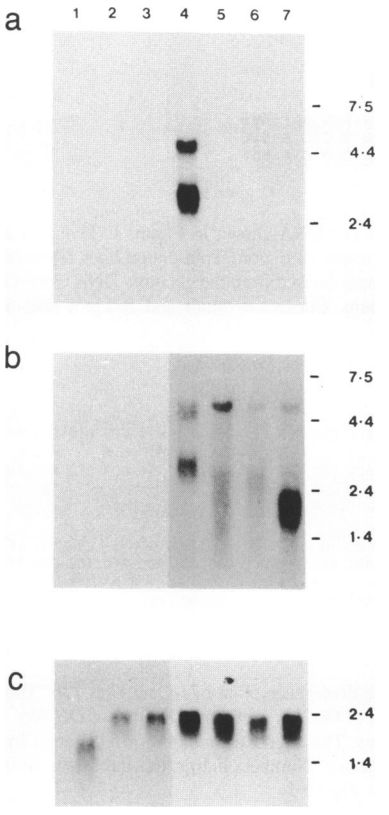

Figure 5. Northern blot analysis of *Oct-11* and *Oct-2* expression in various mouse tissues and cell lines. Two μg polyA$^+$ RNA prepared from 1. Hind leg muscle, 2. Kidney, 3. 14.5 day p.c. embryo, 4. P3/NS1/1-Ag4.1 myeloma cell line, 5. P3.X63 Ag8 myeloma eell line, 6. Ps.X63.653 myeloma cell line, 7. P3U1 myeloma cell line, were electrophoresed in a formaldehyde/agarose gel and blotted to Hybond N$^+$. The same Northern blot was hybridized and then stripped and rehybridized with the probes indicated (**a**) Oct-11, (**b**) Oct-2 and (**c**) actin. Size of RNA molecular weight markers (in kb) (BRL-Gibco) are shown to the right of each panel.

Chromosomal localization of *Oct-11* genes in the mouse

Southern analysis of genomic DNA indicated that there are two *Oct-11* genes in the mouse genome but only one in the rat (data not shown). The mouse chromosomal locations of the two murine

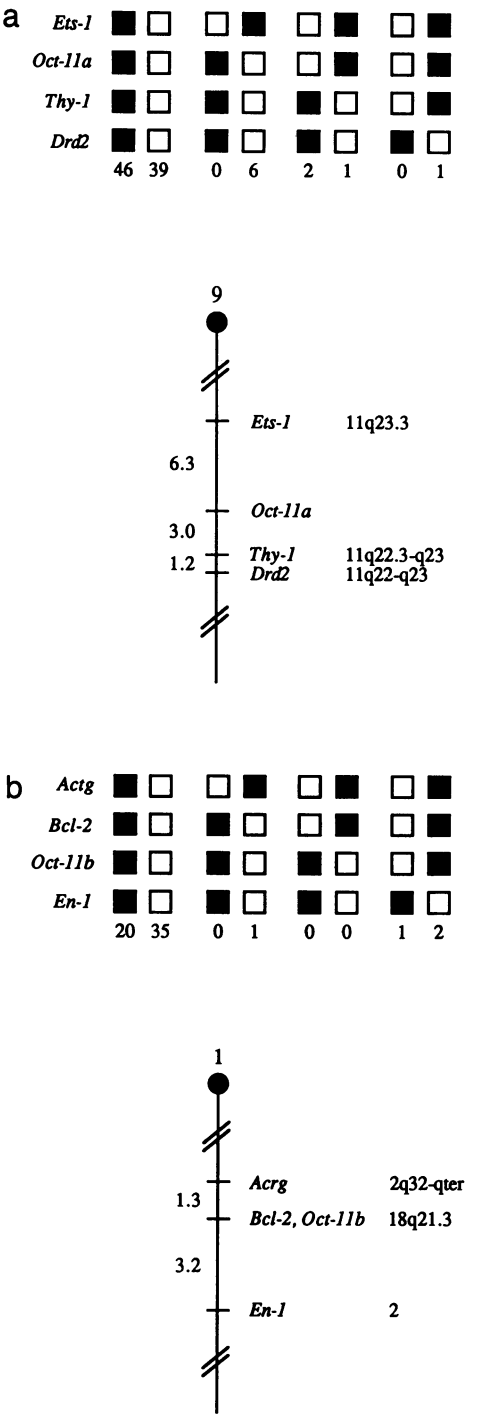

Figure 6. Chromosomal mapping of the mouse *Oct-11* genes. **a**. The *Oct-11a* locus maps to mouse chromosome 9. **b**. The *Oct-11b* locus maps to mouse chromosome 1. The *Oct-11a* and *Oct-11b* loci were mapped by interspecific backcross analysis. The segregation patterns of *Oct-11a*, *Oct-11b* and flanking genes in the backcross animals that were typed in common is shown at the top of parts a and b. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J×*M.spretus*) F_1 parent. The shaded boxes represent the presence of the *M.spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. Partial chromosome 9 and chromosome 1 linkage maps showing the location of *Oct-11a* and *Oct-11b* in relation to linked genes are shown. Recombination distances between loci in centiMorgans are shown to the left of the chromosome and the positions of loci on human chromosomes are shown to the right.

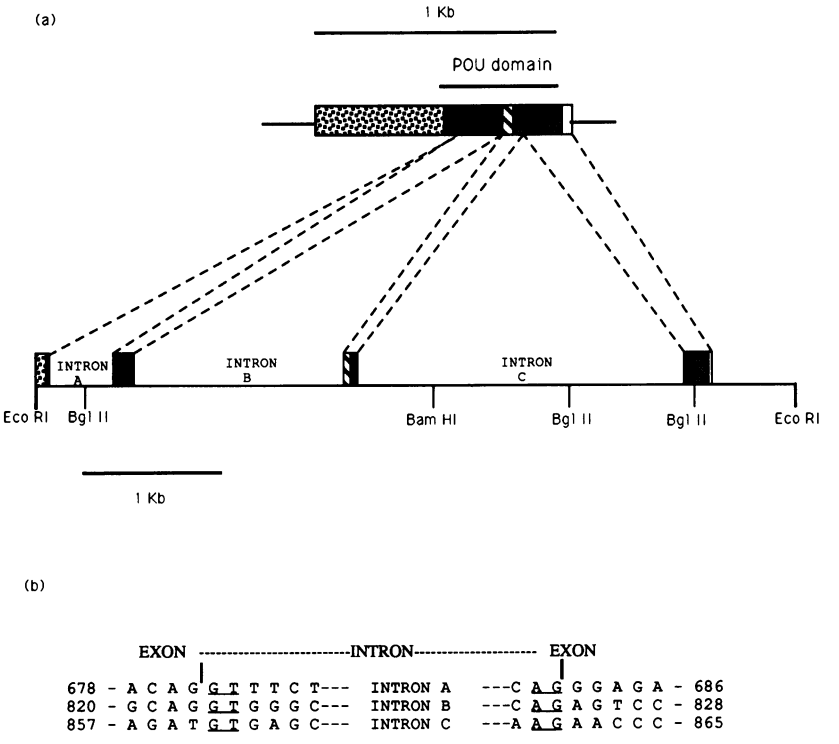

Figure 7. Intron/Exon structure of the *Oct-11a* POU domain. (**a**) Schematic representation of the *Oct-11* cDNA shown in Figure 1. Domains are as follows: a. The glutamine rich amino-terminus, b. POU-spp domain, c. the linker, d. the POU-hd and e. carboxy terminus region. Thin dotted lines represent the exon/intron junctions relative to the cDNA. The lower half of the figure is a schematic view of the 5.6 kb EcoRI fragment derived from the genomic DNA clone GE5.1. Restriction sites and exons are shown. (**b**) Sequences of splice acceptor and donor sites of the *Oct-11* POU domain. Consensus donor and acceptor sites are underlined.

Oct-11 genes were determined by interspecific backcross analysis using progeny derived from matings of [(C57BL/6J×*Mus spretus*) F_1×C57BL/6J] mice. This interspecific backcross mapping panel has been typed for over 700 loci that are well distributed among all the autosomes as well as the X chromosome (35). C57BL/6J and *Mus spretus* DNAs were digested with several enzymes and analysed by Southern blot hybridization for informative restriction fragment length variants (RFLVs) using the *Oct-11* probe. Several *Mus spretus*-specific PstI RFLVs were identified and used to follow the segregation of the *Oct-11* loci in backcross mice. The mapping results indicated that the two *Oct-11* genes, which we have called *Oct-11a* and *Oct-11b*, are unlinked in the mouse genome. *Oct-11a* is located in the proximal region of mouse chromosome 9 linked to *Ets-1*, *Thy-1* and *Drd2*. Although 95 mice were analysed for every marker and are shown in the segregation analysis (see Figure 6a), up to 162 mice were typed for some markers. Each locus was analysed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number exhibiting recombinant chromosomes to the total number of mice analysed for each pair of loci and most likely order of gene order are: centromere – *Ets-1* – 6/95 – *Oct-11a* – 3/99 – *Thy-1* – 2/162 – *Drd2*. The recombination frequencies [expressed as genetic distances in centimorgans (cM) +/– the standard error] are *Ets-1* – 6.3 +/– 2.5 – *Oct-11a* – 3.0 +/– 1.7 – *Thy-1* – 1.2 +/– 0.9 *Drd2*. The placement of these loci relative to other chromosome 9 markers and a comparison of the interspecific backcross map with the composite intraspecific backcross map has been reported previously (37). *Oct-11a* is located in a region of mouse chromosome 9 homologous with the long arm of human

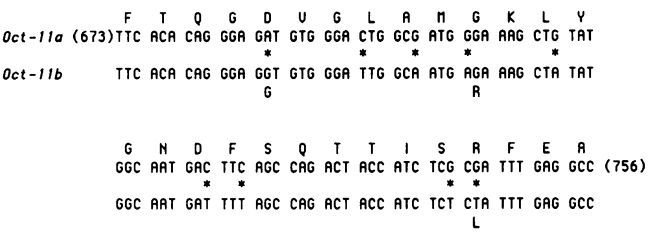

Figure 8. Sequence divergence of *Oct-11a* and *Oct- 11b*. The sequence of part of the POU domain of *Oct-11a* is compared to that of *Oct-11b*. Asterisks indicate sequence differences. The amino acids in *Oct-11b* different from those encoded by *Oct-11a* are indicated. Numbers in brackets indicate positions in the *Oct-11a* cDNA sequence in Figure 1.

chromosome 11. This suggests that the human homologue of the *Oct-11a* locus will reside on chromosome 11 in the region of q23.

A 8.7 kb *M.spretus*-specific PstI RFLV was used to follow the segregation of the *Oct-11b* locus in backcross mice. The mapping results indicated that *Oct-11b* is located in the middle of mouse chromosome 1 tightly linked to *Acrg*, *Bcl-2* and *En-1*. Although 59 mice were analysed for every marker and are shown in the segregation analysis (Figure 6b), up to 127 mice were typed for markers. Each locus was analysed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant to the total number of mice analysed for each pair of loci and the most likely order are: centromere – *Acrg* – 1/77 – *Oct-11b* – 0/59 – *Bcl-2* – 4/127 – *En-1*. The recombination frequencies [expressed as genetic distances in centiMorgans (cM) +/– the

standard error] are *Acrg* − 1.3 +/− 1.3 [*Oct-11b*, *Bcl-2*] − 3.2 +/− 1.6 − *En-1*. The fact that no recombination between *Oct-11b* and *Bcl-2* was observed in 59 mice suggests the two loci reside within 5 cM (upper 95% confidence limit). The genetic distance between [*Oct-11b* and *Bcl-2*] and *En-1* was calculated using the data for *Bcl-2* and *En1*]. The placement of these loci relative to other chromosome 1 markers and a comparison of the interspecific backcross map has been reported previously (39).

Structure of the murine *Oct-11* genes

We screened a BALB/c murine genomic library and identified hybridising clones representing both the *Oct-11a* and *Oct-11b* genes. Analysis of the structure of the POU domain region of the *Oct-11a* gene (Figure 7) revealed the presence of three introns. The splice site sequences of *Oct-11a* are shown in Figure 7a. These conform well to consensus sequences (51). All three introns are conserved in the POU class II gene *Oct-2* (46). This strengthens the argument for a common ancestor for this class of POU domain gene. Two of these introns (A and C, Figure 7) are also present at similar positions in the mammalian POU domain gene, *Pit-1* (52) and the nematode gene *Unc-86* (53). *Oct-11b*, in contrast, does not have any of the introns we identified in *Oct-11a* and is therefore probably derived by retroposition. *Oct-11b* has diverged at 10 nucleotides in a total of 90 within the POU specific domain compared with the structural gene and is therefore a clearly distinct but related sequence (Figure 8). This high level of divergence, together with the absence of *Oct-11b* from rats (data not shown) suggests that this gene may be under the reduced selective pressure typical of pseudogenes.

DISCUSSION

We have cloned and characterized cDNA for *Oct-11*, a novel member of the POU domain gene family. Analysis of the sequence of this gene reveals that it is most closely related to *Oct-1* and *Oct-2* and thus is the third member of the POU class II family isolated. The amino-terminal region of the Oct-11 protein is more closely related to Oct-2 than Oct-1 perhaps indicating a conserved function or more recent divergence of these two genes. This domain has previously been shown to be important for transcriptional activation (44, 45). In contrast to the amino-terminal region, the POU domain is highly conserved. All members of the family are related by sequence homology within this domain that is composed of the POU specific domain, the linker and the POU homeodomain. All three of these domains may contribute to DNA binding (54, 55), however the interaction of the POU specific domain with the major groove suggests that this region confers DNA binding specificity by making sequence-specific contacts with DNA (54). As predicted from the high sequence homology of Oct-11 with Oct-1 and Oct-2, Oct-11 can bind sequence specifically to an octamer site *in vitro* (Figure 3). The POU domain has also been shown to be important for protein−protein interactions (55). These can homo- (16, 56) or hetero-dimeric (57, 58) in nature. For example, Oct-1 can homodimerize or heterodimerize with Oct-2a or Pit-1, and both the POU-specific and POU-homeodomains are required for these interactions (55). It seems likely that Oct-11 will be involved in similar interactions.

The ability of the Oct-1 POU domain to interact with the Herpes virus protein VP16 is well established (8, 59). Regions of the POU domain that interact with other components of the transcriptional apparatus have been inferred from mutational analyses of Oct-1 and Oct-2. The second predicted alpha-helix within the Oct-1 POU homeodomain appears to be critical for VP16 interaction (8). It may be significant that the mouse Oct-1 homeodomain sequence is diverged at four positions compared with *Xenopus*, chicken and human Oct-1. Two of the four changes are within the second helix (30). Considering that Oct-2 does not interact with VP-16 efficiently (8, 59) it will be of great interest to establish whether the murine Oct-1 and Oct-11 proteins are capable of complexing with VP16.

We used PCR to examine a range of tissues and cell lines for *Oct-11* expression. *Oct-11* mRNA was detected in the adult thymus and testes and in 14.5 day p.c. embryos. However, no detectable signal was obtained by nonhem blot analysis of RNA isolated from these tissues. This suggests that the expression was at a low level. Alternatively, significant levels of expression may be limited to a subset of cells or developmental stages. It is possible that the PCR signal in testis was due to blood contamination, however we did not detect *Oct-11* expression in well-vascularised tissues such as kidney and liver. Expression was also absent from the spleen which is composed of up to 50% B cells, although *Oct-11* is expressed in the plasma cell line P3/NS1/1-Ag4.1 (see below). This suggests it is not normally expressed during B cell differentiation although we have not examined early B cells. Furthermore *Oct-11* expression could not be detected in any B cell line analysed (see Table 1). Although expression could be detected in the thymus we have not yet identified the specific cell type involved. We found *Oct-11* to be abundantly expressed in P3/NS-1/1-Ag4.1 (P3/NS), a non-antibody secreting myeloma cell line which is a derivative of the mineral oil-induced myeloma MOPC 21 (60). Two antibody-secreting hybridomas derived from fusion of P3/NS with mouse spleen cells also express *Oct-11* abundantly (data not shown). However P3U1 (61), P3X63 Ag8 (49, 50, 61, 62) and P3/X63.653 (63) which are also derived from MOPC 21 do not express *Oct-11* detectably. This suggests that some change occurred specifically in the P3/NS cell line to activate transcription of the *Oct-11* gene. A possible mechanism for the activation is suggested by the observation that two genes, *c-mos* and the interleukin 6 receptor, are known to be activated in the P3/NS cell line due to the insertion of an intracisternal A particle into regulatory regions (64−67). It is possible that *Oct- 11* is activated in a similar manner in this cell line.

Two unlinked genes were found homologous to *Oct-11* in the mouse genome in contrast to rats and humans which have only one (data not shown). These genes were mapped, using an interspecific backcross, to mouse chromosome 9 (*Oct-11a*) and mouse chromosome 1 (*Oct-11b*). One of these genes, *Oct-11b*, is likely to be a pseudogene having diverged significantly from *Oct-11a* (Figure 8). The active gene, *Oct-11a*, is localized to part of mouse chromosome 9 syntenic with human chromosome 11q. In particular, *Oct-11* is predicted to map close to a region (11q23) frequently translocated in acute lymphoblastic leukaemias (68). Further work is underway to map the human gene in detail but preliminary results support the localisation to human chromosome 11.

All three members of the POU class-II gene family have now been mapped in the mouse. The three genes are unlinked; *Oct-1* is on chromosome 1 (69), *Oct-2* is on chromosome 7 (70) and *Oct-11* on chromosome 9. Interestingly these regions of the mouse genome are thought to be paralogous or ancestrally related (71). This further emphasises the evolutionary relatedness of the class II genes. The isolation of *Oct-11*-like genes from *Xenopus laevis* (72) indicates that the separation of the class II genes occurred

before the divergence of mammals and amphibians. Elucidation of the genomic organization of the *Oct-11a* POU domain revealed a striking concordance of intron/exon junctions with *Oct-2*. Two of the junctions are also in common with the POU domain genes *Pit-1* (54) and *Unc-86* (55). This provides further evidence that the POU domain gene family arose from a series of gene duplications.

Oct-11 can bind to an octamer site, *in vitro*, suggesting that *in vivo* it may be able to interact with similar target genes as Oct-1 and Oct-2. It will be interesting to establish how the specificity of the interaction of these proteins with DNA is determined in cells, such as P3/NS, which express all three of the POU-domain class II genes.

ACKNOWLEDGEMENTS

This work was funded by the Cancer Research Campaign, the Medical Research Council, the Leukaemia Research Fund and the National Cancer Institute (DHHS contract NO1-CO-74101 with ABL). We thank D.Angle and D.A.Swing for excellent technical assistance, P.Denny, F.Bedford, G.May and D.Latchman for helpful discussions and M.Muller-Immergluck for critical reading of the manuscript.

REFERENCES

1. Herr,W., Sturm,R.A., Clerc,R.G., Corcoran,L.M., Baltimore,D., Sharp,P.A., Ingraham,H.A., Rosenfeld,M.G., Finney,M. and Ruvkun,G. (1988) *Genes Dev.* **2**, 1513–1516.
2. Scholer,H.R. (1991) *Trends Genet.* **7**, 323–329.
3. Clerc,R.G., Corcoran,L.M., LeBowitz,D., Baltimore,D. and Sharp,P.A. (1988) *Genes Dev.* **2**, 1570–1581.
4. Garcia-Blanco,M. A., Clerc,R.G. and Sharp,P.A. (1989) *Genes. Dev.* **3**, 739–745.
5. Ko,H.S., Fash,P., McBride,W. and Staudt,L.M. (1988) *Cell* **55**, 135–144.
6. Sturin,R.A. and Herr,W. (1988) *Nature* **336**, 601–604.
7. Verrijzer,C.P., Kal,A.J. and van der Vliet,P.C. (1990) *Genes Dev.* **4**, 1964–1974.
8. Stern,S., Tanaka,M. and Herr,W. (1989) *Nature* **341**, 624–630.
9. Billin,A.N., Cockerill,K.A. and Poole,S.J. (1991) *Mech. Dev.* **34**, 75–84.
10. Bodner,M., Castrillo,J., Theill,L.E., Deerinck,T., Ellisman,M. and Karin,M. (1988) *Cell* **55**, 505–515.
11. Burglin,T., Finney,M., Coulson,A. and Ruvkun,G. (1989) *Nature* **341**, 239–243.
12. Ingraham,H.A., Chen,R.P., Mangalam,H.J., Elsholtz,H.P., Flynn,S.E., Lin,C.R., Simmons,D.M., Swanson,L. and Rosenfeld,M.G. (1988) *Cell* **55**, 519–529.
13. Johnson,W.A. and Hirsh,J. (1990) *Nature* **343**, 467–470.
14. Li,S., Crenshaw,E.B., Rawson,E.J., Simmons,D.M., Swanson,L.W. and Rosenfeld,M.G. (1990) *Nature* **347**, 528–533.
15. Rosner,M.H., Vigano,M.A., Ozato,K., Timmons,P., Poitier,F., Rigby,P.W.J. and Staudt,L.M. (1990) *Nature* **345**, 686–692.
16. LeBowitz,J.H., Clerc,R.G., Brenowitz,M. and Sharp,P.A. (1989) *Genes Dev.* **3**, 1625–1638.
17. Scholer,H. R., Hatzopoulus,A.K., Balling,R., Suzuki,N. and Gruss,P. (1989) *EMBO J.* **8**, 2543–2550.
18. Schreiber,E., Matthias,P., Muller,M.M. and Schaffner,W. (1988) *EMBO J.* **7**, 4221–4229.
19. Sturm,R.A., Das,G. and Herr,W. (1988) *Genes Dev.* **2**, 1582–1599.
20. Muller,M.M., Ruppert,R., Schaffner,W. and Matthias,P. (1988) *Nature* **336**, 544–551.
21. Schiendereit,C., Cromlish,J.A., Gerster,T., Kawakami,K., Balmaceda,C., Currie,A. and Roeder,R.G. (1988) *Nature* **336**, 551–557.
22. Staudt,L., Clerc,R.G., Singh,H., LeBowitz,J.H., Sharp,P.A. and Baltimore,D. (1988) *Science* **241**, 577–580.
23. Okamoto,K., Okazawa,H., Okuda,A., Sakai,M., Muramatsu,M. and Hamada,H. (1990) *Cell* **60**, 461–472.
24. Scholer,H.R., Ruppert,S., Suzuki,N., Chowdhury,K. and Gruss,P. (1990) *Nature* **344**, 435–439.
25. He,X., Treacy,M.N., Simmons,D.M., Ingraham,H.A., Swanson,L.W. and Rosenfeld,M.G. (1989) *Nature* **340**, 35–41.
26. Meijer,D., Graus,A., Kraay,R., Langeveld,A., Mulder,M.P. and Grosveld.G. (1990) *Nucleic Acids Res.* **18**, 7357–7365.
27. Monuki,E.S., Kuhn,R., Weinmaster,G., Trapp,B.D. and Lemke,G. (1990) *Science* **249**, 1300–1303.
28. Suzuki,N., Rohdewohld,H., Neuman,T., Gruss,P. and Scholer,H.R. (1990) *EMBO J.* **9**, 3723–3732.
29. Wirth,T., Priess,T. Annweiler,A. Zwilling,S. and Oeler,B. (1991) *Nucleic Acids Res.* **19**, 43–51.
30. Goldsborough,A., Ashworth,A. and Willison,K. (1990) *Nucleic Acids Res.* **18**, 1634.
31. Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
32. Nilsson,B., Abrahmsen,L. and Uhlen,M. (1985) *EMBO J.* **4**, 1075–1080.
33. Studier,F.W. (1991) *J. Mol. Biol.* **219**, 37–44.
34. Denny,P., Swift,S., Connor,F. and Ashworth,A. (1992) *EMBO J.* **11**, 3705–3712.
35. Copeland,N.G. and Jenkins,N.A. (1991) *Trends Genet.* **7**, 113–118.
36. Jenkins,N.A., Copeland,N.G., Taylor,B.A. and Lee,B.K. (1982) *J. Virol.* **43**, 26–36.
37. Kingsley,D.M., Jenkins,N.A. and Copeland,N.G. (1989) *Genetics* **123**, 165–172.
38. Bunzow,J.R., Van Tol,H.H., Grandy,D.K., Albert,P., Salon,J., Christie,M., Machida,C.A., Neve,K.A. and Civelli,O. (1988) *Nature* **336**, 783–787.
39. Copeland,N.G., Silan,C.M., Kingsley,D.M., Jenkins,N.A., Cannizzaro,L.A., Croce,C.M., Huebner,K. and Sims,J.E. (1991) *Genomics* **9**, 44–50.
40. Green,E.L. (1981) In *Genetics and Probability in Animal Breeding Experiments*. Oxford University Press, New York, pp. 77–113.
41. Kozak,M. (1987) *J. Mol. Biol.* **196**, 947–950.
42. Smith,D.P. and Old,R.W. (1990) *Nucleic Acids Res.* **18**, 369.
43. Petryniak,B., Staudt,L.M., Postema,C.E., McCormack,W.T. and Thompson,C.B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1099–1103.
44. Gerster,T., Balmaceda,C.G. and Roeder,R.G. (1990) *EMBO J.* **9**, 1635–1643.
45. Tanaka,M. and Herr,W. (1990) *Cell* **60**, 375–386.
46. Hatzopoulus,A.K., Stoykova,A.S., Erselius,J.R., Goulding,M., Neuman,T. and Gruss,P. (1990) *Development* **109**, 349–362.
47. Ashworth,A., Skene,B., Swift,S. and Lovell-Badge,R. (1990) *EMBO J.* **9**, 1529–1534.
48. Cowan,N.J., Secher,D.S. and Milstein,C. (1974) *J. Mol. Biol.* **90**, 691–701.
49. Horibata,K. and Harris,W. (1970) *Exp. Cell. Res.* **60**, 61–77.
50. Kohler,G. and Milstein,C. (1975) *Nature* **256**, 495–497.
51. Shapiro,M.B. and Senapathy,P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
52. Li,S., Crenshaw,E.B., Rawson,E.J., Simmons,D.M., Swanson,L.W. and Rosenfeld,M.G. (1990) *Nature* **347**, 528–533.
53. Finney,M., Ruvkun,G. and Horvitz,H.R. (1988) *Cell* **55**, 757–769.
54. Aurora,R. and Herr,W. (1992) *Mol. Cell. Biol.* **12**, 455–467.
55. Verrijzer,C.P., van Oosterhout,J.A.W.M. and van der Vliet,P.C. (1992) *Mol. Cell. Biol.* **12**, 542–551.
56. Ingraham,H.A., Flynn,S.E., Voss,J.W., Albert,V.R., Kapiloff,M.S., Wilson,L. and Rosenfeld,M.G. (1990) *Cell* **61**, 1021–1033.
57. Treacy,M.N., He,X. and Rosenfeld,M.G. (1991) *Nature* **350**, 577–584.
58. Voss,J.F., Wilson,L. and Rosenfeld,M.G. (1991) *Genes Dev.* **5**, 1309–1320.
59. Gerster,T. and Roeder,R.G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6347–6351.
60. Svasti,J. and Milstein,C. (1972) *Biochem J.* **128**, 427–444.
61. Lemke,H., Hammerling,G., Hohmann,C. and Rajewsky,K. (1978) *Nature* **271**, 249–251.
62. Koprowski,H., Gerhard,W. and Croce,C.M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2985–2988.
63. Kearney,J.F., Radbruch,A., Liesegang,B. and Rajewsky,K. (1979) *J. Immunol.* **123**, 1548–1553.
64. Cohen,J.B., Unger,T., Rechavi,G., Canaani,E. and Givol,D. (1983) *Nature* **306**, 797–799.
65. Gattoni-Celli,S., Hsiao,W.-L.W. and Weinstein,I.B. (1983) *Nature* **306**, 795–796.
66. Hawley,R.G., Shulman,M.J., Murialdo,H., Gibson,D.M. and Hozumi,N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7425–7429.
67. Sugita,T., Totsuka,T., Saito,K., Yamasaki,K., Taga,T., Hirano,T. and Kishimoto,T. (1990) *J. Exp. Med.* **171**, 2001–2009.
68. Botti,A.C and Vernal,R.S. (1990) *Anticancer Res.* **10**, 519–525.
69. Hsieh,C.L., Sturn,R., Herr,W. and Francke,U. (1990) *Genomics* **6**, 666–672.
70. Siracusa,L.D., Rosner,M.H., Vigano,M.A., Gilbert,D.J., Staudt,L.M., Copeland,N.G. and Jenkins,N.A. (1991) *Genomics* **10**, 313–326.
71. Nadeau,J. and Kosowsky,M. (1991) *Mammalian Genome* **1**, S433–S460.
72. Baltzinger,M., Stiegler,P. and Remy,P. (1990) *Nucleic Acids Res.* **18**, 6131.