The EMBO Journal vol.12 no.4 pp.1387–1401, 1993

Gtx: a novel murine homeobox-containing gene, expressed specifically in glial cells of the brain and germ cells of testis, has a transcriptional repressor activity *in vitro* for a serum-inducible promoter

Issei Komuro[1], Martin Schalling[2], Lothar Jahn[1], Rolf Bodmer[3], Nancy A.Jenkins[4], Neal G.Copeland[4] and Seigo Izumo[1,5]

[1]Molecular Medicine and Cardiovascular Divisions, Beth Israel Hospital, and Department of Medicine and Program in Cell and Developmental Biology, Harvard Medical School, 300 Brookline Ave, Boston, MA 02215, [2]Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, [3]Department of Biology, University of Michigan, Ann Arbor, MI 48105 and [4]Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA
[5]Corresponding author

Communicated by R.Krumlauf

Although it is likely that a highly complex network of transcription factors acts in concert during mammalian brain development, relatively few such genes have been characterized to date. We describe here a novel murine homeobox gene, denoted *Gtx*, which in adult animals is specifically expressed within glial cells of the central nervous system, including the forebrain, and in germ cells of the testis. *Gtx* resides on chromosome 7 and does not cosegregate with any previously mapped homeobox gene. The amino acid sequence of the predicted protein encoded by *Gtx* is highly divergent from that of any other known homeobox genes. The *Gtx* homeodomain contains unique residues at positions predicted to contact DNA bases. It did not bind to known target sites for other homeobox genes *in vitro* but bound with high affinity to the MEF-2 motif, a binding site for the serum response factor-related proteins. GTX efficiently competed with RSRF to bind the MEF-2 element *in vitro*. Co-transfection of *Gtx* prevented the serum-induced activation of the MEF-2-containing reporter genes. Although the true biological role of *Gtx* is not known, these results suggest that *Gtx* is a novel cell-type specific homeobox gene that has the potential to act as a transcriptional repressor for a subset of serum-inducible genes.

Key words: DNA binding protein/homeobox/MEF-2/RSRF/tissue-specific expression

Introduction

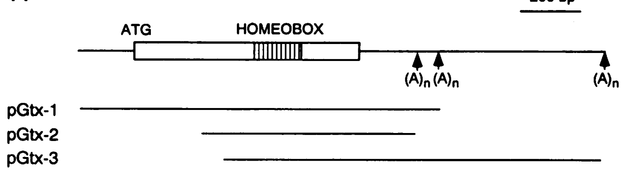
The coordinated activation of regulatory genes is a fundamental mechanism for determining the temporal and spatial patterns of development. The molecular control of differentiation processes depends on lineage-restricted transcription factors that regulate tissue-specific genes. Although much progress in unravelling these regulatory events has been made in *Drosophila* (reviewed in St Johnston and Nüsslein-Volhard, 1992), the hierarchy of genes governing developmental processes, such as lineage commitment and cell type specification, is largely unknown in mammalian development.

Among the genes that govern the development of *Drosophila*, the sequential activation of homeotic and segmentation genes controls the identity, polarity and number of segments (reviewed in Akam, 1987; Scott and Carroll, 1987). Many such genes, including the *Antennapedia (Antp)*, *Engrailed (En)* and *Paired (Prd)* families, contain a characteristic 180 bp sequence motif called the homeobox (reviewed in Scott *et al.*, 1989; Affolter *et al.*, 1990a; McGinnis and Klumrauf, 1992). The carboxyl-terminal portion of the homeobox domain has structural similarity to the helix–turn–helix DNA-binding domain of some regulatory proteins in yeast and prokaryotes (Harrison and Aggarwal, 1990). Recent NMR and X-ray crystallographic studies on the protein–DNA complex of three homeodomains have shown that the amino-terminal arm and the third helix of the homeodomain make specific contacts with bases in the minor and major grooves of DNA, respectively (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991). Thus, homeodomain-containing genes act as sequence-specific transcription factors that transactivate or repress the expression of other genes, although very few targets of homeobox genes have been identified to date (reviewed by Hayashi and Scott, 1990).

Many homologs of invertebrate homeodomain proteins have been identified in mammals, including mice and humans (Scott *et al.*, 1989; McGinnis and Klumrauf, 1992). The best-studied vertebrate homeobox-containing genes are *Antp*-like *Hox* genes, which exist in four major clusters in the mouse genome (Duboule and Dolle, 1989; Graham *et al.*, 1989). Each cluster exhibits intriguing similarities to the complement of genes within the *Drosophila Antp* and *Bithorax* clusters, not only in homeodomain sequences but also in the temporal order of activation, anterior boundary of expression during embryogenesis, and possible role in segmentation (Kessel and Gruss, 1990; McGinnis and Krumlauf, 1992). Ectopic expression of *Hox 1.1* in mouse causes a homeotic transformation in the anterior vertebral column (Kessel *et al.*, 1990), and ectopic overexpression of the mouse *Hox 2.2* and human *HOX 4.2* genes in *Drosophila* mimics in part the effect of overexpression of the homologous genes, *Antp* and *Deformed* (*Dfd*), respectively (Malicki *et al.*, 1990; McGinnis *et al.*, 1990). Furthermore, null alleles of *Hox 1.5* (Chisaka and Capecchi, 1991) and *Hox 1.6* (Lufkin *et al.*, 1991) display morphological defects in head and thorax, reminiscent of loss-of-function mutants of the *Drosophila labial* and *Dfd* genes. These data suggest that the mammalian *Hox*-type homeobox genes may function similarly to the invertebrate homologs during development and that they appear to convey positional information rather than specification of cell-types.

A homeobox gene whose expression is restricted to specific cell lineages would be of particular interest as a

A



B

```
   1     C CGC CGC CCA TTC AGC GCA ACA GCC GTC GGT CCT CTC GCT TTC CCG    46

  47   TAG GGG CCG TCG GCG TTC GTT TGA AAC GCG GTC CAC CCG TCC CAG CGT    94

  95   AGC CGG CGC TCT TCG GCG CCG CGC GCA AAC TTC CCG AGC CGG CGG GTG   142

 143   CGG GCG GTG GCA GCG GGG CCC GGA TGG GCG CCC GGG TCG GAG GCG GCG   190

 191   GCG CCC ATG GAC GCT AAC CGC CCG GGT GCG TTT GTG CTG AGC AGC GCG   238
   1           Met Asp Ala Asn Arg Pro Gly Ala Phe Val Leu Ser Ser Ala    14

 239   CCT TTA GCC GCG CTG CAC AAC ATG GCT GAG ATG AAG ACG TCG CTG TTC   286
  15   Pro Leu Ala Ala Leu His Asn Met Ala Glu Met Lys Thr Ser Leu Phe    30

 287   CCC TAC GCG CTG CAG GGC CCG GCG GGC TTC AAG ACA CCC GCC CTA GGC   334
  31   Pro Tyr Ala Leu Gln Gly Pro Ala Gly Phe Lys Thr Pro Ala Leu Gly    46

 335   AGC CTT GGC GCG CAG TTG CCT CTA GGC ACT CCG CAC GGC ATC AGC GAC   382
  47   Ser Leu Gly Ala Gln Leu Pro Leu Gly Thr Pro His Gly Ile Ser Asp    62

 383   ATC CTG GGA CGG CCG GTG GGC GCA GCG GGT GGC GGC CTC CTA GGA AGT   430
  63   Ile Leu Gly Arg Pro Val Gly Ala Ala Gly Gly Gly Leu Leu Gly Ser    78

 431   CTG CCC CGT CTC AAC GGG CTC GCC TCG TCT GCA GGT GTC TAC TTC GGG   478
  79   Leu Pro Arg Leu Asn Gly Leu Ala Ser Ser Ala Gly Val Tyr Phe Gly    94

 479   CCC GCA GCC GCC GTG GCT CGG GGC TAC CCC AAG CCG CTG GCG GAA CTG   526
  95   Pro Ala Ala Ala Val Ala Arg Gly Tyr Pro Lys Pro Leu Ala Glu Leu   110

 527   CCT GGG CGC CCG CCC ATC TTC TGG CCT GGG GTG GTG CAG GGC TCT CCC   574
 111   Pro Gly Arg Pro Pro Ile Phe Trp Pro Gly Val Val Gln Gly Ser Pro   126

 575   TGG AGG GAC CCG CGA CTG GCC GGC TCC GCC CAA GCC GGC GGG GTC CTG   622
 127   Trp Arg Asp Pro Arg Leu Ala Gly Ser Ala Gln Ala Gly Gly Val Leu   142

 623   GAT AAG GAT GGC AAG AAG AAA CAC TCG CGG CCG ACT TTC TCC GGC CAG   670
 143   Asp Lys Asp Gly Lys Lys Lys His Ser Arg Pro Thr Phe Ser Gly Gln   158

 671   CAG ATC TTC GCG CTG GAG AAG ACT TTC GAG CAG ACC AAG TAT TTG GCA   718
 159   Gln Ile Phe Ala Leu Glu Lys Thr Phe Glu Gln Thr Lys Tyr Leu Ala   174

 719   GGC CCA GAG CGC GCG CGG CTT GCC TAC TCT CTG GGC ATG ACC GAG AGC   766
 175   Gly Pro Glu Arg Ala Arg Leu Ala Tyr Ser Leu Gly Met Thr Glu Ser   190

 767   CAA GTG AAG GTG TGG TTC CAG AAT CGG CGG ACC AAG TGG CGC AAG CGG   814
 191   Gln Val Lys Val Trp Phe Gln Asn Arg Arg Thr Lys Trp Arg Lys Arg   206

 815   CAC GCG GCA GAG ATG GCG TCG GCT AAA AAG AAG CAA GAC TCG GAT GCC   862
 207   His Ala Ala Glu Met Ala Ser Ala Lys Lys Lys Gln Asp Ser Asp Ala   222

 863   GAG AAG CTG AAG GTG GGT GGC TCA GAC GCG GAG GAC GAT GAC GAA TAC   910
 223   Glu Lys Leu Lys Val Gly Gly Ser Asp Ala Glu Asp Asp Asp Glu Tyr   238

 911   AAC CGG CCC CTG GAC CCC AAC TCC GAT GAC GAG AAG ATC ACG CGG CTT   958
 239   Asn Arg Pro Leu Asp Pro Asn Ser Asp Asp Glu Lys Ile Thr Arg Leu   254

 959   CTC AAA AAG CAC AAA CCC TCG AAC TTG GCG CTC GTT AGC CCG TGT GGT  1006
 255   Leu Lys Lys His Lys Pro Ser Asn Leu Ala Leu Val Ser Pro Cys Gly   270

1007   GGC AGC GCG GGG GAC GCC TTG TGA GGA TGC GGC CAG GCC GGA GAA CCC  1054
 271   Gly Ser Ala Gly Asp Ala Leu                                      277

1055   GAG AAC CGG GAC TCG CGG CAT GCC CCG ACG CCA GCC GCC CAG CCG CAG  1102

1103   TGT GTA TAT ATA TTT TTA CAG AAT AAG TTA TAA AGC GGA CGT TGG CGC  1150

1151   GGC CTT GGC GTG ATG GCG GAG TAC GGG GTT TGG GCC GAT CAC TTT GTA  1198

1199   TAA TCA ATA AAT TAT TTA ACA CGT CCT CGT CGG AGC CGT GGC TCC C(A)n 1244
```

C

```
                  HELIX 1         HELIX 2        HELIX 3
a.a.#  1         10        20        30        40        50        60
         ● ●   o                   o          o            o  ●o  ● ● o●o  o   %homology
  Gtx  KKHSRPTFSG QQIFALEKTF EQTKYLAGPE RARLAYSLGM TESQVKVWFQ NRRTKWRKRH
HOX11  --KP-TS-TR L--CE---R- HRQ----SA- --A--KA-K- -DA---T--- -------RQT  58%
  NK3  --R--AA--H A-V-E--RR- A-QR--S--- -SEM-K--RL --T---I--- ---Y-TKRKQ  55
Chox3  PRRA-TA-TY E-LV---NK- RA-R--SVC- -LN--L--SL --T---I--- ------K-Q-  53
  NK1  PRRA-TA-TY E-LVS--NK- KT-R--SVC- -LN--L--SL --T---I--- ------K-QN  50
 H2.0  RSW--AV--N L-RKG--IQ- Q-Q--ITKPD -RK--AR-NL -DA------- ---M---HTR  50
  NK4  -RKP-VL--Q A-VLE--CR- RLK---T-A- -EII-QK-NL SAT---I--- ---Y-SKRGD  43
  NK2  -RKR-VL-TK A-TYE--RR- R-QR--SA-- -EH--SLIRL -PT---I--- -H-Y-TKRAQ  42

 Antp  R-RG-Q-YTR Y-TLE---E- HFNR--TRRR -IEI-HA-CL --R-I-I--- ---M--K-EN  43
  Prd  QRRC-T---A S-LDE--RA- -R-Q-PDIYT -EE--QRTNL --ARIQ---S ---ARL--Q-  42
  Eve  VRRY-TA-TR D-LGR---E- YKEN-VSR-R -CE--AQ-NL P--TI----- ---M-DKRQR  40
  Bcd  PRRT-T--TS S--AE--QH- L-GR--TA-R L-D-SAK-AL GTA---I--K ---RRHKIQS  38
   En  E-RP-TA--S E-LAR-KRE- NENR--TERR -QQ-SSE--L N-A-I-I--- -K-A-IK-ST  38
  Ftz  S-RT-Q-YTR Y-TLE---E- HFNR-ITRRR -IDI-NA-SL S-R-I-I--- ---M-SK-DR  38
  Cut  S-KQ-VL--E E-KE--RLA- ALDP-PNVGT IEF--NE--L ATRTITN--H -H-MRLKQQV  28
 Oct1  RRKK-TSIET NIRV----S- LENQKPTSE- ITMI-DQ-N- EKEVIR---C ---Q-EKRIN  28

CONS.   KR R  Y    Q   LE  F     Y        R   A  L L      QIKIWFQ NRR K K
        RK    F                             S            V V          R R
```

candidate for a 'cell-type specification' gene. The POU (*Pit-1, Oct-1, Unc 86*) gene family has a divergent homeodomain from that of *Hox* class. Several members of the POU family are expressed in the developing central nervous system (CNS) and show restricted expression patterns in the adult brain (He *et al.*, 1989). The *Pit-1* gene, which is expressed in the anterior pituitary gland, is necessary for formation of pituitary cells and activates pituitary-specific genes (Ingraham *et al.*, 1988; Castrillo *et al.*, 1991).

Drosophila genes belonging to the *NK* family (Kim and Nirenberg, 1989) have homeodomains divergent from the *Antp, En, Prd* or *Even-skipped* classes and seem to specify cell fates in specific tissues. Expression of the *S59 (NK-1)* gene is restricted to the somatic mesoderm, to subsets of CNS and to a small region of the midgut (Dohrmann *et al.*, 1990). The *msh-2(NK-4)* gene is expressed in all mesoderm cells in the segmented part of the embryo during germband elongation, but soon afterwards its expression becomes restricted to the dorsal mesoderm (Bodmer *et al.*, 1990). A rat homeobox-containing gene, *TTF-1*, which is closely related to *Drosophila NK-2*, is expressed in the thyroid and lung anlage and in restricted neuroblast populations, suggesting that *TTF-1* may have a role in cell differentiation in these three systems (Guazzi *et al.*, 1990; Lazzaro *et al.*, 1991). It is possible that the homeobox-containing genes which have sequences divergent from the *Hox* class genes and are expressed in restricted tissues may play important roles in tissue differentiation.

In order to search for new homeobox genes that may play a role in tissue differentiation, we screened a mouse embryonic cDNA library using a *Drosophila* homeodomain probe whose sequence is divergent from the *Hox* class genes. We report here the isolation and characterization of a novel murine gene whose homeodomain sequence differs significantly from that of any other known homeobox genes. In adult animals, this gene appears to be expressed exclusively in glial cells of the CNS including cerebrum, midbrain, cerebellum and spinal cord, and germ cells of the testis. To indicate this restricted expression, this new gene is denoted *Gtx*, for Glial- and testis-specific homeobox gene. *Gtx* resides on the distal portion of mouse chromosome 7 and does not cosegregate with any previously mapped homeobox genes. The homeodomain of *Gtx* does not bind to known binding sites for *Antp, En, Prd* or *Bicoid* classes of the homeobox genes (Hayashi and Scott, 1990). Instead, GTX binds to the MEF-2 motif (Gossett *et al.*, 1989), which is a binding site for the 'serum response factor-related protein' (RSRF), a recently isolated transcription factor of the 'MADS box' family (Pollock and Treisman, 1991; Yu *et al.*, 1992). *In vitro* translated GTX efficiently competed with RSRF to bind the MEF-2 element. Interestingly, co-transfection of *Gtx* with a reporter construct under the control of the MEF-2 site prevented the serum-induced activation of the reporter gene.

Results

Isolation of a novel homeobox cDNA, Gtx

We screened a λgt10 cDNA library prepared from 8.5 day mouse embryos (Fahrner *et al.*, 1987) under low stringency to isolate new homeobox-containing genes. The homeobox region of *Drosophila msh2 (NK-4)* gene (Bodmer *et al.*, 1990) was used as a hybridization probe because this gene has a homeodomain sequence very divergent from that of known homeobox genes. We isolated a new murine homeobox-containing cDNA, hereafter referred to as *Gtx*. The initial *Gtx* clone contained a 0.8 kb insert (pGtx-2, Figure 1A). To obtain a full-length *Gtx* cDNA, a mouse neonatal brain cDNA library was screened and several hybridizing clones were obtained. The sequencing of three overlapping cDNA clones revealed identical coding sequence and three alternative polyadenylation sites (Figure 1A).

The *Gtx* cDNA sequence has an open reading frame that is predicted to encode a polypeptide of 277 amino acids (Figure 1B). There are two methionine codons in-frame with the predicted homeobox domain. Both ATG codons are in a favorable context for the Kozak consensus sequence. In most cases, the first ATG that conforms to Kozak consensus is the predominant initiation site (Kozak, 1987); therefore we assigned the first ATG as the translation start site. An in-frame stop codon is located 129 nt upstream of the first ATG codon.

The *Gtx* homeodomain shows a very limited homology (30−50%) to most homeodomain sequences and its similarity to *Antp* and *En* are 43 and 38%, respectively (Figure 1C). The *Gtx* homeodomain is more related to the homeodomain of human *HOX*11 (58%) and *Drosophila NK*-3 (55%). However, *Gtx* is unlikely to be the mouse homolog of *HOX*11 or *NK* genes, because the sequence similarities between *Gtx* and *HOX*11 or *NK* family members are less than 60%. Sequence similarities of the homeodomains between homologous genes usually exceed 80−90% even among most divergent species (Scott *et al.*, 1989).

The amino- and carboxyl-terminal regions of the protein encoded by *Gtx* show no significant sequence homology to any of the known homeobox containing genes. However, the amino-terminal region of the protein is very rich in glycine and proline; of residues 36−94, 25% are glycine, and of residues 95−130, 25% are proline (Figure 1B). The glycine- and proline-rich amino-terminal domain could function either in transcriptional regulation or in interaction with other proteins required for efficient DNA target site selection *in vivo* (Mitchell and Tjian, 1989; Mermod *et al.*, 1989). The carboxyl-terminal region of *Gtx* is rich in charged

Fig. 1. Nucleotide and deduced amino acid sequence of *Gtx*. (**A**) Schematic diagram of the *Gtx* cDNAs. The open box indicates the coding region and solid bars the untranslated regions. Three polyadenylation sites are shown. (**B**) Nucleotide and amino acid sequence of pGtx-1 cDNA. The predicted amino acid sequence is presented under the nucleotide sequence. The large box indicates the location of the homeodomain. Proline and glycine residues found in the region preceding the homeobox are marked with solid and dashed lines, respectively. The region following the homeodomain is rich in acidic amino acids (boxed) and basic amino acids (bold lines). The putative first polyadenylation signal is shown in bold. The nucleotide sequence is available from EMBL/GenBank under accession number L08074. (**C**) Comparison of the GTX homeodomain with other known homeobox-containing genes. Various known homeobox sequences are compared with the GTX homeobox sequence. Amino acid identity with GTX is indicated by a dash and the percent identity with GTX is given at the right. The consensus is derived from known homeoboxes (Scott *et al.*, 1989). Symbols on the top represent putative contact sites with DNA bases (closed circles) and with DNA sugar−phosphate backbone (open circles), based on the crystal structure of the DNA−homeodomain complex of *En* and *MAT*a2 (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991). The boundaries of three helixes are also based on the reported crystal structure of the homeodomain.

amino acids. Of amino acids 215−259, 27% are basic and 33% are acidic (Figure 1B).

The Gtx gene is localized in mouse chromosome 7

To determine whether there are any other genes closely related to *Gtx* in the mouse genome, Southern blot analysis was performed using mouse genomic DNA. Under high stringency, the pGtx-1 cDNA probe detected a single major band and a few weak bands in each lane, suggesting that *Gtx* is likely to exist as a single copy and that there may be some *Gtx*-related genes in the mouse genome (Figure 2A). When a *Gtx* cDNA was hybridized at high stringency with genomic DNAs from a variety of animal species, a distinctive hybridization was observed in many species including human, monkey, rat, dog, cow, rabbit and chicken (data not shown), indicating that *Gtx* is an evolutionarily conserved gene in vertebrates.

The chromosomal location of *Gtx* in the mouse was determined by interspecific backcross analysis using progeny derived from matings of [(C57BL/6J × *Mus spretus*) F_1 × C57BL/6J] mice (Figure 2B). This interspecific backcross mapping panel has been typed for over 1000 loci that are well distributed among all the autosomes as well as the X chromosome (Copeland and Jenkins, 1991). C57BL/6J and *M.spretus* DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) using a full length *Gtx* probe. A 8.2 kb *M.spretus* *Hin*cII RFLP (see Materials and method) was used to follow the segregation of the *Gtx* locus in backcross mice. The mapping results indicated that *Gtx* is located in the distal region of the mouse chromosome 7 linked to *Pkcb, Il-4r, Igf-2* and *Int-2*. Although 138 mice were analyzed for every marker and are shown in the segregation analysis (Figure 2B), up to 180 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order are: centromere, *Pkcb*-(3/180)-*Il-4r*-(13/157)-*Gtx*-(2/156)-*Igf-2*-(0/179)-*Int-2*. The recombination frequencies [expressed as genetic distances in centiMorgans (cM) ± the standard error] are *Pkcb*-(1.7 ± 1.0)-*Il-4r*-(8.3 ± 2.2)-*Gtx*-(1.3 ± 0.9)-[*Igf-2, Int-2*]. No recombinants were detected between *Igf-2* and *Int-2* in 179 animals typed in common, suggesting that the two loci are within 1.7 cM of each other (upper 95% confidence limit).

Expression of Gtx in glial cells of the CNS

To examine the tissue distribution of *Gtx* gene expression, total cellular RNA, prepared from various tissues of 14 day embryos, as well as 2 and 12 day postnatal and adult mice, was analyzed by Northern hybridization (Figure 3A). *Gtx* expression is tissue-specific and developmentally regulated. A major transcript of *Gtx* (~1.5 kb in size) was observed only in the brain (cerebrum and midbrain), cerebellum, spinal cord and testis. In the brain, *Gtx* was more abundantly expressed in the adult than in embryos or neonates. In addition to the 1.5 kb transcripts, two larger transcripts were observed in CNS and testis, which may correspond to the cDNAs with different polyadenylation sites (Figure 1A) or to alternatively spliced transcripts. *Gtx* transcripts were not

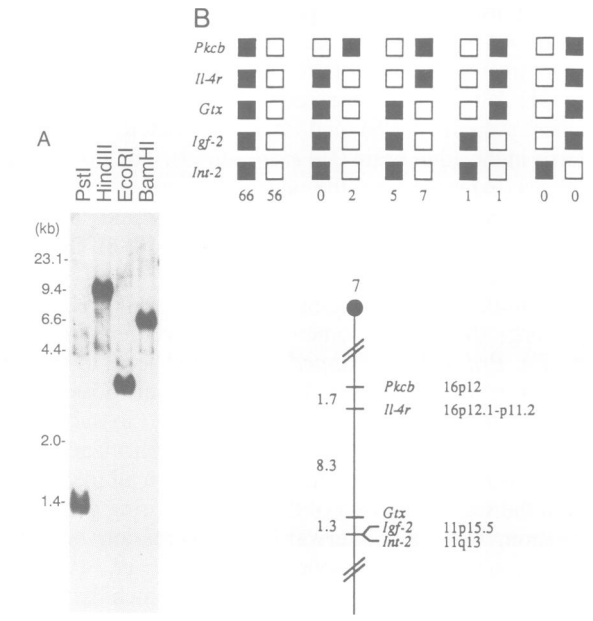


Fig. 2. Genomic analysis of the mouse *Gtx* gene. (**A**) Southern blot analysis. Twenty micrograms of mouse genomic DNA was digested with *Bam*HI, *Eco*RI, *Hin*dIII and *Pst*I and electrophoresed in 0.8% agarose gels. After Southern blotting, a reinforced nitrocellulose filter was hybridized at high stringency with the full-length *Gtx* cDNA probe. The final wash of the filter was in 0.1 × SSC and 0.1% SDS at 65°C. (**B**) *Gtx* maps in the distal region of mouse chromosome 7. The location of *Gtx* was mapped by interspecific backcross analysis. The segregation patterns of *Gtx* and flanking genes in 138 backcross animals that were typed for all loci are shown at the top of the figure. For individual pairs of loci, more than 138 animals were typed (see text). Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M.spretus*) F_1 parent. The filled boxes represents the presence of a C57BL/6J allele and white boxes represent the presence of a *M.spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 7 linkage map showing the location of *Gtx* in relation to linked genes is shown at the bottom of the figure. Recombination distances between loci in centimorgans are shown to the left of the chromosome and the positions of loci in human chromosomes, where known, are shown to the right. References for the human map positions of loci mapped in this study can be obtained from the GDB (Genome Data Base), a computerized database of human linkage information maintained by the William H.Welch Medical Library of The Johns Hopkins University (Baltimore, MD).

detected in other adult tissues (including heart, kidney, liver, lung, skeletal muscle, ovary, skin, intestine and spleen) or in cultured cell lines, such as HeLa, COS or C2 muscle cells, or in PC12 cells (Figure 3 and data not shown).

To localize *Gtx* expression sites better, *in situ* hybridization analysis was performed (Figure 3B). In day 14 embryos, *Gtx* transcripts were located lateral to the midline on a horizontal section of the midbrain where only a few cells were moderately labeled (not shown). In 18 day embryos, more abundant expression was detected primarily in the spinal cord, medulla oblongata and mesencephalon (Figure 3Ba, arrow heads). In addition, the labeling was detected in a more rostral brain region, most probably the hypothalamic area (thin arrow in Figure 3Ba). Transient expression was also observed in the mucosa of colon (thick arrow in Figure 3Ba and d) and in the developing glomeruli of the kidney in 18 day embryos (Figure 3Bc and e, thin arrows). In adult mice, however, no transcripts were detected in colon or kidney (data not shown). There were no specific

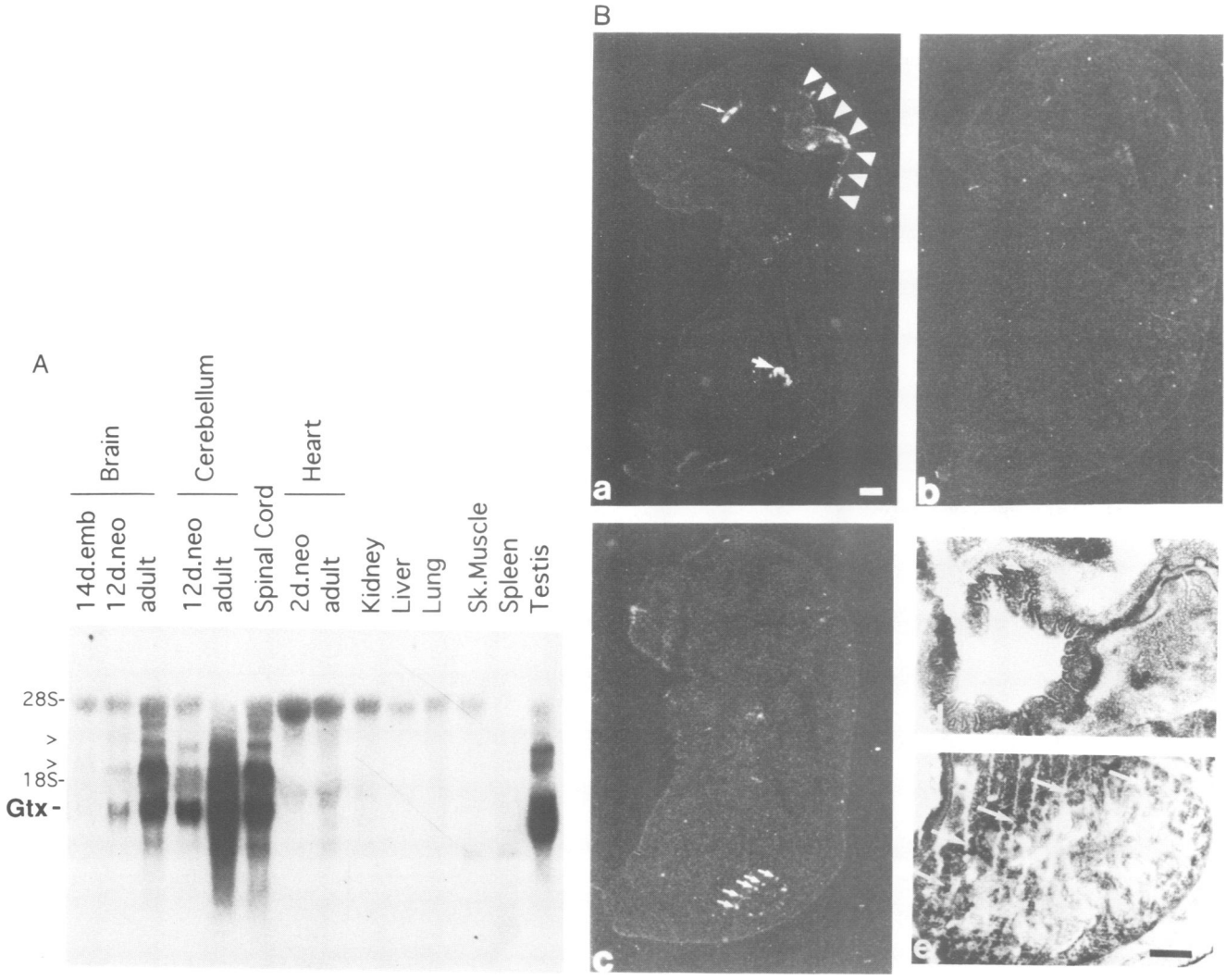


Fig. 3. Developmental and tissue-specific expression of *Gtx*. (**A**) Northern blot analysis of *Gtx* mRNA from various mouse tissues. Twenty micrograms of total RNA isolated from each tissue was loaded in each lane and hybridized with a riboprobe generated from pGtx-1. The positions of 28S and 18S rRNA are indicated (the riboprobe used cross-hybridized with 28S and 18S rRNA). The major transcript (~1.5 kb in size) is marked by Gtx and two other larger transcripts are marked by arrowheads. The quality of RNA and equal loading in each lane were confirmed by ethidium bromide staining of the gel (not shown). (**B**) *In situ* hybridization of whole mouse embryos. Darkfield autoradiographs of 18 day old mouse embryo sectioned close to the midline (a and b) and at a more lateral level (c) are shown. (**a**) Distinctive labeling of the spinal cord, medulla oblongata and mesencephalon (arrowhead) is observed with a *Gtx* antisense probe. In addition, a small area just below the thalamus, most probably the developing hypothalamus, is intensely labeled (thin arrow). Expression was also observed in mucosa of the colon (thick arrow), where it appeared to be confined to the epithelial part of the mucosa as observed by hematoxylin–eosin staining of the same section (**d**, thick arrows). (**b**) No labeling is seen in a serial section using a *Gtx* sense probe. (**c**) Weak but distinct labeling is observed in a dot-like pattern in the periphery of the kidney parenchyma (thin arrows) with the antisense probe. This pattern is consistent with labeling of developing glomeruli as determined by hematoxylin–eosin staining of the same section (**e**, thin arrows). Bars in (a) and (e) indicate 1 mm and 0.4 mm respectively.

signals when sense probes were used for hybridization (Figure 3Bb).

In the adult brain (Figure 4), the expression of *Gtx* was concentrated in areas with high nerve fiber content. The message was particularly abundant in the adult corpus callosum (short arrows in Figure 4a, c, e and g), anterior commissure (arrow heads in Figure 4a and c), internal capsule (Figure 4e, arrow head), thalamic and hypothalamic areas (Figure 4e and g), brain stem (Figure 4i), and spinal cord (not shown). In the cerebellum, labeling was most abundant over the white matter fiber tracts (Figure 4i, arrow heads). In addition, a low level of message was detected in most brain regions (Figure 4a, c, e, g and i). Using emulsion autoradiography under a higher power magnification, the intense labeling was located overlying individual cell bodies in the corpus callosum and in the white matter of the cerebellum (data not shown).

The pattern of distribution of *Gtx* transcripts is consistent with the expression in glial cells rather than in neurons. In order to identify what types of glial cells express GTX protein, a primary culture of glial cells was prepared from neonatal rat brain for immunochemical analysis. In this glial culture, ~70% of the cells reacted with an antibody against glial fibrillary acidic protein (GFAP), indicating that they are astroglia. Double labeling of cells with an anti-GTX (Figure 5, left panel) and anti-GFAP (Figure 5, right panel) showed that both GFAP-positive and GFAP-negative cells showed nuclear staining by the anti-GTX antibody. These GFAP-negative and GTX-positive cells reacted with an anti-galactocerebroside antibody, which specifically recognizes

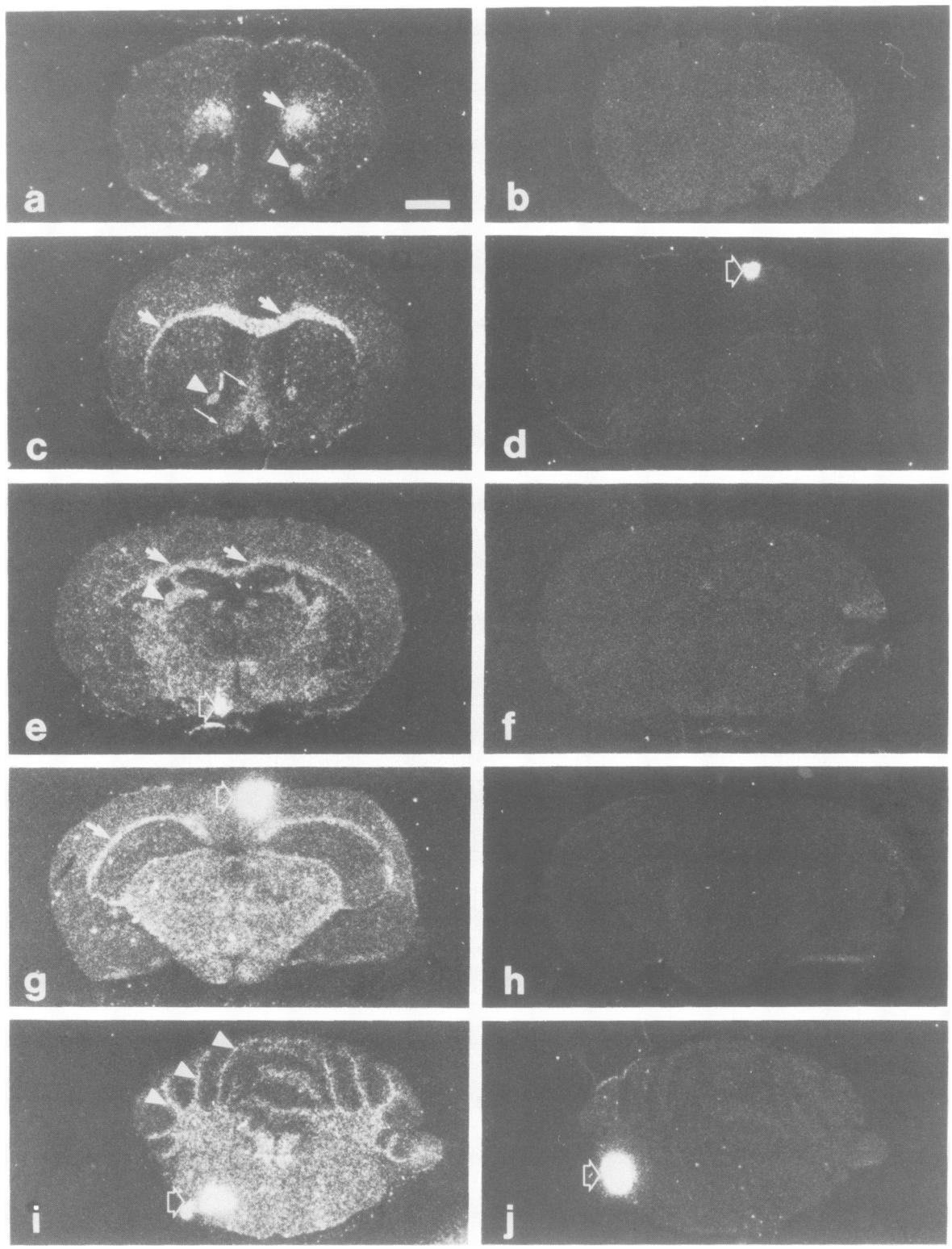


Fig. 4. *In situ* hybridization of adult mouse brain. Darkfield autoradiographs of a series of sections throughout the adult mouse brain hybridized with antisense (**a**, **c**, **e**, **g** and **i**) and sense (**b**, **d**, **f**, **h** and **j**) *Gtx* oligonucleotide probes. Sections are spaced ~2 mm apart except between g and i, where the distance is larger. (a) Intense labeling is present over the forceps minor of the corpus callosum (short arrow) and the anterior part of the anterior commissure (arrowhead). (c) The corpus callosum (short arrows), the anterior commissure (arrowhead) as well as the nuclei of the diagonal band (thin arrows) exhibit strong labeling. (e) Labeling is present over the corpus callosum (short arrows), the internal capsule (arrowhead) and generally in the hypothalamic area and the median forebrain bundle. In (g), the corpus callosum (short arrow) and the entire thalamic and hypothalamic region express high mRNA levels whereas the hippocampus and all cortical regions have low mRNA levels. (i) Labeling is present over the entire medulla oblongata and extending into the spinal cord (not shown). In the cerebellum, labeling is concentrated to the fiber tracts of the white matter (arrowheads), whereas only weak labeling is seen over all three cortical layers. No labeling was detected when the sense oligo was hybridized to adjacent sections (b, d, f, h and j). Open arrows indicate silver grain aggregation artifacts. Bar in (a) indicates 1 mm.

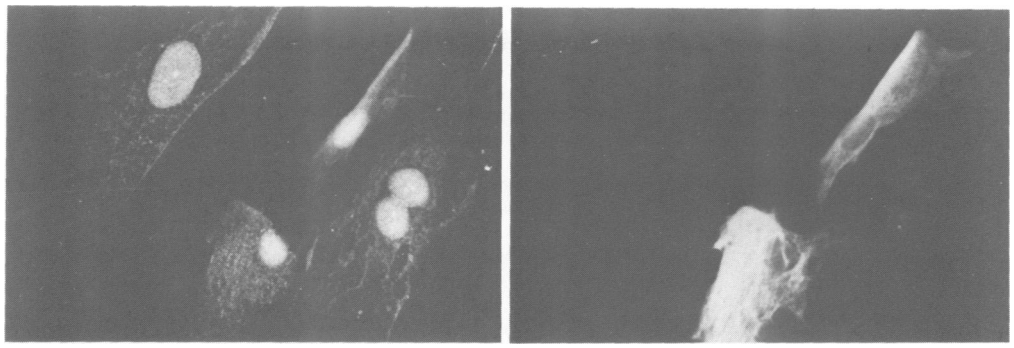


Fig. 5 Expression of GTX protein in cultured glial cells. Double immunofluorescent staining of the same field was performed using antibodies directed against GST-GTX fusion protein (left panel) and GFAP (right panel). Note that both astroglia (GFAP-positive cells) and oligodendroglia (GFAP-negative cells) display a nuclear staining pattern with the anti-GTX antibody (left panel).

oligodendrocytes (data not shown). In contrast, cells reacted with anti-neurofilament antibody did not show nuclear staining by the anti-GTX antibody (data not shown). These results indicate that the GTX protein is expressed in both astroglia and oligodendrocytes but not in neuronal cells in culture.

Gtx is specifically expressed in male germ cells

In the testis, no expression was detected at 1 or 12 days of age (Figure 6a). At day 17, *Gtx* RNA was first detected in a subset of tubuli, with expression limited to the inner two-thirds of the tubular wall (Figure 6c, arrows). At 28 days (Figure 6e), all tubules were labeled and this pattern remained in the adult testis (Figure 6g, see panel h for a low magnification). No labeling could be detected overlying Leydig cells or over the most peripheral layer of the tubules (Figure 6g and h).

Mutation at the *W* locus on mouse chromosome 5 causes absence of gonadal germ cells that can be traced back to primordial germ cells (Mintz and Russell, 1957). The testes of W/W^v compound heterozygous males are normal with respect to their somatic cells but severely deficient in germ cells. Compared with the intense labeling by *Gtx* probe in the normal testis (Figure 6h, left), the testis from W/W^v mice that lack germ cells showed no significant grain accumulation above background (thick arrows in Figure 6h). Sections adjacent to those shown in Figure 6g and h both hybridized well with a probe for WT-1 (Pelletier *et al.*, 1991) which is expressed by Sertoli cells (data not shown). Thus, *Gtx* is specifically expressed in germ cells of the testis.

The Gtx homeodomain binds to MEF-2-like sequences with high affinity in vitro

Proteins encoded by homeobox-containing genes act as transcription factors in which the homeodomain is responsible for sequence-specific recognition of DNA (reviewed in Hayashi and Scott, 1990). We attempted to identify GTX homeodomain target sequences *in vitro* by the method of 'selected and amplified binding' (SAAB) which utilizes repeated cycles of electrophoretic mobility shift assay (EMSA) and PCR (Blackwell and Weintraub, 1990). The homeobox region of *Gtx* was subcloned into the pGEX-3X vector (Smith and Johnson, 1988) and glutathione-S-transferase fusion protein (GST-GTX) was produced in *Escherichia coli* and purified by glutathione−sepharose chromatography. GST-GTX was incubated with a radio-labeled mixture of highly degenerate (4^{15}-fold) synthetic oligomers (51mer), in which an internal 15 bp stretch was completely random in sequence. After the third cycle of SAAB, a shifted band was observed (data not shown). DNA was extracted from the shifted band and cloned into a pBluescript vector. After transformation, plasmid DNA was prepared from 50 independent colonies. DNA from 48 of these colonies bound to the GST-GTX fusion protein in EMSA analysis. DNA sequencing of these 48 clones, however, showed no clear consensus sequence (I.Komuro, and S.Izumo, unpublished results), probably because the SAAB was done only for three cycles in conditions of protein excess.

In order to select a high affinity binding site for the *Gtx* homeodomain from these sequences, we performed an EMSA using different concentrations of GST-GTX (0.1, 0.3, 1, 3 and 10 nM). Of the 48 clones screened, clone 44, which contained a stretch of 11 A or T residues (Figure 7C, top), bound to GTX with the highest apparent affinity (data not shown). A search for homologous sequences in the GenBank database revealed that the sequence of clone 44 was nearly identical to a consensus binding site called MEF-2 (Figure 7C), originally identified in the enhancer of the muscle creatine kinase (MCK) gene (Gossett *et al.*, 1989) as well as in many muscle genes and in serum-inducible genes (Pollock and Treisman, 1991 and references therein).

To confirm that GTX binds to the MEF-2 site, an oligonucleotide corresponding to the MCK enhancer was synthesized and tested by EMSA. Specific binding of GST-GTX was detected with as little as ~0.1 nM of the protein (Figure 7A). A 100-fold molar excess of cold oligonucleotide containing the MEF-2 site completely abolished the binding (Figure 7A, right), while oligonucleotides containing non-specific sequences did not (data not shown). GST-GTX also bound to the MEF-2-like sequences of myosin light chain (MLC) 1/3 (Donoghue *et al.*, 1988), MLC2 (Braun *et al.*, 1989) and brain creatine kinase (CKB) (Hobson *et al.*, 1988) (Figure 7B and C). The concentrations of the fusion protein necessary to bind to these MEF-2-related oligomers were 0.1−0.3 nM in all cases (data not shown). Binding of GTX to the MEF-2 motif is sequence-specific, because GTX bound to the mutated MCK MEF-2 sequence, mt-4 (CTATAAATAA), but not to mt-6 (CTAAACATAAA) (data not shown); this is similar to the DNA binding specificity of the bona fide MEF-2 activity in muscle nuclear extracts (Cserjesi and Olson, 1991).

In contrast, GTX did not bind to BS-2 or $(TAA)_5$ sequences (the known binding sites of many *Antp*-type

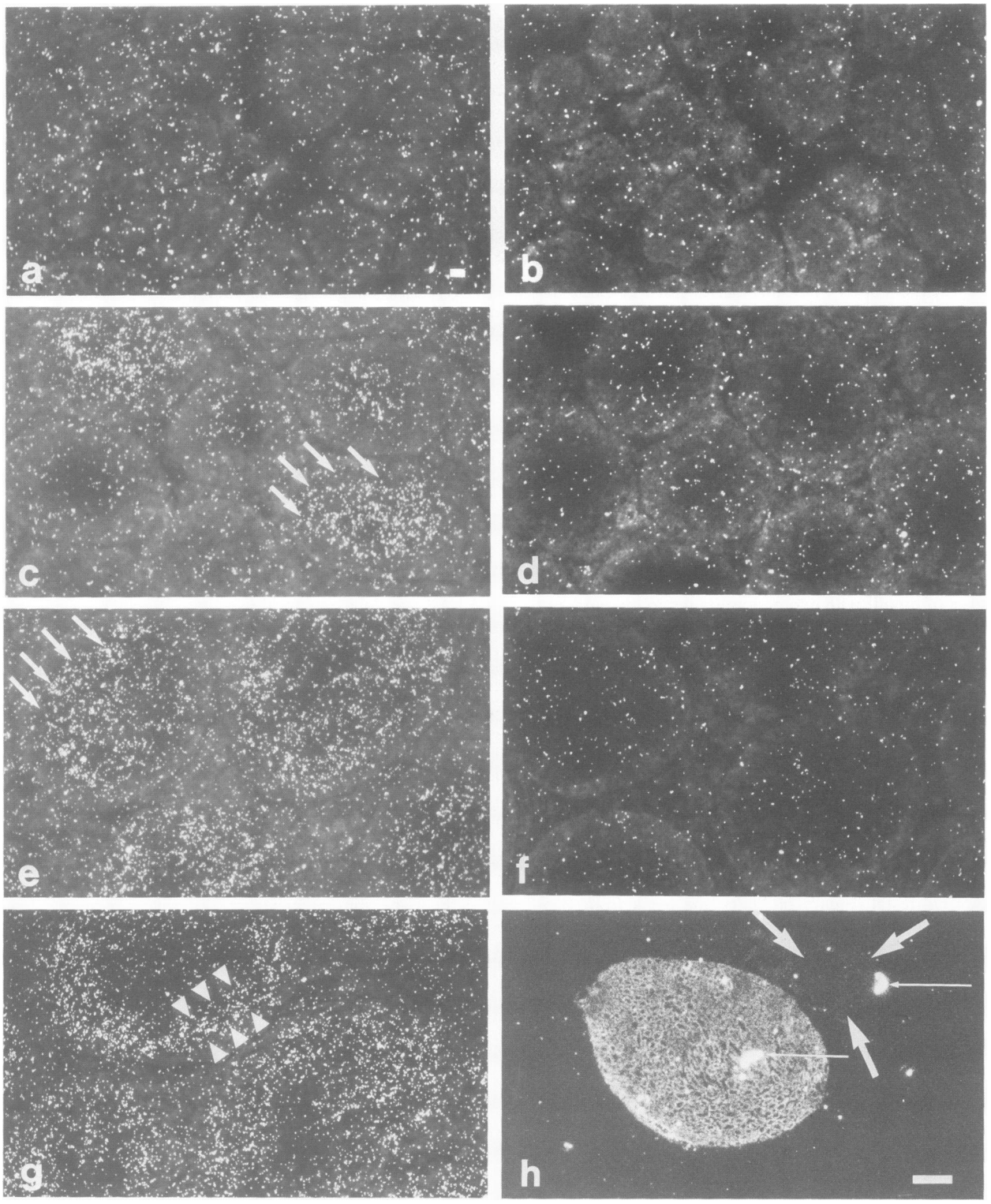


Fig. 6. *In situ* hybridization of mouse testis. Emulsion autoradiography of sections from murine testis following *in situ* hybridization with *Gtx* antisense (a, c, e, g and h) and sense probes (b, d and f), respectively. Mice were sacrificed at day 12 (a and b), 17 (c and d) and 28 (e and f). (g) represents an adult and (h) is an age-matched adult W/W^v mutant mouse. (**a** and **b**) No labeling is observed with either probe. (**c**) Intense labeling is present over some (arrows) but not all tubules. Grains cluster over more central parts of tubules, excluding the basal one-third of the tubular wall and the interstitial space. (**d**) No labeling is detected with the sense probe in an adjacent section. (**e**) At 28 days, all tubules are labeled with grains overlying most of the tubular wall (arrows) excluding only the most peripheral layer. (**f**) No labeling is seen with the sense probe. (**g**) Labeling in the adult is similar to that of 28 day old in (e). (**h**) The adult testis (same as in g) and an age-matched W/W^v mutant testis (thick arrows) are shown in a low magnification. Intense labeling by *Gtx* probe is present only over the normal testis whereas W/W^v is devoid of labeling. Thin arrows denote artifacts. Bars in (a) and (h) indicate 10 μm and 1 mm, respectively.

homeoproteins) (Beachy *et al.*, 1988; Muller *et al.*, 1988), the CArG box/SRE (Treisman, 1990), the TTF-1 site (the target of the mouse NK-2-like homeoprotein) (Guazzi *et al.*, 1990), or the E box (Blackwell and Weintraub, 1990) (Figure 7B and C). GTX bound weakly to the NP sequence (Figure 7B), a well characterized binding site for many types of

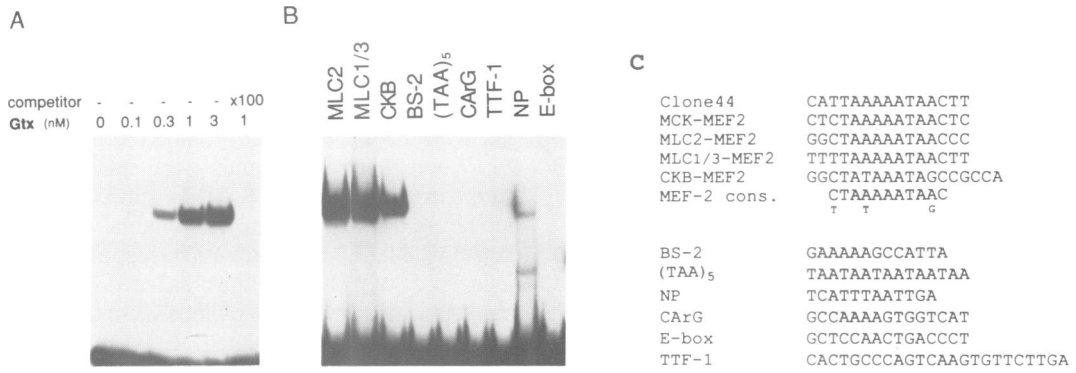


Fig. 7. GTX binds to MEF-2 motif but not to other homeobox target sequences. (**A**) GST-GTX protein at the indicated concentrations was incubated with ^{32}P-labeled oligonucleotides containing the MCK MEF-2 sequence. Note that 50% of maximum binding (EC_{50}) occurs at sub-nanomolar concentrations of GTX. Addition of a 100-fold molar excess of cold MCK MEF-2 oligos abolished the shifted band. (**B**) EMSA using GST-GTX (3 nM) and various oligos are shown in panel C. Note that GTX binds strongly to MEF-2-containing oligos and very weakly to NP but not to other sequences.

homeodomain, including *Antp, Ubx, Ftz, Zen, En, Dll, Nk-1, Eve* and *Mix-1* (Hayashi and Scott, 1990; Treisman *et al.*, 1992), but its apparent affinity was 50- 100-fold lower than that of the MEF-2 site.

DNase I footprinting analysis showed that the GST-GTX protein generated a footprint on both strands of the CKB promoter (Figure 8). The footprints span a region of ~12 bp centered on a CTATAAA sequence (the lower strand), which is the core of MEF-2 binding sequence (Gossett *et al.*, 1989).

GTX competes with RSRF for binding to the MEF-2 sequence

Recently, the genes encoding transcription factors related to RSRF have been isolated (Pollock and Treisman, 1991; Yu *et al.*, 1992). RSRF proteins, which contain a DNA binding motif called the 'MADS box' (Schwartz-Sommer *et al.*, 1990; Treisman and Amerer, 1992), have also been shown to bind to MEF-2 and its related sequences with high affinity (Pollock and Treisman, 1991; Yu *et al.*, 1992). We therefore examined whether RSRF and GTX can compete with each other for DNA binding. *In vitro* translated C4, a member of RSRF family, bound to the CKB MEF-2 motif (Figure 9, lane 4), and the excess cold probe competed the binding (lane 5). Addition of increasing amounts of *in vitro* translated GTX protein efficiently competed with RSRF-C4 for binding to the MEF-2 sequence (lanes 6–9), while addition of a 10-fold excess of unprogrammed reticulocyte lysates did not diminish C4–DNA complex (not shown). A similar competition for DNA binding was observed when GST-GTX was added to *in vitro* translated RSRF (lanes 11 and 12), but not when 100 nM of GST protein (without GTX) was added (lane 10). Thus, the MEF-2 sequence can be a high affinity binding site for two different classes of DNA binding proteins *in vitro*: GTX, a homeodomain-containing protein, and RSRF, a MADS box-containing protein.

Gtx transrepresses expression of the MEF-2 containing promoters

At present, we do not know whether the MEF-2 motif is a real target for GTX *in vivo*. However, CKB and *Gtx* are both expressed in the brain, and the MEF-2-like sequence of the CKB promoter is an important *cis*-element for its expression (Horlick *et al.*, 1990). Therefore, we carried out co-transfection experiments using reporter plasmids

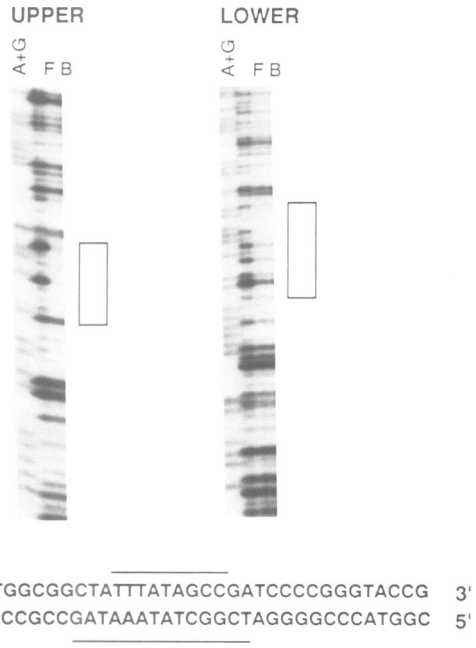


Fig. 8. DNase I footprinting by GTX. A 114 bp fragment containing the CKB MEF-2 motif was incubated with GST-GTX and was subjected to DNase I footprinting analysis. UPPER and LOWER indicate the upper and lower strands of the sequence shown on the bottom. (A+G) indicates Maxam–Gilbert sequence ladder. F and B indicate the absence and presence, respectively, of the fusion protein during DNase I digestion. Protected regions are indicated by boxes and their sequence is shown on the bottom with over- and underlines.

containing the CKB MEF-2 sequence as a model system to assess the transcriptional activity of GTX. We subcloned *Gtx* cDNA into an expression vector driven by constitutively active promoters, the cytomegalovirus (CMV) or the human elongation factor (EF) promoters (Mizushima and Nagata, 1990) (Figure 10A, right). To create a control plasmid, a *Gtx* cDNA was subcloned in the antisense orientation (anti-Gtx). As reporter plasmids, we used a luciferase gene driven by a 200 bp rat CKB promoter (pCKB) (Hobson *et al.*, 1990) and luciferase constructs driven by a thymidine kinase (TK) minimal promoter with or without one copy of the CKB MEF-2 sequence (Figure 10A, left).

It has been shown that promoters containing MEF-2

binding sites are inducible by serum, presumably due to the action of serum-inducible factors like RSRF (Pollock and Treisman, 1991). We therefore tested whether the CKB promoter, which contains a MEF-2-like motif, is serum-inducible. When pCKB was transfected into cells grown in 10% serum, luciferase activity was ~3 times higher than that of cells transfected under low serum conditions (0.5%) (compare lanes 1 and 3 of Figure 10B). Addition of serum 24 h after transfection in low serum conditions also caused an ~3-fold increase in expression of the CKB−luciferase fusion gene (lane 2). Similar activation by serum was observed with the TK-MEF reporter gene (lanes 4−6) but not with the TK minimal promoter construct (lanes 7 and 8). When reporter plasmids were transfected under 10% serum conditions, the TK-MEF showed ~10 times more luciferase activity than the TK minimal promoter (compare lane 8 with lane 6 of Figure 10B), suggesting that the serum-inducibility of pCKB might be mediated via the MEF-2 motif by serum-inducible factors which seem to exist in many cell types (Pollock and Treisman, 1991).

Interestingly, when the *Gtx* expression plasmid, EF-*Gtx*, was co-transfected with the pCKB reporter construct in high serum conditions, luciferase activity was significantly suppressed (Figure 10C, compare lane 1 with lanes 2 and 3). Similar results were obtained using TK-MEF (lanes 10−12), suggesting that the MEF-2 motif is sufficient to confer transcriptional repression by GTX. When 5 μg of the EF-*Gtx* or 0.5 μg of CMV-*Gtx* were co-transfected with TK-MEF reporter under high serum conditions, the luciferase activity was suppressed to a level similar to that seen when the reporter was transfected under low serum conditions (compare lane 4 with lanes 9, 12 and 14). Since expression vectors without the *Gtx* insert or with *Gtx* in an anti-sense orientation had no such effects on the TK-MEF or CKB promoters (Figure 10C, lanes 1, 10 and 16), the repressional activity is specific for *Gtx*. The repressional activity of *Gtx* was not due to the general transcriptional repression by, for example, interfering TATA box binding factors (Ohkuma *et al.*, 1990), because *Gtx* overexpression by itself had no effects on the TK-minimal promoter (Figure 10C, lanes 17−20) or on the CMV promoter (lanes 21 and 22), which lack the MEF-2 motif. Similar transcriptional repression was observed in primary glial cells and differentiated P19 cells

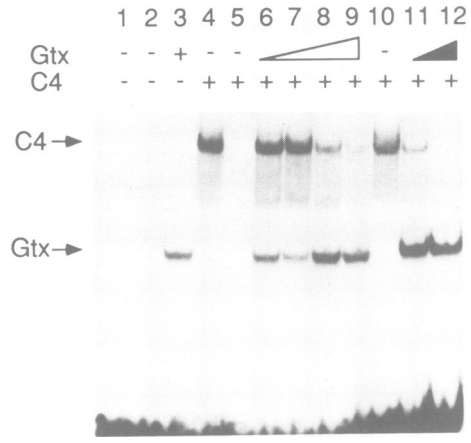


Fig. 9. GTX competes with RSRF for DNA binding *in vitro*. EMSA was performed using a ^{32}P-labeled CKB MEF-2 oligomer and *in vitro* translated RSRF-C4 and *Gtx*, as well as bacterially produced GTX (GST-GTX). Lanes: 1, ^{32}P-labeled CKB-MEF-2 oligomer probe alone; 2, 1 μl of nonprogrammed rabbit reticulocyte lysate (RRL) as a control; 3, 1 μl of *Gtx*-programmed RRL; 4, 1 μl of RSRF-C4-progammed RRL; 5, 1 μl of C4-programmed RRL plus a 100-fold excess of cold CKB MEF-2 oligomer; 6−9, 1 μl of C4-programmed RRL plus 1 μl (lane 6), 2 μl (lane 7), 5 μl (lane 8) or 10 μl (lane 9) of *Gtx*-programmed RRL; 10−12, 1 μl of C4-programmed RRL plus GST alone (100 nM) as a control (lane 10), or with 3 nM (lane 11) or 10 nM (lane 12) of GST-GTX fusion protein.

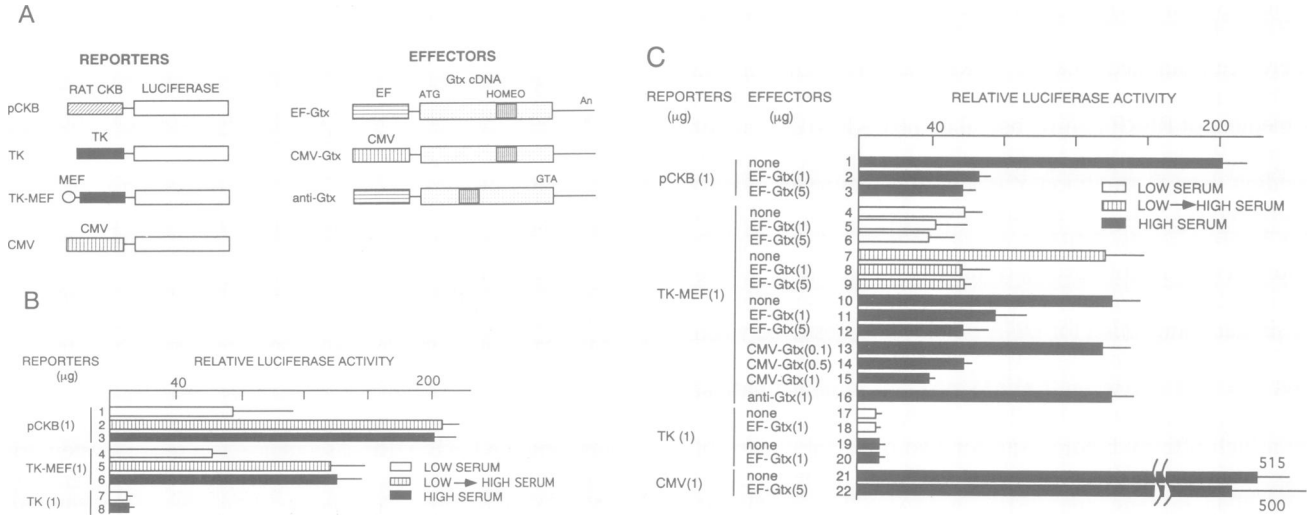


Fig. 10. Transfection analysis of *Gtx* and MEF-2-containing promoters. (**A**) Structure of the luciferase constructs (Reporters) and GTX expression plasmids (Effectors) used in this experiment. pCKB contains ~200 bp of the rat CKB promoter. The solid box in the reporter indicates the TK minimum promoter and the open circle indicates the CKB MEF-2 sequence (22 bp). CMV indicates the cytomegalovirus enhancer−promoter. (**B**) Serum inducibility of the reporter constructs containing the CKB MEF-2 sequence. One microgram of the indicated reporters were transfected into HeLa cells under low (0.5%) serum (open bars) or high (10%) serum (black bars) conditions. Twenty-four hours after transfection, some of the dishes in low serum medium were changed to high serum medium (hatched bars). Forty hours after transfection, luciferase activity was measured. (**C**) Repression by *Gtx* of the reporter genes containing the CKB MEF-2 sequence. One microgram of the indicated reporters and 0.5−5 μg of the indicated effectors were mixed and transfected into HeLa cells under low, high, and low to high serum conditions as in (B). pEF-BOS CAT plasmid was included in each transfection as an internal control for transfection efficiency and the results shown represent luciferase activity divided by CAT activity for each sample. Results are mean ± SD of triplicate transfection determinations. Similar results were obtained in experiments using either COS cells, P19 cells or primary cultured glial cells (data not shown).

treated with retinoic acid (data not shown). Thus, *Gtx* can transrepress promoters containing the serum-inducible, MEF-2-like element in many cell types *in vitro*.

Discussion

Gtx is a novel homeobox-containing gene

The murine homeobox gene, *Gtx*, reported here differs considerably from previously described vertebrate *Hox* genes in several aspects. First, the *Gtx* homeodomain is highly divergent from that of other invertebrate and vertebrate homeodomains. Second, it is unusual among mammalian homeobox genes in that it is expressed in the adult forebrain. Moreover, *Gtx* expression seems restricted to specific cell types such as glial cells and male germ cells. Third, *Gtx* maps to murine chromosome 7 and does not cosegregate with any known homeobox-containing genes. Fourth, GTX protein does not bind to the recognition sequences of *Antp*-, *Eve*-, *En*- or *Prd*-type homeodomains but binds to the MEF-2 binding sequence *in vitro*. Fifth, GTX competes with RSRF for the MEF-2 site *in vitro* and transrepresses serum-inducible promoters which contain the MEF-2 motif.

Gtx has a different DNA binding specificity from Antp-type homeoboxes

Although a very large number of homeobox genes have been identified, there are very few examples of vertebrate homeobox genes for which potential targets have been identified (Hayashi and Scott, 1990). At present, we do not know the physiological targets of *Gtx in vivo*. Although GTX protein avidly binds to the MEF-2 motif *in vitro*, it can also bind to other sequences *in vitro* as judged by SAAB analysis (I.Komuro and S.Izumo, unpublished results). It is possible that we might have missed other sites that may have higher affinity than the MEF-2 site *in vivo*. On the other hand, GTX is not a 'promiscuous DNA binder' because it did not bind to the mutated MCK MEF-2 (mt-6), $(TAA)_5$ motif or the BS-2 site, and it bound only very weakly to the NP site which is bound by many homeodomains including *Antp, En, Ftz, Eve, Zen, Ubx* and *Abd-B* (Hayashi and Scott, 1990; Treisman *et al.*, 1992). All of these homeodomains that bind to the NP sequence have Gln at residue 50 of the homeodomain (Table I). It has been shown that residue 50, which is located in the center of the third helix (Figure 1C), plays a key role in determining the target site selection, because the *Prd* homeodomain, which has only 40% overall similarity to the *Antp* homeodomain with Ser at position 50, does not bind to the NP site, while a single substitution of Ser50 to Gln50 changes the DNA binding specificity of *Prd* to that of the *Antp*-type (NP site) (Treisman *et al.*, 1989). Similarly, mutating Lys50 of the *Bcd* homeodomain to Gln50 switches DNA binding specificity to *Bcd* from the Bicoid site to the NP site (Hanes and Brent, 1989).

GTX has Gln at residue 50, which would predict that it would have *Antp*-type DNA binding specificity. Why does GTX then bind poorly to the NP site? A potential clue to this question comes from recent structural studies of the homeodomain–DNA complex of *En*, *Antp* and *MATα2* (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991). All three homeodomains contact the DNA sugar–phosphate backbone at eight residues: Phe8 (or Tyr8), Tyr25, Arg31, Gln44, Trp48, Arg54, Lys55 (or Arg55) and Lys57 (or Arg57). All eight sugar–phosphate contact residues are conserved in the *Gtx* homeodomain, suggesting that GTX 'docks' to DNA in the same manner as other homeodomains. The structural studies also indicated that the third helix of the homeodomain contacts the major groove of DNA at residues 47, 50, 51 and 54, and the amino-terminal arm of the homeodomain contacts the minor groove of DNA at residues 3 and 5. Intriguingly, these six residues that make specific side-chain contacts with DNA bases are highly conserved among all homeodomains that bind to the NP site (Table I). In contrast, GTX differs at three out of six of these residues: His3, Val47 and Thr54. While Val47 and Thr54 can be found in other homeodomains that bind to the NP site (Table I), His3 is unique. In fact, *Gtx* is the only gene among all known homeobox-containing genes in eukaryotes (of which there are more than 100) that has His at position 3. Thus, it is possible that the DNA binding characteristics of GTX may be due to the presence of unique amino acid residues at the positions predicted to contact DNA bases. It would be interesting to determine whether mutation of His3 to Arg would make an otherwise intact GTX change its DNA binding specificity.

Table I. Sequence comparison of predicted DNA base-contacting residues that bind to NP, MEF-2, Prd-3 and Bicoid sites

	DNA Base Contacting Residues					
Residue No.	3	5	47	50	51	54
NP site (ATTTAATTGA)						
Antp	R	R	I	Q	N	M
En	R	R	I	Q	N	A
Ubx	R	R	I	Q	N	M
Ftz	R	R	I	Q	N	M
Zen	R	R	I	Q	N	M
Caudal	K	R	I	Q	N	A
Dll	K	R	I	Q	N	S
S59	R	R	I	Q	N	T
Eve	R	R	V	Q	N	M
Mix-1	R	R	V	Q	N	A
NP consensus	R K	R	I V	Q	N	M A
MEF-2 site (CTAAAAATAAC)						
Gtx	H	R	V	Q	N	T
Prd-3 Site (ACGTCAAA)						
Prd	R	R	V	S	N	A
Gsb-p	R	R	V	S	N	A
Gsb-d	R	R	V	S	N	A
Bicoid site (TCTAATCCC)						
Bcd	R	R	V	K	N	A
Otd	R	P	I	K	N	R

Gtx is a tissue-specific homeobox gene

Most of the known homeobox genes are expressed more abundantly in the fetus than in the adult. *Gtx* expression was detected at early stages of development (8.5 day mouse embryos), although we have not been able to localize *Gtx* transcripts at this stage by *in situ* hybridization. In 18 day mouse embryos, *Gtx* is expressed in part of the forebrain (hypothalamus) as well as in mid- and hindbrain. *Gtx* was expressed at much higher levels in the adult brain than in the fetal or neonatal brain. In the adult CNS, expression of *Gtx* was most abundant in areas rich in oligodendrocyte cell bodies and subpopulations of astrocytes. Immunocytochemical studies using cultured primary glial cells confirmed the presence of GTX protein in these cell types.

Thus, *Gtx* may not only regulate early embryonic progenitor cell differentiation but also may play a role in postnatal differentiation and maintenance of glial cells, although it would be necessary to create null alleles of *Gtx* by homologous recombination in order to determine the developmental role(s) of *Gtx*.

Unlike the *Hox* class genes, *Gtx* was abundantly expressed in the forebrain, including corpus callosum, anterior commissure, internal capsule, thalamus and hypothalamus. The *Hox* class genes are also expressed in the developing CNS, but never anterior to the myelencephalon (reviewed in Kessel and Gruss, 1990). Very recently, several divergent classes of the homeobox-containing genes including *Dlx*, the mouse *Distalless* homolog (Price *et al.*, 1991), the *NKx*-2 family members (Price *et al.*, 1992), *Emx* and *Otx*, the mouse homolog of *Drosophila empty spiracles* and *orthodenticle*, respectively (Simeone *et al.*, 1992), are shown to be expressed in the forebrain. The transcriptional functions of these genes are not known. It is of interest that all of these murine homeobox genes have highly conserved homologs (amino acid identity of the homeodomains >90%) in *Drosophila*, while a *Drosophila* homolog of *Gtx* has yet to be identified.

Gtx is also abundantly expressed in adult testis. This expression of *Gtx* in the adult testis could result from expression either in germ cells or in somatic cells or both. To distinguish between these possibilities, we analyzed the expression of the *Gtx* gene in testis from W/W^v mutant mice. Strong expression of the *Gtx* gene was observed by *in situ* analysis in wild-type adult testis but was not detectable in mutant mice (Figure 6h), suggesting that the *Gtx* gene is expressed only in germ cells. We also analyzed expression of the *Gtx* gene during development of the juvenile testis. The appearance of the various spermatogenic cells in the prepubertal testis occurs in a defined temporal sequence (Bellve *et al.*, 1977). *Gtx* transcripts were not detected until postnatal day 17. Since Sertoli cells and spermatogonia in early meiotic prophase comprise a high percentage of the cells in the seminiferous epithelium at days 8–13, our observation suggests that *Gtx* is not expressed in these cells. Thus the *Gtx* gene seems to be activated at the pachytene stage of meiotic prophase in male germ cells. A similar expression pattern has been reported for the *Hox*1.4 gene (Rubin *et al.*, 1986). The physiological roles of two different classes of homeobox genes, *Antp*-type (*Hox*1.4) and divergent type (*Gtx*), which are co-expressed in the same cells at the same stage, is unknown.

Gtx maps closely to the imprinted region of mouse chromosome 7

Besides its unique expression pattern and functional properties, *Gtx* defines a new homeobox gene locus, chromosome 7, to which no known homeobox genes have been mapped previously. The majority of mammalian genes containing homeoboxes are most closely related to the *Antp* family. These genes are situated in four major clusters in the mouse genome, localized on chromosome 6 (*Hox-1*), 11 (*Hox-2*), 15 (*Hox-3*) and 2 (*Hox-4*) (reviewed in McGinnis and Krumlauf, 1992). *En1* and *En2* map to chromosome 1 and 5, respectively (Hill *et al.*, 1987; Joyner and Martin, 1987). *Hox7.1* maps to the proximal region of chromosome 5, 2–3 cM distant from the mouse *En2* gene (Hill *et al.*, 1989; Robert *et al.*, 1989). We have recently isolated two more novel murine homeobox genes that are expressed in tissue-specific manner and do not cosegregate with known *Hox* clusters (I.Komuro, H.Inagaki, N.G.Copeland and S.Izumo, unpublished observation).

We have compared our interspecific map of chromosomal 7 with a composite mouse linkage map that reports the map location of many uncloned mouse mutations (compiled by M.T.Davisson, T.H.Roderick, A.L.Hillyard, and D.P.Doolittle and provided from GBASE, a computerized database maintained at The Jackson Laboratory, Bar Harbor, ME). *Gtx* mapped to a region of the composite map that lacks known mutations with a phenotype that might be expected for an alteration in the *Gtx* locus (data not shown). The placement of *Gtx* 1.3 cM away from *Igf-2* and *Int-2*, which map to human chromosome 11p15.5 and 11q13 respectively, suggests that *Gtx* will also map to human chromosome 11 although its precise position on chromosome 11 is difficult to predict from these mapping studies.

Interestingly, *Igf-2* and a closely linked gene *H19* have both been shown to be imprinted genes in the mouse (Bartolomei *et al.*, 1991; DeChiara *et al.*, 1991). H19 may also be imprinted in humans (Zhang and Tycko, 1992). It remains to be determined if *Gtx* is an imprinted gene as well.

Gtx can act as a transcriptional repressor for MEF-2 site-containing promoters

Although homeobox genes are thought to act as sequence-specific transcription factors, there are relatively few examples in which transcriptional activities of vertebrate homeobox genes are examined. At present, we do not know the physiological targets of *Gtx in vivo*. Interestingly, the MEF-2 motif to which GTX binds with apparently high affinity *in vitro* is also a high affinity binding site for RSRF. The MEF-2 site is found in the promoters and enhancers of many serum-inducible and muscle-specific genes (Gossett *et al.*, 1989; Pollock and Treisman, 1991). The activity of reporter constructs containing the MEF-2 site is serum-inducible (this report and Pollock and Treisman, 1991). When a Gtx expression plasmid was co-transfected with the MEF-2 driven reporter genes, it specifically repressed the serum-induced activation of the reporter gene. In contrast, reporter genes lacking the MEF-2 element were not repressed by *Gtx*. These results suggest that GTX, at least in these experimental conditions, may compete with serum-inducible transactivators that bind to the MEF-2 motif. It is possible that RSRF may be one such protein, since GTX competed with RSRF for binding to the CKB/MEF-2 motif *in vitro*.

However, it is also possible that *Gtx* may inhibit the activity of serum-inducible transactivators not by a direct competition for the target sequence, but by other mechanisms. For example, Phox 1, a human *Prd*-type homeoprotein, enhances DNA-binding activity of the serum response factor (a MADS box protein) and potentiates transcription of the serum-response element-containing genes, without direct binding to the serum response element (Grueneberg *et al.*, 1992). It is conceivable that *Gtx* may also interact with RSRF and may facilitate dissociation of RSRF from the MEF-2 site. This might account for the marked decrease in C4/RSRF–DNA complex even in the presence of excess probe when increasing amounts of GTX were added to DNA binding reactions (Figure 9). However, the precise mechanisms of repression of the MEF-2 element by *Gtx* needs further investigation.

Several reports have demonstrated intimate relationships

between growth factors and homeobox-containing genes. Some peptide growth factors have been shown to induce expression of homeobox genes, which in turn transactivate other genes to establish the body axis during *Xenopus* development (Rosa, 1989; Ruiz i Altaba and Melton, 1989). Recently *Evx-1*, a mouse homeobox gene most closely related to the *Drosophila* pair-rule gene, *even-skipped*, was reported to activate the gene encoding cytoactin, a morphoregulatory molecule, via a serum-inducible TRE/AP-1 site (Jones *et al.*, 1992). Although we do not know the true biological role of *Gtx*, the result of our *in vitro* experiments raises a possibility that some homeobox-containing genes, such as *Gtx*, may function as inhibitors for growth factor-inducible genes in terminally differentiated adult tissues. Further experiments are necessary to test this hypothesis.

Materials and methods

cDNA library screening and DNA sequencing

The probes used for library screening were prepared by PCR from the *msh2/NK-4* homeodomain (Bodmer *et al.*, 1990). A λgt10 library prepared from 8.5 day mouse embryo mRNA (Fahrner *et al.*, 1987) was used for the first screening. Standard procedures (Sambrook *et al.*, 1989) were used to prepare nylon replicas (NEN) from each plate. Low stringency hybridizations were performed with 35% formamide, 1 M NaCl, 10% dextran sulfate and 1% SDS at 37°C. Screening of a mouse neonatal brain cDNA library (Stratagene) with the embryonic probe (pGtx-2) was under high stringency, in 50% formamide, 1 M NaCl and 1% SDS at 42°C for hybridization, and in 0.1 × SSC, 0.1% SDS at 65°C for the final washing. Inserts from the λgt10 clones of interest were subcloned into pBluescript. For sequencing, nested deletions were constructed using exonucleases III and VII, and double stranded DNAs were sequenced by the dideoxy chain-termination method (Sanger *et al.*, 1977) using Sequenase (USB) according to the manufacturer's directions.

Interspecific backcross mapping

Interspecific backcross progeny were generated by mating (C57BL/6J × *M.spretus*) F_1 females and C57BL/6J males as described by Copeland and Jenkins (1991). A total of 205 N_2 progency were obtained; a random subset of these N_2 mice were used to map *Gtx* (see text for details). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer and hybridization were performed essentially as described by Sambrook *et al.* (1989). All blots were prepared with Zetabind nylon membrane (AMF-Cuno). The pGtx-1 cDNA clone (Figure 1A) was labeled with [α-^{32}P]dCTP using a nick translation labeling kit (Boehringer Mannheim); washing was done to a final stringency of 0.5 × SSCP, 0.1% SDS, 65°C. A 12.0 kb fragment was detected in *Hin*cII-digested C57BL/6J DNA and an 8.2 kb fragment was detected in *Hin*cII-digested *M.spretus* DNA. The presence or absence of the 8.2 kb *M.spretus*-specific *Hin*cII fragment was followed in backcross mice.

A description of the probes and RFLPs for the loci linked to *Gtx*, including protein kinase C beta (*Pkcb*), interleukin-4 receptor (*Il-4r*), insulin-like growth factor-2 (*Igf-2*) and mammary tumor integration site-2 (*Int-2*), has been reported previously (Pritchard *et al.*, 1991; Foroni,L *et al.*, submitted). Recombination distances were calculated as described by Green (1981) using the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

DNA and RNA extractions and blotting analysis

Total cellular RNA was isolated by the LiCl/urea method (Auffray and Rougeon, 1980). RNA was size-separated on 1% agarose–formaldehyde gels, transferred to nitrocellulose membranes (Schleicher and Schuell) and hybridized with a [^{32}P]UTP-labeled cRNA probe at 65°C in buffer containing 50% formamide followed by serial washing with a final wash in 0.1 × SSC, 0.1% SDS at 65°C. Genomic DNA was extracted from Balb/c mouse liver. The extraction procedure and Southern blot analysis were as described by Sambrook *et al.* (1989).

In situ hybridization analysis

Four oligonucoleotides were synthesized (Milligen) and purified on a 6% polyacrylamide–8 M urea gel. The oligonucleotides correspond to both the sense and antisense strand of the *Gtx* cDNA nucleotides 346–391 (Figure 1B). The antisense strand DNAs of nucleotides 863–907 and 1154–1198 were also synthesized. The oligonucleotides were tailed at the 3′ end with [^{35}S]dATP (1000–1500 Ci/mmol, NEN) to a specific activity of 4 × 10^9 c.p.m./μg using terminal deoxynucleotidyltransferase (Bethesda Research Laboratories). Mouse tissues (brain, testis, heart, liver, spleen, kidney, adrenal gland, ovary and uterus) and fetuses were rapidly dissected out on frozen ice and cut at 14 μm thickness in a cryostat, thawed onto ProbeOn plus slides (Fisher Biochemicals) and processed as described by Schalling *et al.* (1991) for *in situ* hybridization. Briefly, sections were incubated at 42°C for 15–18 h with 10^6 c.p.m. of the labeled probe per 100 μl in a solution containing 50% formamide, 4 × SSC, 1 × Denhardt's solution, 1% sarcosyl, 0.02 M sodium acetate (pH 7.0), 10% dextran sulfate, 500 μg/ml salmon sperm DNA and 200 mM dithiothreitol. Sections were rinsed in 1 × SSC at 55 or 60°C for 1 h with five changes of buffer, dried, exposed to Hyperfilm β-Max X-ray film (Amersham) for 5–20 days and then dipped in NTB-2 nuclear track emulsion (Kodak). After exposure for 50–70 days, slides were developed in D19 (Kodak) for 3 min and fixed for 3 min. Sections were either mounted directly in glycerol–phosphate (3:1) or counterstained with hematoxylin–eosin and mounted in Permount (Fisher Biochemicals).

Production of bacterial GTX fusion protein

In order to produce the GTX homeodomain protein in bacteria, the cDNA sequence corresponding to amino acids 141–221 was amplified by PCR. The PCR conditions were as follows: denaturation at 94°C for 1 min, annealing at 46°C for 30 s and elongation at 72°C for 1 min in a commercial buffer (Perkin–Elmer Cetus). The PCR amplified product was gel-purified and ligated in-frame to pGEX-3X (Smith and Johnson, 1988). The GTX–glutathione transferase fusion protein (GST-GTX) was produced in *E.coli* and was purified by binding to a glutathione–Sepharose 4B column (Pharmacia). The fusion protein was eluted from the column by adding 10 mM glutathione. The eluted GST-GTX preparation was >95% pure as judged by Coomassie blue staining of the SDS–polyacrylamide gel.

Identification of potential GTX binding sites in vitro

The strategy used for isolation of potential binding sites of the GTX homeodomain is a modification of the SAAB method (Blackwell and Weintraub, 1990). Synthetic 51 nt oligonucleotides were designed with the sequence CAAGCCGGCGGGATCCTG-$(N)_{15}$-GGA**GAATTC**GAGAACCGG where the internal 15 nt segment was fully degenerate for each of the four bases and was flanked by an 18 nt 5′ primer containing a *Bam*HI site (underlined) and by an 18 nt 3′ end sequence containing an *Eco*RI site (bold type). ^{32}P-labeled double-stranded probes were generated using these degenerate oligonucleotides and an 18 nt oligo complementary to the 3′ end sequence of the 51mers as a primer for Klenow enzyme. The protein–DNA binding reaction was performed by incubating 10 nM GST-GTX and the labeled degenerate oligo probes in a buffer containing 10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 1 μg of poly(dI-dC) and 5% glycerol for 30 min on ice. The mixture was electrophoresed on native 10% polyacrylamide gels in 1 × TAE buffer at 4°C for 2–3 h at 300 V. The shifted band was excised and DNA was eluted from the gel slice. GST-GTX-bound DNA templates were amplified by 35 cycles of PCR using the 5′ and 3′ primers (18 nt long). PCR amplified products were purified and used for the second cycle of SAAB analysis. After three cycles of SAAB, the PCR products were digested with *Bam*HI and *Eco*RI, cloned into pBluescript and used to transform XL1-Blue (Stratagene). Resulting individual bacterial colonies were picked to amplify the inserts directly using the end-labeled PCR primers. The DNA binding activity of the individual PCR amplified inserts was examined by EMSA using different concentrations (0.3–10 nM) of the fusion protein. The sequences of the oligonucleotides used as probes or competiters in EMSA were: MCK, CTCTAAAAAT-AACTC/GAGTTATTTTTAGAG; CKB, GGCTATAAATAGCCGCCA/TGGCGGCTATTTATAGCC; MLC2, GGCTAAAAATAACCC/GGGTTATTTTTAGCC; MLC1/3, TTTTAAAAATAACTT/AAGTTATTTTTAAAA; $(TAA)_5$, TAATAATAATAATAA/TTATTATTATTATTA; NP, TCAATTAAATGA/TCATTTAATTGA; TTF-1, CACTGCCCAGTCAAGTGTTCTTGA/TCAAGAACACTTGACTGGGCAGTG; CArG, GCCAAAAGTGGTCAT/ATGACCACTTTTGGC; and E-box, GCTCCAACTGACCCT/AGGGTCAGTTGGAGC.

DNase I footprint analysis

To generate a probe for DNase I footprint analysis, the double-stranded CKB oligomer was subcloned into the *Bam*HI site of pUC19. A 114 bp probe labeled on only one strand was prepared by PCR using an end-labeled sequence primer. 100 nM of GST-GTX and 4 ng of the labeled probe were incubated in the same buffer as that used for EMSAs for 30 min on ice. 100 μg/ml DNase I was added and incubated for 1 min at room temperature, and the reaction was then stopped by adding 0.5 M EDTA. The digestion

products were extracted with phenol–chloroform and were analyzed by 6% DNA sequencing gels. Maxam–Gilbert sequencing reaction of the same fragment was done as described by Sambrook *et al.* (1989).

In vitro transcription and translation

In vitro transcription of the full coding sequences of RSRF-C4 and *Gtx* was carried out using T7 and T3 RNA polymerase, respectively. *In vitro* translation was performed in rabbit reticulocyte lysates (Promega). The amount of *in vitro* transcribed mRNA for each protein was titrated in order to obtain similar amounts of proteins in parallel translations, as determined by SDS–PAGE and fluorography of [^{35}S]methionine-labeled translation products (data not shown). For DNA-binding competition by GTX and RSRF, each mRNA was translated in the absence of labeled methionine.

Transfection analysis

For the TK-MEF–luciferase fusion gene, a double-stranded CKB AT-rich sequence (22mer) was subcloned into the *Bam*HI site of pUC19 and then isolated as an *Sst*I–*Hin*dIII fragment. This was subcloned upstream of the luciferase gene driven by a TK minimal promoter (−81 to +52 nt), pT81Luc, (Nordeen, 1988). For pCKB-Luc, the rat CKB promoter sequence from −195 to +5 nt was excised with *Hin*dIII and *Bgl*II from the p*wt* construct (Hobson *et al.*, 1990) and subcloned into the promoterless luciferase-containing gene, pXP2 (Nordeen, 1988). For pEF-BOS-Gtx, the cDNA insert was excised from pGtx-1, blunt-ended and subcloned into the blunt-ended *Xba*I site of pEF-BOS (Mizushima and Nagata, 1990). For pCMV-Gtx, the *Gtx* cDNA was excised with *Eco*RI, and subcloned into the *Eco*RI site of pcDNAI (Invitrogen). All constructs were verified by DNA sequencing. Transfections were performed by the protocol of Chen and Okayama (1987) using 1 μg reporter plasmid and 0.1–5 μg of effector plasmids. Luciferase activity was measured 48 h after transfection as previously described (Nordeen, 1988). Amounts of plasmid DNA were kept constant in all transfection experiments by replacing any effector plasmids not added by pEF-BOS plasmid. One microgram of pEF-BOS-CAT was also included in each transfection as an internal control for transfection efficiency, and all results were normalized by the chloramphenicol acetyltransferase (CAT) activity in each sample.

Immunofluorescent studies

Primary cultures of astroglia and oligondendroglia were obtained using neonatal rat cerebral tissues according to a standard protocol (McCarthy and de Vellis, 1980). A polyclonal antibody was made by injecting GST-GTX fusion protein into chickens. Chicken egg IgG was purified as described previously (Polson *et al.*, 1985). Indirect immunofluorescent studies were carried out using polyclonal anti-GTX antibody and anti-GFAP, anti-neurofilament (68 kDa) and anti-galactocerebroside monoclonal antibodies (Boehringer) as described by Harlow and Lane (1988). Anti-chicken IgG antibody conjugated with Texas Red and anti-mouse IgG antibody conjugated with fluorescein isothiocyanate were used, as secondary antibodies.

Acknowledgements

We thank B.L.M. Hogan for the cDNA library, P.A.Benfield for the CKB plasmid, R.Treisman for the RSRF plasmid, Y.-N.Jan, R.D.Rosenberg, W.Grossman, D.Housman and A.Graybiel for encouragement, R.Krumlauf, B.Neel, A.Greene and S.L.Alper for critical reading of the manuscript, D.J.Gilbert for excellent technical assistance, and S.Fukuoka for database analysis. This research was supported in part by Jacob D.Indursky Memorial Fund, the National Cancer Institute under contract N01-CO-74101 with ABL and a grant from the Muscular Dystrophy Association. I.K. is supported by the Bayer/Japan Heart Foundation Scholarship and the Medical Foundation Fellowship. M.S. is supported by an EMBO Longterm Fellowship and grants from the Wenner Glen Foundation and the Swedish Society of Medicine. S.I. is an Established Investigator of the American Heart Association.

References

Affolter,M., Schier,A. and Gehring,W.J. (1990) *Curr. Opin. Cell Biol.*, **2**, 485–495.
Akam,M. (1987) *Development*, **101**, 1–22.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.
Bartolomei,M.S., Zemel,S. and Tilghman,S.M. (1991) *Nature*, **351**, 153–155.
Beachy,P.A., Krasnow,M.A., Gavis,E.R. and Hogness,D.S. (1988) *Cell*, **55**, 1069–1081.
Bellve,A.R., Cavicchia,J.C., Millette,C.F., O'Brien,D.A., Bhatnagar,Y.M. and Dym,M. (1977) *J. Cell Biol.*, **74**, 68–85.
Blackwell,T.K. and Weintraub,H. (1990) *Science*, **250**, 1104–1151.
Bodmer,R., Jan,L.Y. and Jan,Y.N. (1990) *Development*, **110**, 661–669.
Braun,T., Tannich,E., Buschhausen-Denker,G. and Arnold,H.H. (1989) *Mol. Cell. Biol.*, **9**, 2513–2525.
Castrillo,J.L., Theill,L.E. and Karin,M. (1991) *Science*, **253**, 197–199.
Chen,C. and Okayama,H. (1987) *Mol. Cell. Biol.*, **7**, 2745–2752.
Chisaka,O. and Capecchi,M. (1991) *Nature*, **350**, 473–479.
Copeland,N.G. and Jenkins,N.A. (1991) *Trends Genet.*, **7**, 113–118.
Cserjesi,P. and Olson,E.N. (1991) *Mol. Cell. Biol.*, **11**, 4854–4862.
DeChiara,T.M., Robertson,E.J. and Efstratiadis,A. (1991) *Cell*, **64**, 849–859.
Dohrmann,C., Azpiazu,N. and Frasch,M. (1990) *Genes Dev.*, **4**, 2098–2111.
Donoghue,M., Ernest,H., Wentworth,B., Nadal-Ginard,B. and Rosenthal,N. (1988) *Genes Dev.*, **2**, 1779–1790.
Duboule,D. and Dolle,P. (1989) *EMBO J.*, **8**, 1497–1505.
Fahrner,K., Hogan,B.L.M. and Flavell,R.A. (1987) *EMBO J.*, **6**, 1265–1271.
Gossett,L.A., Kelvin,D.J., Sternberg,E.A. and Olson,E.N. (1989) *Mol. Cell. Biol.*, **9**, 5022–5033.
Graham,A., Papalopulu,N. and Krumlauf,R. (1989) *Cell*, **57**, 367–378.
Green,E.L. (1981) In Green,E.L. (ed.), *Genetics and Probability in Animal Breeding Experiments*. Macmillan, New York, pp. 77–113.
Grueneberg,D.A., Natesan,S., Alexandre,C. and Gilman,M.Z. (1992) *Science*, **257**, 1089–1095.
Guazzi,S., Price,M., De Felice,M., Damante,G., Mattei,M.G. and Di Lauro,R. (1990) *EMBO J.*, **9**, 3631–3639.
Hanes,S.D. and Brent,R. (1989) *Cell*, **57**, 1275–1283.
Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Harrison,S.C. and Aggarwal,A.K. (1990) *Annu. Rev. Biochem.*, **59**, 933–969.
Hayashi,S. and Scott,M. (1990) *Cell*, **63**, 883–894.
He,X., Treacy,M.N., Simmons,D.M., Ingraham,H.A., Swanson,L.W. and Rosenfeld,M. (1989) *Nature*, **340**, 35–42.
Hill,R.E., Hall,A.E., Sime,C.M. and Hastie,N.D. (1987) *Cytogenet. Cell. Genet.*, **44**, 171–174.
Hill,R.E., Jones,P.F., Rees,A.R., Sime,C.M., Justice,M.J., Copeland,N.G., Jenkins,N.A., Graham,E. and Davidson,D.R. (1989) *Genes Dev.*, **3**, 26–37.
Hobson,G.M., Mitchell,M.T., Molloy,G.R., Pearson,M.L. and Benfield,P.A. (1988) *Nucleic Acids Res.*, **16**, 8925–8945.
Horlick,R.A., Hobson,G.W., Patterson,J.H., Mitchell,M.T. and Benfield,P.A. (1990) *Mol. Cell. Biol.*, **10**, 4826–4836.
Ingraham,H.A., Chen,R., 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.
Jones,F.S., Chalepakis,G., Gruss,P. and Edelman,G.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2091–2095.
Joyner,A.L. and Martin,G.R. (1987) *Genes Dev.*, **1**, 29–38.
Kessel,M. and Gruss,P. (1990) *Science*, **249**, 374–379.
Kessel,M., Balling,R. and Gruss,P. (1990) *Cell*, **61**, 301–308.
Kim,Y. and Nirenberg,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7716–7720.
Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
Kozak,M. (1987) *Nucleci Acids Res.*, **15**, 8125–8148.
Lazzaro,D., Price,M., De Felice,M. and Di Lauro,R. (1991) *Development*, **113**, 1093–1104.
Lufkin,T., Dierich,A., LeMeur,M., Mark,M. and Chambon,P. (1991) *Cell*, **66**, 1105–1119.
McCarthy,K.D. and de Vellis,J. (1980) *J. Cell Biol.*, **85**, 890–902.
Malicki,J., Schughart,K. and McGinnis,W. (1990) *Cell*, **63**, 961–967.
McGinnis,N., Kuziora,M.A. and McGinnis,W. (1990) *Cell*, **63**, 969–976.
McGinnis,W. and Krumlauf,R. (1992) *Cell*, **68**, 283–302.
Mermod,N., O'Neill,E.A., Kelley,T.J. and Tjian,R. (1989) *Cell*, **58**, 741–753.
Mintz,B. and Russel,E.S. (1957) *J. Exp. Zool.*, **134**, 207–237.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Mizushima,S. and Nagata,S. (1990) *Nucleic Acids Res.*, **18**, 5322.
Muller,M., Affolter,M., Leupin,W., Otting,G., Wuthrich,K. and Gehring,W. (1988) *EMBO J.*, **7**, 4299–4304.
Nordeen,S.K. (1988) *BioTechniques*, **6**, 454–458.
Ohkuma,Y., Horikoshi,M., Roeder,R.G. and Desplan,C. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2289–2293.

Otting,G., Qian,Y.Q., Billeter,M., Muller,M., Affolter,M., Gehring,W.J. and Wuthrick,K. (1990) *EMBO J.*, **9**, 3085–3092.
Pelletier,J., Schalling,M., Buckler,J.A., Rogers,A., Haber,D.A. and Housman,D. (1991) *Genes Dev.*, **5**, 1345–1356.
Pollock,R. and Treisman,R. (1991) *Genes Dev.*, **5**, 2327–2341.
Polson,A., Coetzer,T., Kruger,J., von Maltzahn,E. and van der Merwe,K.J. (1985) *Immunol. Invest.*, **14**, 323–327.
Price,M., Lemaistre,M., Pischetola,M., Di Lauro,R. and Duboule,D. (1991) *Nature*, **351**, 748–751.
Price,M., Lazzaro,D., Pohl,T., Mattei,M.G., Ruther,U., Olivo,J.C., Duboule,D. and Di Lauro,R. (1992) *Neuron*, **8**, 241–255.
Pritchard,M.A. *et al.* (1991) *Genomics*, **10**, 801–806.
Robert,B., Sassoon,D., Jacq,B., Gehring,W. and Buckingham,M. (1989) *EMBO J.*, **8**, 91–100.
Rosa,F.M., (1989) *Cell*, **57**, 965–974.
Rubin,M.R., Toth,L.E., Patel,M.D., D'Eustachio,P. and Chi Nguyen-Huu,M. (1986) *Science*, **233**, 663–667.
Ruiz i Altaba,A. and Melton,D.A. (1989) *Nature*, **341**, 33–38.
St Johnson,O. and Nüsslein-Volhard,C. (1992) *Cell*, **68**, 201–219.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schalling,M., Franco-Cereceda,A., Hemsen,A., Dagerlind,A., Seroogy,K., Persson,H., Hokfelt,T. and Lundberg,J.M. (1991) *Neuroscience*, **41**, 753–766.
Schwarz-Sommer,Z., Huijser,P., Nacken,W., Saedler,H. and Sommer,H. (1990) *Science*, **250**, 931–936.
Scott,M.P. and Carroll,S.B. (1987) *Cell*, **51**, 689–698.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
Simeone,A., Acampora,D., Gulisano,M., Stornaiuolo,A. and Boncinelli,E. (1992) *Nature*, **358**, 687–690.
Smith,D.B. and Johnson,K.B. (1988) *Gene*, **67**, 31–40.
Treisman,J., Gonczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) *Cell*, **59**, 553–562.
Treisman,J., Harris,E., Wilson,D. and Desplan,C. (1992) *BioEssays*, **14**, 145–150.
Treisman,R. (1990) *Semin. Cancer Biol.*, **1**, 47–58.
Treisman,R. and Ammerer,G. (1992) *Curr. Opin. Genet. Dev.*, **2**, 221–226.
Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) *Cell*, **67**, 517–528.
Yu,Y.T., Breitbart,R.E., Smoot,L.B., Lee,Y., Mahdavi,V. and Nadal-Ginard,B. (1992) *Genes Dev.*, **6**, 1783–1798.
Zhang,Y. and Tycko,B. (1992) *Nature Genetics*, **1**, 40–44.

Received on November 6, 1992; revised on January 5, 1993