Embryonic expression and cloning of the murine GATA-3 gene

Kathleen M. George[1], Mark W. Leonard[1], Matthew E. Roth[1], Ken H. Lieuw[1], Dimitris Kioussis[2], Frank Grosveld[2,3] and James Douglas Engel[1,*]

[1]Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Il 60208-3500, USA
[2]Laboratory of Molecular Immunology and Laboratory of Gene Structure and Expression, National Institute for Medical Research, Mill Hill, UK
[3]Department of Cell Biology and Genetics, Erasmus University, Rotterdam, Holland

*Author for correspondence

SUMMARY

We describe the embryonic expression pattern as well as the cloning and initial transcriptional regulatory analysis of the murine (m) *GATA-3* gene. In situ hybridization shows that mGATA-3 mRNA accumulation is temporally and spatially regulated during early development: although found most abundantly in the placenta prior to 10 days of embryogenesis, mGATA-3 expression becomes restricted to specific cells within the embryonic central nervous system (in the mesencephalon, diencephalon, pons and inner ear) later in gestation. GATA-3 also shows a restricted expression pattern in the peripheral nervous system, including terminally differentiating cells in the cranial and sympathetic ganglia. In addition to this distinct pattern in the nervous system, mGATA-3 is also expressed in the embryonic kidney and the thymic rudiment, and further analysis showed that it is expressed throughout T lymphocyte differentiation.

To begin to investigate how this complex gene expression pattern is elicited, cloning and transcriptional regulatory analyses of the *mGATA-3* gene were initiated. At least two regulatory elements (one positive and one negative) appear to be required for appropriate tissue-restricted regulation after transfection of *mGATA-3*-directed reporter genes into cells that naturally express GATA-3 (T lymphocytes and neuroblastoma cells). Furthermore, this same region of the locus confers developmentally appropriate expression in transgenic mice, but only in a subset of the tissues that naturally express the gene.

Key words: PNS, CNS, T cells, kidney, mGATA-3 cloning, expression, mouse

INTRODUCTION

The cascade of events governing tissue specification during vertebrate development is generally believed to be initiated by the response of regionally distinct germ layer cells to extracellular signaling cues. These signals, in turn, activate transcription factor proteins which act on genes required for the elaboration of specific cell fates. However, the molecules and mechanisms involved in vertebrate tissue specification are still being defined. To understand how cells achieve their ultimate developmental fate, it is necessary to determine how signals transduced to the nucleus mediate specific transcription factor regulation of downstream target genes.

Transcription factor GATA-1 is known to be an essential regulatory protein for murine erythropoiesis, while the physiological function of other members of this growing multifactor family is far less well understood. Six distinct members of the GATA family have been identified in vertebrate organisms. Individual family members share greater identity between species than do all of the family members expressed within one species (Yamamoto et al., 1990; Zon et al., 1991). Furthermore, the tissue distribution of each GATA family member appears to be highly conserved among different vertebrates.

These observations suggest that GATA factors share evolutionarily conserved roles in vertebrate gene regulation (Arceci et al., 1993; Dorfman et al., 1992; Kelley et al., 1993; Ko et al., 1991; Trainor et al., 1990; Wilson et al., 1990; Yamamoto et al., 1990; Yang et al., 1994; Zon et al., 1990, 1991).

Within a single species, the amino acid sequence identity of the GATA factors is highest within the DNA-binding domain (Yamamoto et al., 1990) and likely, as a consequence, family members display overlapping binding specificity in vitro (Ko and Engel, 1993; Merika and Orkin, 1993). The GATA factors share a common DNA sequence recognition motif composed of two (C_4) zinc fingers: only the carboxyl finger is required for the GATA nucleotide core site-specific DNA binding (Martin and Orkin, 1990; Omichinski et al., 1993; Yang and Evans, 1992; Yang et al., 1994); the amino finger may specify nuclear localization of the factor (Yang et al., 1994), and may additionally impart alternative DNA-binding site specificity (Whyatt et al., 1993). Outside of the DNA-binding domain, the sequence of the GATA factors within a single species diverges significantly.

Confirmation that members of the GATA family play key roles in gene regulation has been rigorously documented only for GATA-1 (Evans and Felsenfeld, 1989; Tsai et al., 1989).

Analysis of mice reconstituted from GATA-1-ablated ES cells showed that these defective cells fail to contribute to erythropoiesis in chimeric animals (Pevny et al., 1991); differentiation of the same GATA-1-deficient ES cells also fail to give rise to mature erythrocytes in vitro (Simon et al., 1992). GATA-1 is expressed at all developmental stages of erythropoiesis, as well as in mast cells, megakaryocytes and Sertoli cells of the testis (Ito et al., 1993; Martin et al., 1990; Romeo et al., 1990; Whitelaw et al., 1990; Yamamoto et al., 1990; Yomogida et al., 1994). GATA-2 is more broadly expressed than GATA-1 (Dorfman et al., 1992; Yamamoto et al., 1990), while GATA-4, GATA-5 and GATA-6, the most recently identified family members, appear to constitute a distinct subfamily that is expressed in the developing heart and gut (Arceci et al., 1993; Kelley et al., 1993; Lavernere et al., 1994; Tamura et al., 1993). Thus with the exception of GATA-2, the vertebrate GATA factor family members described to date appear to be distinct in their individual expression patterns.

GATA-3 was first identified as an abundantly expressed mRNA in chicken, mouse and human T lymphocytes and in the embryonic chicken and murine brain (Ko et al., 1991; Kornhauser et al., 1994; Leonard et al., 1993; Yamamoto et al., 1990). Although no target genes have yet been identified in the nervous system, GATA-3 activity has been shown to be required for stimulation of the T cell receptor (TCR) genes (Joulin et al., 1991; Ko et al., 1991; Redondo et al., 1990, 1991), the *CD8α* gene (Hambor et al., 1993; Landry et al., 1993) and the HIV-1 LTR (Yang and Engel, 1993) in T cells.

The initial step in defining the role(s) that a regulatory protein may play in development is to document where and when it is both first and then persistently expressed during normal embryogenesis; once the normal expression pattern is established, one may then focus subsequent studies on specific candidate tissues or target genes. Using in situ hybridization, we show that mGATA-3 is expressed at previously identified as well as new anatomical sites during murine embryonic development: in addition to expression in the thymic rudiment and at the earliest stages of T lymphocyte differentiation, mGATA-3 was found to be expressed in highly restricted groups of cells within the placenta, in the peripheral and central nervous systems (PNS and CNS) and in the embryonic liver, kidney and adrenal gland.

To initiate transcriptional regulatory analyses, we isolated and structurally characterized the *mGATA-3* gene. We then employed both transient transfection and transgenic mice to determine the position of *mGATA-3* regulatory sequences. These data suggested that *cis*-regulatory elements required for tissue-restricted expression of this factor lie within 3 kbp surrounding the transcriptional initiation site. The transgenic embryos displayed an appropriately regulated mGATA-3 expression pattern in most tissues. However, this presumed *mGATA-3* transcriptional regulatory domain appeared to be missing element(s) required in vivo for specification of CNS transcriptional control.

MATERIALS AND METHODS

In situ hybridization

In situ hybridization was performed as described previously (Wilkinson, 1987) using equal cts/minute of T7 RNA polymerase-derived [^{35}S]UTP-labeled, single-stranded riboprobes. The mGATA-3 antisense probe was a linearized exoIII-deletion of the full-length cDNA, mc5b1 (Ko et al., 1991). For autoradiography, slides were coated with Kodak NTB2 film emulsion and kept under desiccant at 4°C 1-2 weeks. After developing, the sections were counterstained with cresyl violet.

Oligonucleotides: RT/PCR and primer extension

Quantitative RT-PCR analysis was performed with mGATA-3 primers and mS16 primers, as previously described (Leonard et al., 1993). Oligonucleotides used for primer extension were mG3PE1 (5′...GAGTAGCAAGGAGCGTAGAGGAGGA...3′, corresponding to nt +187 to +163) and mG3PE2 (5′...CTTTGCGGGATAGTT-TAGCAA...3′, corresponding to nt +814 to +794). The sense strand corresponding to both oligonucleotide sequences is shaded in Fig. 5C.

Cloning

Approximately 3×10^6 pfu of a λFIXII BALB/c genomic library (a gift from D. I. H. Linzer) was screened at high stringency with a random-primed full-length mGATA-3 cDNA. Purified phage were subdivided based on whether they hybridized to a probe containing either 5′ or 3′ cDNA sequences. The longest clones, which corresponded to the 5′ or 3′ end, were subcloned into pGEM7Zf(+) or pBluescript vectors (Promega, Inc. and Stratagene, Inc., respectively). The clones were characterized by Southern blotting and exon-intron boundaries were identified. The exon-intron boundaries, as well as the region upstream of the cap site, were sequenced by dideoxy-chain termination of the double-stranded plasmids (Choi and Engel, 1986; Sanger et al., 1977). The EMBL gene bank accession number for the sequence shown in Fig. 5C is Z33620.

Flow cytometry

CD4- and CD8-positive thymocyte subpopulations were isolated using two-color analysis as previously described (Robey et al., 1991). Cells were then lysed for RNA RT/PCR analysis (above) as described (Leonard et al., 1993).

Plasmid constructions

pR1.3HNC

Plasmid subclone R1.3 of genomic recombinant λ7e (Fig. 7A) was used as the base plasmid to substitute the neomycin resistance (*neo^R*) open reading frame for mGATA-3 translation initiation coding sequences. The 5′ primer (mG3USPC1Hind; 5′...GCAGG-AAGCTTGCGAAGACCT...3′) corresponds to nt +518 to +538 (spanning the *Hin*dIII site, underlined) in intron 1 (Fig. 7C), while the 3′ primer (mG3DSPC1Nco; 5′...CACCTCCATGGCCTCG-GCTGT...3′) matched the flanking sequences of the translation initiation codon (underlined; Ko et al., 1991) but converted sequences surrounding the mGATA-3 translation initiation codon to a *Nco*I site. The *neo^R* gene was amplified from pMC1NeoA^+ (Stratagene) using primers NeoA^+5′PC1 (5′...AGCCACCATGGGGATCGGCCATT-GA...3′; which converts the *neo* gene translation initiation codon to a *Nco*I site; underlined) and NeoA^+3′PC1 (5′...GGCTGCAGATC-GATGGATCCGAAC...3′) which introduces a *Cla*I site (underlined) at the 3′ end of the neo^R coding sequences. The PCR amplified and restriction enzyme digested mGATA-3 *Hin*dIII-*Nco*I fragment was then ligated to the amplified and restriction enzyme digested neo^R *Nco*I (partial)-ClaI fragment, and the resulting *Hin*dIII-neo^R-*Cla*I fragment was used to replace the internal *Hin*dIII-*Cla*I fragment of R1.3 (Fig. 7C; the *Cla*I site is in intron 2, not shown).

308pCAT

pR1.3HNC (above) was digested to completion with *Nco*I and partially digested with *Nsi*I (nt +47 to +52; Fig. 5C). After repairing the ragged ends, the *CAT* structural gene isolated from pJFCAT1 (Fridovich-Keil et al., 1991) was then ligated to this vector.

308pI1CAT

pRI.3HNC was digested with *Nco*I (nt +1002 to +1004; Fig. 5C) and partially filled (using Klenow polymerase with dCTP and dATP), and then ligated to the *CAT* structural gene from pJFCAT1 to generate 308pI1CATI2. The plasmid was then digested with *Sac*I, and the largest fragment was isolated and religated to generate 308pI1CAT. 5′ deletions of the proximal promoter (from −308 in a 5′ to 3′ direction, referred to in the text) were generated by *Exo*III-S1 nuclease digestion of this plasmid.

ΔSac2

308pI1CAT was digested with *Sac*II (sites underlined in Fig. 5C), and the large fragment (containing the internally deleted mGATA-3 sequences plus vector) was isolated and ligated.

308pI1mCAT

The inverted GATA site in the first intron of 308pI1CAT (nt +233 to +240; boxed in Fig. 5C) was mutated by insertion of a *Sal*I linker into the natural *Eco*RV site (5′-TTGAT<u>GGTCGACC</u>ATC-3′; the added linker sequences are underlined).

(I1+) and (I1−)308pCAT

I1 was amplified by PCR, using primers incorporating *Sph*I sites on the ends to facilitate subcloning into pGEM7. Oligonucleotide primers used were: mG3INSDS (5′...GGATGCATGCCTG-CAAGGGAGAGAA...3′) and mG3INSUS (5′...GGATGCATGCG-GTTGGTATTGTGAC...3′). The PCR products were digested with *Sph*I, and ligated to 308pCAT which had been digested with *Sph*I (a unique *Sph*I site is in the residual polylinker at the 5′ end of 308pCAT). The subclones were sequenced to confirm the two orientations of I1.

1754pI1CAT

308pI1CAT (above) was digested with *Xho*I (in the residual multiple cloning site of pGEM7), *Cla*I linkers were added, and the resulting plasmid was digested with *Cla*I and *Hin*dIII (located in I1, above). The largest fragment was isolated and then ligated to a 2.3 kb genomic *Cla*I (−1754)-*Hin*dIII (+524) fragment isolated from genomic clone λ7e to extend the contiguous mGATA-3 genomic sequence of 308pI1CAT to nt −1754 to generate 1754pI1CAT (Fig. 5C). 5′ deletions of this subclone were generated by *Exo*III-S1 nuclease digestion of 1754pI1CAT (Henikoff, 1984), and sequenced to determine the extent of digestion. The clones 1203pI1CAT, 1475pI1CAT and 1754pI1CAT used in the transient transfection analysis (endpoints shown as rightward arrows in the 5′ flanking sequences in Fig. 5C; see also Fig. 7) were generated from this set of 5′ deletions.

2052pI1LacZ

The *lacZ* gene was isolated by digestion of pNASSβ (Clontech Inc.) with *Not*I and then filled using Klenow polymerase plus dNTPs. The *lacZ* insert was then ligated to pR1.3HNC which had been digested with *Nco*I and then treated with S1 nuclease to create intermediate 308pI1lacZ. This plasmid was then digested with *Xho*I (in the residual pGEM4 polylinker) plus *Hin*dIII and then ligated to the 2.6 kbp *Xho*I-*Hin*dIII fragment of genomic clone λ7e (from nt −2052 to +523; Fig. 5C) to generate 2052pI1lacZ.

Cell culture and transient transfection assays

Cell lines were purchased from American Type Culture Collection (BW5147.3, CH27), or were gifts from D. I. H. Linzer (BALB/c 3T3), R.I. Morimoto (NIH 3T3), V. Patel (MEL), or B. Mirkin (C1300). 3T3 cells were grown in DME containing 10% heat-inactivated fetal bovine serum plus penicillin/streptomycin. C1300 cells were grown in DME containing 10% heat-inactivated fetal calf serum, 1 mM sodium pyruvate (Sigma), and penicillin/streptomycin. BW5147.3

cells were grown in DME containing 10% heat-inactivated horse serum and penicillin/streptomycin.

3T3 cells were transfected in triplicate using the calcium phosphate procedure (Ausubel et al., 1989; Graham and van der Eb, 1973). Approximately 1.7 pmole of each test plasmid and 1 μg pRSV.LUC were transfected. pBluescript was added to bring total amount of DNA to 10 μg. Briefly, cells were passaged the day before transfection at a density of 5×10^5 cells/100 mm plate. DNA was suspended 250 mM CaCl$_2$. 2× Hepes-buffered saline (50 mM Hepes acid pH 7, 280 mM NaCl, and 1.5 mM Na$_2$HPO$_4$) was added and vortexed. After 30 minutes, the precipitate was added to plates of cells. Cells were washed in PBS and fresh media was added approximately 16-20 hours after transfection.

BW5147.3 and C1300 were transfected in duplicate using DEAE-dextran (Choi and Engel, 1986). Briefly, 10^7 cells were incubated at room temperature for 30 minutes with the DNA (the same concentrations used for 3T3 transfections) and either 500 μg/ml (BW5147.3) or 100 μg/ml (C1300) DEAE-Dextran (M_r 5×10^5; Pharmacia). Cells were then washed and plated in 20 ml of complete media.

Cell lysates were prepared 48 hours after transfection by the freeze-thaw procedure (Rosenthal, 1987). Luciferase assays were used to determine transfection efficiency and CAT assays were performed using equal amounts of extract (Ausubel et al., 1989). The percentage of acetylation for each extract was determined by quantification of the ^{14}C-acetylated chloramphenicol on thin-layer chromatography plates using a Molecular Dynamics PhosphorImager. The per cent acetylation was then normalized to luciferase activity to correct for transfection efficiency.

Transgenic mice: transient embryo analysis and embryo staining

A DNA fragment (2052LacZ) containing the *mGATA-3* gene directing β-galactosidase synthesis (above) was removed from vector sequences by restriction enzyme digestion followed by preparative agarose gel electrophoresis. The DNA was finally purified using Gene Clean (Bio 101) followed by an Elutip column (Schleicher and Schuell). The DNA was resuspended into microinjection buffer consisting of TE (10 mM Tris-HCl, 0.2 mM EDTA, final pH = 7.5) and microinjected into fertilized oocytes from (C57Bl10 × CBA) F$_1$ mice. Eggs surviving injection were then transferred into oviducts of recipient pseudopregnant females (Dillon and Grosveld, 1993; Hogan et al., 1986).

Embryos were then isolated at days 10, 11 and 12 of gestation and fixed in 1% formaldehyde, 0.2% gluteraldehyde, 2 mM MgCl$_2$, 5 mM EGTA and 0.02% NP-40 in PBS. They were then washed with PBS and stained in the dark, overnight at ambient temperature in 1 mg/ml X-gal (Whiting et al., 1991). Transgenesis was determined by Southern blotting of placentae recovered from individual embryos.

RESULTS

mGATA-3 expression during embryonic development

RNA in situ hybridization was performed to define the sites and times of mGATA-3 expression during embryonic development. Embryos from 8.0 days p.c. (e8.0) through 14.5 days (e14.5) of gestation were examined, and specific organs and tissues were identified by comparison to depicted sections (Kaufman, 1992). Tissue sections were hybridized to sense or antisense transcripts prepared from mGATA-3 cDNA clone mc5b1 (Ko et al., 1991). The probes did not contain sequences corresponding to the zinc fingers of mGATA-3 to preclude the possibility of cross hybridization to other members of the GATA factor family (Yamamoto et al., 1990).
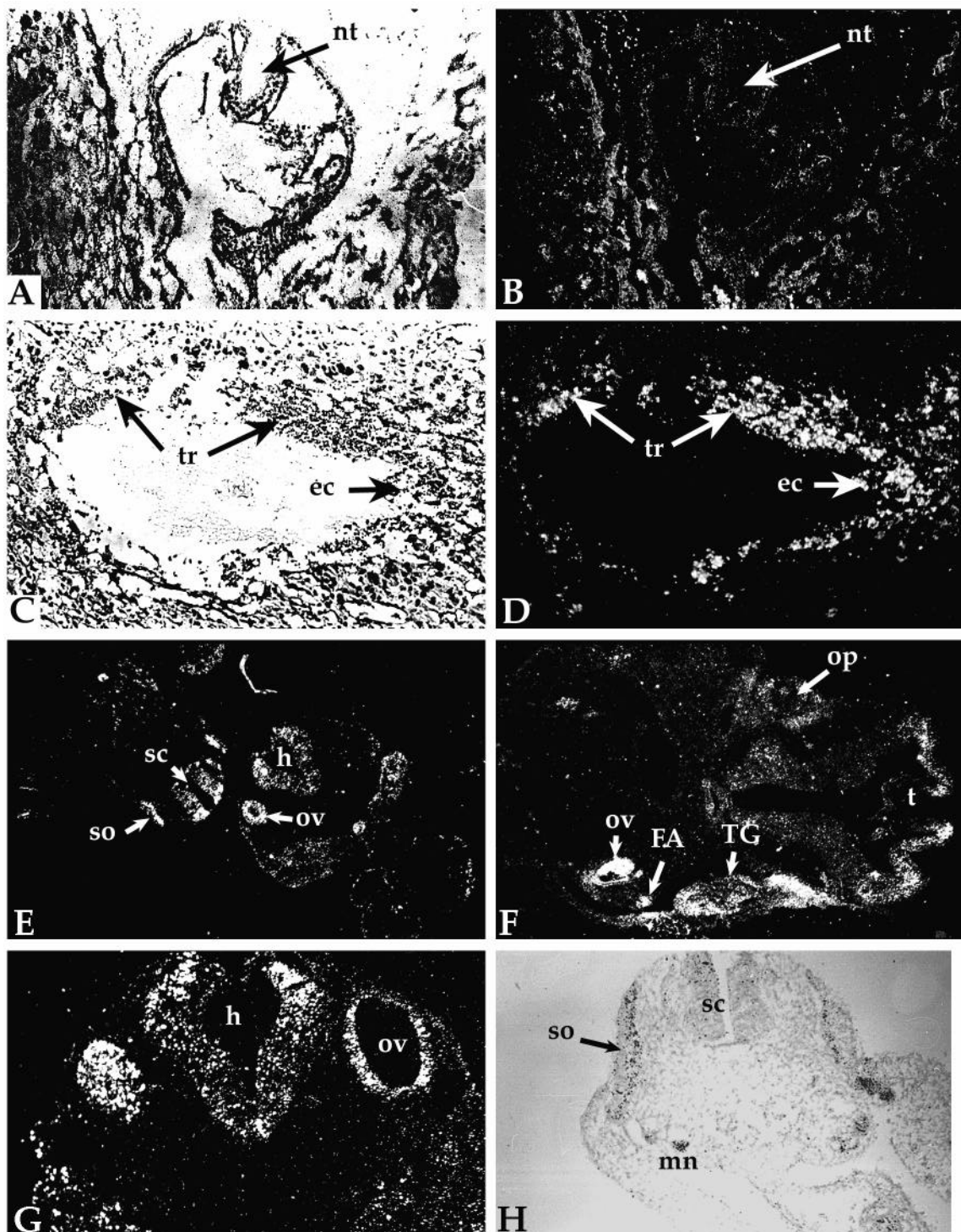
Fig. 1. mGATA-3 expression in mouse embryogenesis at e8-e10.5. Whole decidua were sectioned and in situ hybridization was performed (Materials and Methods). (A,B) Bright- and dark-field illuminations of the same cross-section through an e8.0 embryo. mGATA-3 mRNA is not detectable in embryonic tissues in this section, or in equivalent aged embryos. The unclosed neural tube is indicated (nt). Significant signal is detected in extraembryonic cells. (C,D) Strong hybridization is detected in the ectoplacental cone (ec) and trophoblast cells (tr). No embryonic tissue is present in this section. (E) A cross-section in different planes of two separate e10 embryos is shown. The upper left embryo shows expression in the spinal cord (sc) and somites (so), the lower right embryo shows expression in the otic vesicle (ov) and hindbrain (h). (F) Expression is detected in another e10 embryo in the front of the telencephalon (t), the otic vesicle, in trigeminal ganglia (TG) and ganglia of the facio-acoustic complex (FA). Expression in the eye (op) is detected at low levels in this section. (G) An oblique cross-section (taken at higher magnification than in E) of an e10 embryo showing expression in the otic vesicle and hindbrain. (H) A cross-section through the caudal end of an embryo showing expression in the spinal cord, somites, and mesonephric ridge (mn).
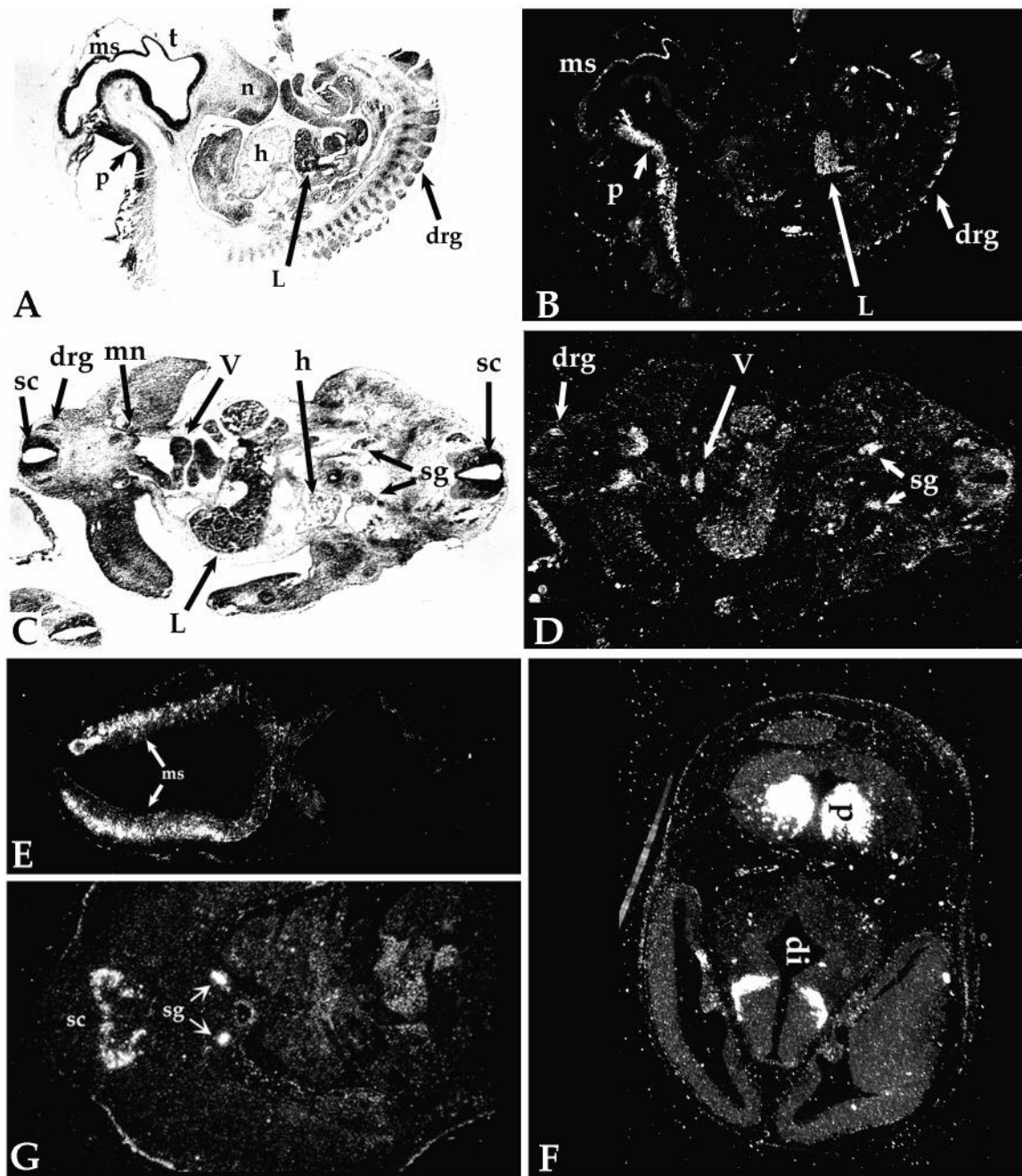
Fig. 2. mGATA-3 expression in mouse embryogenesis at e11.5-e12.5. (A,B) Bright- and dark-field illumination of a sagittal view of an e11.5 embryo. Expression along the pons and pontine flexture (p), and the mesencephalon (ms). The liver (L) also expresses GATA-3, as do the dorsal root ganglia (drg). Telencephalon (t), nose (n) and heart (h) are shown in bright field for orientation. (C,D) Bright- and dark-field illumination of the same section. Expression is found in the sympathetic ganglia (sg), more dorsally in the spinal cord (sc), dorsal root ganglia (drg), vitelline vein (V), liver (L) and the mesonephric ridge (mn). (E,F) Cross-sections through an e12.5 embryo in a progressively rostral-to-caudal direction. (E) Expression is detected in the outer half of the mesencephalon (ms). (F) A cross-section through the head, showing expression restricted to two sets of bilaterally symmetric groups of cells in the ventral half of the diencephalon (di), as well as persistent high expression in the pons (p). (G) mGATA-3 is expressed at high levels in sympathetic ganglia (sg) and spinal cord (sc) expression is now restricted to the ventral half.

At e8, we did not detect expression of mGATA-3 mRNA in the embryo (Fig. 1A,B). Significant expression is, however, detected in the ectoplacental cone and trophoblast cells surrounding the embryonic cavity (Fig. 1C,D), in agreement with the studies of Ng et al. (unpublished data) who found abundant mGATA-3 expression in placental trophoblast cells.

At e10 mGATA-3 is expressed at varying levels throughout the CNS (Fig. 1E-H). Localized expression is detected in the

somites (Fig. 1E,H). In an e10.5 embryo (Fig. 1F), localized mGATA-3 expression is detected in the anterior end of the telencephalon, in the otic vesicle and in ganglia of the peripheral nervous system (PNS), including trigeminal ganglia and ganglia of the facio-acoustic complex. GATA-3 is also expressed in the developing eye (Fig. 1F,G) and in clusters of cells along the mesonephric ridge (Fig. 1H).

By e11, expression is still diffuse throughout much of the CNS, with more localized expression within the walls of the midbrain (the mesencephalon). A sagittal view (Fig. 2A,B) shows that expression is extensive in the pons and pontine flexure, with a discrete rostral boundary. GATA-3 continues to be expressed in the otic vesicle and many ganglia of the PNS (data not shown). A transverse section reveals specific expression in the sympathetic ganglia, spinal cord and dorsal root ganglia, and also in localized regions of the developing heart, liver, mesonephric ridge and the vitelline vein (Fig. 2C,D).

At e12.5, mGATA-3 mRNA has become more restricted in expression within the mesencephalon and accumulates in the outer half (Fig. 2E). At this stage, GATA-3 continues to be expressed in the pons/myelencephalon and becomes restricted to two bilaterally symmetric groups of cells in both halves of the diencephalon (Fig. 2F). Sympathetic ganglia still express GATA-3 mRNA, while expression in the dorsal root ganglia is no longer detected and expression in the spinal cord has become restricted to the ventral half (Fig. 2G).

At e14.5, expression in the diencephalon appears unchanged from e12.5 (Fig. 3A shows a representative sagittal section). mGATA-3 mRNA continues to be present in the pons but in an even more spatially restricted pattern than at e12.5 (Fig. 3A and data not shown). Expression in the mesencephalon persists, but remains removed from the immediate layer of cells overlying the ventricle (Fig. 3A), while expression in the spinal

cord has declined significantly (data not shown). GATA-3 is not detected in the spinal cord at the approximate level of the most rostral cervical vertebra, but is detected in more caudal sections in the ventral half of the spinal cord at significantly lower levels than on previous days of development. mGATA-3 mRNA is still expressed in the inner ear and in ganglia of the PNS (including both cranial and sympathetic ganglia; data not shown). Expression is detected in the thymic rudiment (data not shown), coincident with the time when CD4$^-$CD8$^-$ prethymocytes migrate to this region (Penit and Vasseur, 1989). GATA-3 transcript is also detected in the vomeronasal organ (Fig. 3B), in glomeruli of the kidney and in the adrenal medulla (Fig. 3C,D).

In summary, GATA-3 mRNA expression in the developing mouse embryo was found to be widespread and temporally dynamic. Early in development (prior to e10), GATA-3 was expressed at very low levels, if at all, in the developing embryo. During the next several days of gestation, the relative abundance of GATA-3 mRNA increased and subsequently declined at several sites, and GATA-3-expressing cells became highly restricted to, and persistent in, morphologically and anatomically distinct groups within the CNS and PNS, the kidney and adrenal medulla and in the primitive thymus.

mGATA-3 expression during murine T cell differentiation

GATA-3 is expressed in all T lymphocyte cell lines that have been examined (Ko et al., 1991; Leonard et al., 1993; Yamamoto et al., 1990) and in the rudimentary thymus at the earliest embryonic stages defining the organ (Penit and Vasseur, 1989; data not shown), suggesting that it serves an early and therefore physiologically important function in T cell ontogeny. To test this hypothesis, experiments were under-

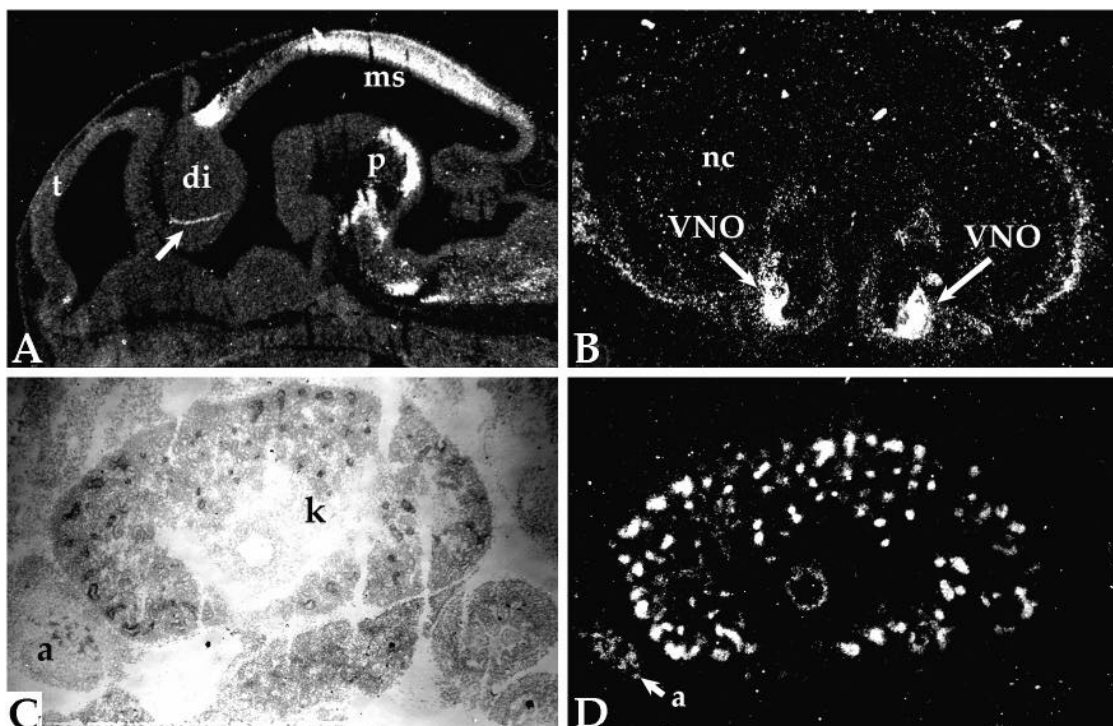Fig. 3. mGATA-3 expression in mouse embryogenesis at e14.5. (A) A sagittal section showing expression in the mesencephalon (ms), diencephalon (indicated by arrow; di) and pons (p). (B) mGATA-3 is expressed in the vomero-nasal organ (VNO), but not in the rest of the nasal cup (nc). (C,D) Bright- and dark-field illuminations of the same section. mGATA-3 is expressed in the kidney (k) and adrenal medulla (a).

taken to ascertain when mGATA-3 expression commenced during normal T lymphocyte differentiation.

Thymocytes were sorted by flow cytometry according to expression of the CD4 and CD8 antigens (Materials and Methods). RNA was recovered from the sorted cells, and RT/PCR was performed using oligonucleotides specific for mGATA-3 (Leonard et al., 1993). mGATA-3 was expressed in all four T lymphocyte fractions (CD4$^-$CD8$^-$, CD4$^+$CD8$^+$, CD4$^+$CD8$^-$, and CD8$^+$CD4$^-$; Fig. 4A, lanes 5-9, respectively). CD4$^+$ cells expressed mGATA-3 at somewhat higher than CD8$^+$ cells, but all lymphocyte subpopulations examined expressed mGATA-3 mRNA. Since CD4$^-$CD8$^-$ T cells define a very early stage of T lymphopoiesis, these data suggest that mGATA-3 is involved in very early aspects of T lymphocyte differentiation.

Fig. 4 also shows that mGATA-3 is expressed in a number of murine cell lines: BW 5147.3, EL4 (mouse T cell lines) and C1300 (a mouse neuroblastoma line; Pons et al., 1982) cells express abundant mGATA-3 mRNA (Fig. 4A, lanes 1 and 2), while BALB/c 3T3 fibroblasts do not (Fig. 4B, lanes 1, 2 and

Fig. 4. Expression of mGATA-3 during T-cell development and in murine cell lines. Total cellular RNA was isolated from T lymphocyte lines BW 5147.3 (A, lanes 1, 2) or EL4 (B, lane 1), neuroblastoma line C1300 (B, lane 2), B lymphocyte line CH27 (A and B, lanes 3), BALB/c 3T3 cells (B, lane 4) or circulating lymphocytes (A, lane 4); the circulating lymphocytes were further fractionated by cell sorting into populations expressing CD8 and CD4 antigens (A, lane 7), neither CD4 nor CD8 (A, lanes 5, 6), only CD4 (A, lane 8) or only CD8 (A, lane 9). ES cells were either propagated normally (panel B, lane 5) or subject to in vitro differentiation for 9 days (B, lane 6) or 12 days (B, lane 7)(Lindenbaum and Grosveld, 1990). RNAs were then reverse transcribed and co-amplified by PCR using internal control (murine S16 mRNA) primers as described (Materials and Methods; Leonard et al., 1993).

4; Landry et al., 1993). mGATA-3 mRNA is prevalent in embryonic stem cells, but its abundance decreased dramatically after in vitro differentiation (Fig. 4B, lanes 5-7; Keller et al., 1993; Lindenbaum and Grosveld, 1990). These studies were extended by immuno-precipitation experiments employing anti-mGATA-3-specific monoclonal antibodies; they confirmed that GATA-3 mRNA expression parallels the accumulation of abundant GATA-3 protein in T cell lines and in C1300 neuroblastoma cells (Ng et al., unpublished data; Yang et al., 1994). The (murine) monoclonal antibodies, however, failed to recognize mGATA-3 in either embryonic whole mounts or tissue sections (M. E. R., data not shown).

Structure of the *mGATA-3* gene

In order to initiate regulatory analysis of the *mGATA-3* gene, the locus was cloned and characterized. A murine BALB/c genomic DNA library (a gift from D. I. H. Linzer) was screened (Maniatis et al., 1982) using an mGATA-3 cDNA clone (plasmid mc5b8; Ko et al., 1991) as probe. Positive clones were plaque-purified using probes corresponding to either the 5′ or 3′ ends of the mGATA-3 cDNA. Four independent *mGATA-3* genomic recombinants (λ7e, λ5a, λ2b, λ10b) were further characterized by restriction enzyme mapping and then subcloned into plasmid vectors. The final map of the *mGATA-3* gene was determined using these individual subclones as probes in Southern blotting of the parental λ recombinants (Fig. 5A).

The overlapping genomic λ clones were found to describe the entire coding region of the *mGATA-3* locus [approximately 23 kilobase pairs (kbp)]; they also contained 13 kbp of 5′ and 8 kbp of 3′ flanking sequence (Fig. 5A). The gene is composed of six exons: the first exon (E1) consists entirely of 5′ untranslated sequence while the second exon (E2) contains the initiation codon for mGATA-3 translation (see below). The amino and carboxy zinc fingers are encoded in exons E4 and E5, respectively, while the 3′ untranslated region and polyadenylation signal are located within E6 (Fig. 5B). The intron-exon boundaries were shown to conform clearly to the general splicing consensus (Fig. 5B; Mount, 1982). Thus the overall organization of the *mGATA-3* gene is quite similar to that of the *mGATA-1* and *cGATA-1* factor genes (Hannon et al., 1991; Tsai et al., 1991).

Primer extension assays were performed to identify the mGATA-3 transcription initiation site. RNA was isolated from BW5147.3 (mouse T lymphoma) and MEL (mouse erythroleukemia) cells; mGATA-3 is abundantly expressed in a variety of murine T cell lines (Fig. 4), but expressed at only very low levels in MEL cells (Leonard et al., 1993). Primer extension (Leonard and Patient, 1991) of BW5147.3 mRNA using primers corresponding to sequences within exon 1 (mG3PE1) or exon 2 (mG3PE2; Materials and Methods), yielded products of 187 and 395 nucleotides (nt), respectively (Fig. 6A, lanes 11 and 12 and Fig. 6B, lanes 1 and 2). The endpoints determined with the two different primers were identical.

S1 nuclease protection assays were also used to confirm the site of mGATA-3 transcriptional initiation. Kinase-labeled oligonucleotide mG3PE1 was annealed to genomic subclone R1.3 (Fig. 6A) containing the (presumptive) first exon and 308 bp of 5′ flanking sequence of the *mGATA-3* locus, and extended using Klenow polymerase. This probe was then
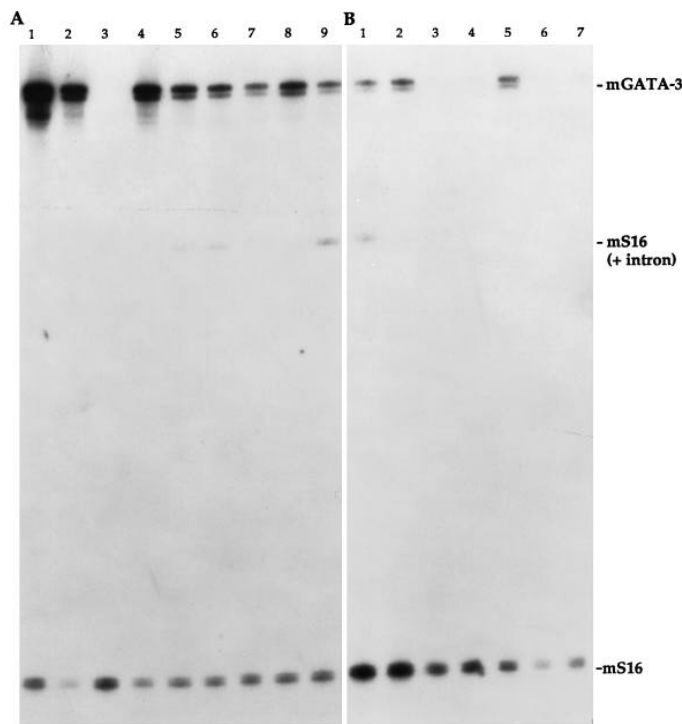
A

E1 E2 E3 E4 E5 E6

-10.   -5.   +.001   5.   10.   15.   20.   25.   30.

S   S   R RS   N S   R   R S   R R   RS   S   R   S R   S   RS   R   S

B   H H   B   B   H B   B B   H   H H   H   B H H

λ7e

λ5a

λ2c

R1.3   λ10b

B

```
                    EXON 1 (noncoding 5' end; 189 nt)...TCAGgttggt...Intron 1 (418 nt)...
ttgcagGTGA....EXON 2 (ATG; 630 nt)...............CACGgtgagt...Intron 2 (2.5 kb)...
tcacagGGAG....EXON 3 (536 nt)...................ACAGgtgagt...Intron 3 (6.5 kb)...
ttccagAAGG....EXON 4 (amino finger; 145 nt).......GCTGgtaagt...Intron 4 (5.6 kb)...
atccagTCGG....EXON 5 (carboxy finger; 125 nt).....TAATgtaagc...Intron 5 (5.5 kb)...
ttttagATT.....EXON 6 (3' end; 828 nt)
```

C

```
  -2052                                              ct cgagacttttt ctcgaagctg tgaactaggg actggctgac tgatgcctta
  -2000 tacaactcat gagtgataca gcaatagtat ttttcccttt taaactctca ttctagatct taatcccggg cctttagca aaaatgtgtg tgcggtcttc
  -1900 tctgcttgtt tatggaatgg tttccggtga atgttagaca gcaactaaat aattaaaagg ttttgagagc gccgtgagct ggaagagatg ctggcagctt
  -1800 gattggccca ggagagatgg gttggtcaaa gagtgaaagt aaaaaatcg atattctaag actgggccgc cctcataaac gtgggaccaa ggggtgcagg
  -1700 gtgaactcag aagggaggga agactcgttt ttaaaaaaaa ggtatcttcg taatttgtat ttagaaattc agccttgatt tcagtcaaag atcagatgta
  -1600 aaaggggggg gggtgtcagg agaagctca aacccaagcc cgcacaggca tttctaaata aattttgtct tcttcaaggg ttgtattgta aaaagaacgc
  -1500 accttccctc tccccgcact ccactggctt caaggtagat attgatggct tttctccaca gtagaataga agaagttact tggaacttttt aagaaagggc
  -1400 ttgtcgggaa tccaagtttt gaataagtgg gaaatattat ggggtaaaga aatgaagagg ggatgacgcc ttcagttcca gtaaccctct aaagccaagc
  -1300 tctctgccct ggagggtgga atgggaacac ctcttacact gccctgggt gcaccgaag tctctttcgg accttgctca tccgggcagt gttttctgga
  -1200 cccaggctgc agaaacctca acgccactag gggatgcagt gactcaaagg acggaacagg gccgcagagg ttggagcccg tggggcgggg cgactgcggc
  -1100 ccggtcgtgg tatctcccga tgattggtcc ttacaaatgg ccccgccctc gtccgcccag cctctctact gggcgtcttc cagcctcctc cctgtgtcag
  -1000 cttgcgccct ccctccagac cctgtccaga gacccaaat gtcaggggag ttgcggctgg gtcacccggg aggtcagggt tggatattat agaggggacc
   -900 gtttttcctcg ggtgaaagca gaatataagta gcacccccgag ctaggccgaag agaaaagtta ggctatcgca gcaaagatca agttagttgt acacggtact
   -800 tcggggacca gtgagcctag gtctcgctaa gccagctttt tatgcatcgc gttgtcacta aggtcaaaag cacacattga tgatatcatt agatccgaaa
   -700 aagacagctc gcaacagtga cgctgttcgt tctggagagg tttgggaaag caagcagaga ccataacaat aacgcgggcg tccgaatcaa agcccaggtc
   -600 ctccgcgcga ttcagcagcc tcccggaaag ctggttcgga ggcaaatctt cagttacttc gccatgaaac tttctaatc tgttttacat gcacatagcc
   -500 cctaagctga tataaactga aaaacgttct ggcttgaatc ctaatttaaa agtttgctca tttttaaaaa agtatcgaaa agtctgctta tgggaaattag
   -400 ctttatgaat ggggcagccg ggtttcactc gtacggaata atctactgag cccgaatgaa tcggctttgc tacatttaaa gggccagaga gcgaattccc
   -300 tcctgcctgt ccctctgcc ccaagcccgg gcgggcctcg cccaccctgc ccgcggcccc tctgccctag gcggcccttg gcggcgcccc tttccggtca
   -200 gtagaggggc gggagagggg gcgcggcggtgt gccggggtgg gtgggggga tgggatgggg gtgaggtga agtcctggag cgggtttggg ttgcagtttc
   -100 cttgtgctga ggatcgtgta ccctcctgcc accgcaggct cctccccctc ggcgcggatg acactagaac ctccttaagt tgcgtcgcgc cacagctgtc
    +1 TGCGAACACT GAGCTGCCTG GCGCCGTCTT GATAGTTTCA GAAAGAATGC ATTCCCTGTA AAAAAAAAT ACTGAGAGAG GGAGAGGAGA AAGAGAGAGA
  +101 GACTGAGAGA GCGAGACATA GAGAGCTACG CAATCTGACC GGGCAGGTCA CACGCCTCCT CCTCCTCCTC TACGCTCCTT GCTACTCAGg ttggtattgt
  +201 gacttgtgac tttttttttt tttttaagtt tgatatcact ccctcctctt ggagggactg tttgtgtgtg tgtgtttggg ggtttgtttg tttggttttt
  +301 tgtaaactaa aagctcagct ttctccttttt ggaggatttg tagtggttgg aaagatcctg gggycggtgt cctccaggga ggacttgtcc caagttttaa
  +401 ttgcttagtc agtcctgagc tcctgcgagc ctggcttcga gaggtgtgct gcggagcttg gcttctttnc nccgggactt cgngtgcgtg tgcacgtccc
  +501 gggtgagtcg gccctgcagg aagcttgcga agacctagtg ccctcccaga attcaggtgt taaaagttct tacacctcat tcaggtctct ctttctctcc
  +601 cttgcagGTG ATCGGAAGAG CAACCGTCTC TGAGCGCCAA GGAATCAGTG TGCAGTGTGG TCACACTCGG ATTCCTCTCT CCCTCCTTTT TTTTTTTTTT
  +701 TTTTTGACCC CTTTATTCCT CCGTGTCTGC TTTTTTTTTG GGGGGGGGGA TCGCCCTCAT TCTTTTCTTT TTCTTCTTTC CCTTCCTTTC TTTTGCTAAA
  +801 CTATCCCGCA AAGATTTTTC TTTCCTCCCT AAACCCTCCT TTTTGCTCTC CTTTTCTATA CCCTTAACTG CAAACAAACC ATTAAACGAC CCCTCTCCTG
  +901 GGCCTCCGAC GGCAGGAGTC CGCGGACCTC CCAGGCCGAC AGCCCTCCCT CTACCCGCGA GGTTCCGGGC CGGGCGAGAG GGCGCGAGCA CAGCCGAGGA
 +1001 CATG
```

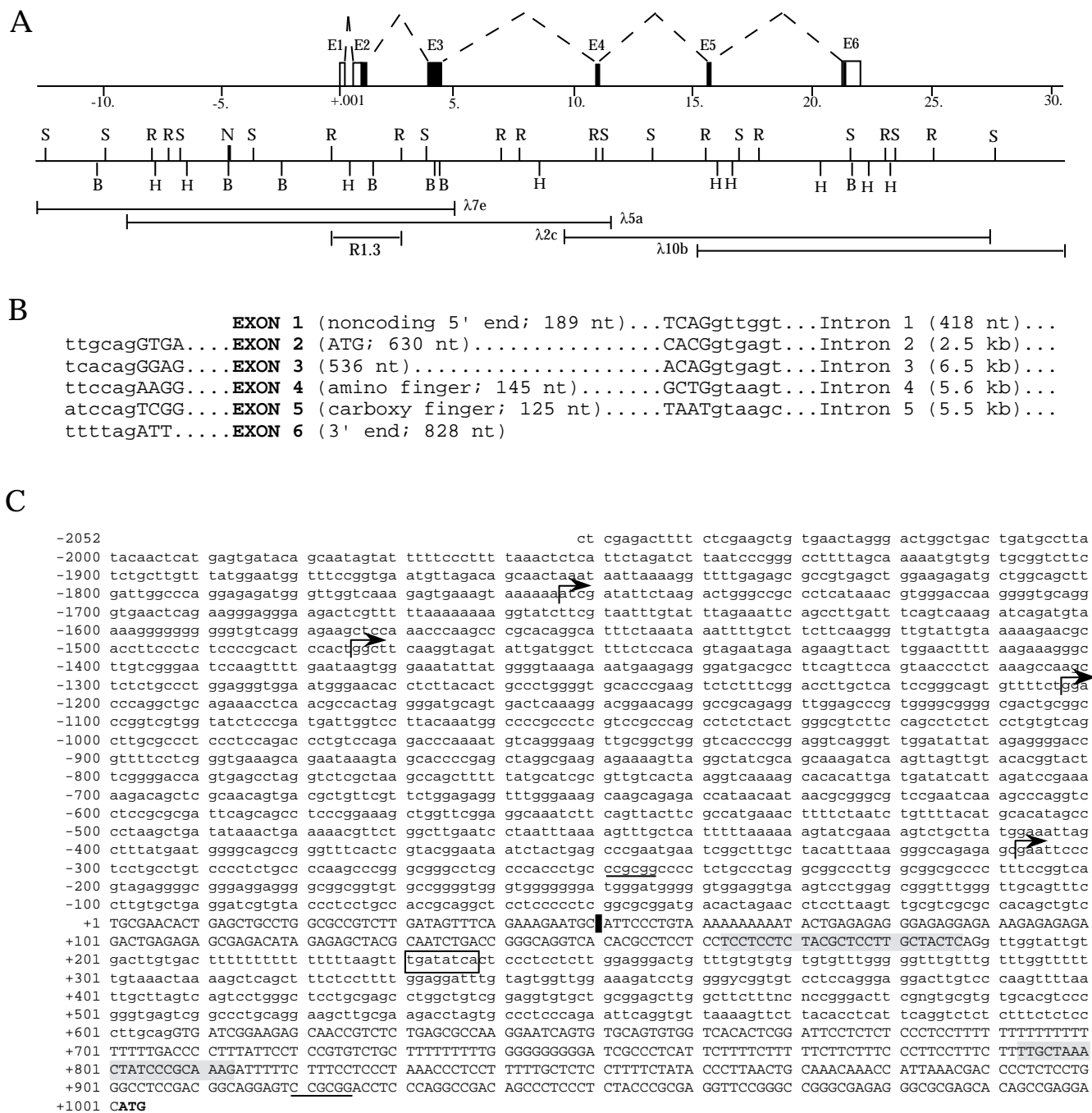Fig. 5. Structure and partial sequence of the murine GATA-3 gene. (A) The overall organization of the *mGATA-3* gene is shown on the top line: E1 through E6 represent the approximate sizes and positions within the locus of each of the *mGATA-3* exons, where filled boxes indicate translated, and open boxes untranslated, sequence. Dashed lines indicate the positions of each of the introns. The second line depicts an abbreviated restriction enzyme map for the *mGATA-3* locus, drawn to the same scale, using restriction enzymes *Sac*I (S), *Bam*HI (B), *Eco*RI (R), *Hin*dIII (H) or *Not*I (N). The four parallel lines beneath the restriction map indicate the approximate sizes and degree of overlap of the four genomic λ clones describing the locus, and the subclone of λ7e (R1.3) used for construction of transient transfection and transgenic reporter genes (Materials and Methods). (B). Intron-exon organization of the *mGATA-3* locus. (C). DNA sequence of *mGATA-3* from the most CAP-site proximal genomic *Xho*I site (−2052) through the translation initiation codon in the second exon (+1002, ATG, heavily underlined). Capital letters indicate nucleotides present in mGATA-3 mRNA (Ko et al., 1991), while lower case letters represent 5′ flanking or first intron sequences. Nucleotide +1 was assigned on the basis of S1 and primer extension assays (Fig. 6). The arrows depict each of the 5′ end of deletion endpoints examined in transient transfection assays (Fig. 7). The heavy vertical bar (at nt +50) indicates the position where the minimum promoter (beginning at nt −308) was linked to the *CAT* gene in transient transfection assay (Fig. 7A). The double GATA-binding site in the first intron referred to in the text is boxed, while the transcription initiation site deletion construct (Δ*Sac*II; Fig. 7B) was generated by cleaving with *Sac*II at the two sites indicated by underlining (centered at −248 and +924). The two sequences corresponding to primer extension oligonucleotides mG3PE1 and mG3PE2 (Materials and Methods) are shaded.
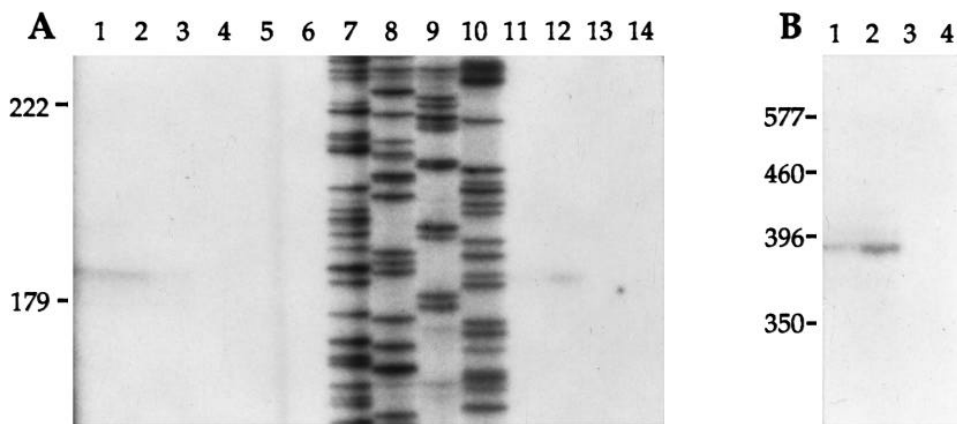
Fig. 6. 5′ end analysis of GATA-3 mRNA. (A) GATA-3 mRNA isolated from murine BW5147.3 T cells (lanes 1-3, 20 μg of cytoplasmic RNA) or murine erythroleukemia cells (lanes 4-6, 40 μg of cytoplasmic RNA) was hybridized to a radiolabelled primer extension product of genomic subclone R1.3 (Fig. 5A) and digested with S1 nuclease as described (Materials and Methods) for 20 minutes (lanes 1, 4), 40 minutes (lanes 2, 5) or 60 minutes (lanes 3, 6). At the same time, primer extension reactions (Leonard and Patient, 1991) were carried out using oligonucleotide mG3PE1 (Materials and Methods) to prime synthesis from either 10 (lane 11) or 20 (lane 12) μg of BW 5147.3 mRNA or 20 (lane 13) or 40 (lane 14) μg of MEL cell mRNA. A sequencing reaction from genomic subclone R1.3 using mPE1 as primer was electrophoresed in parallel (lanes 7-10 are G, A, T and C reactions, respectively). (B) Identical mRNAs to those used for the primer extension reactions shown in lanes 11-14 were carried out using oligonucleotide mG3PE2 (lanes B1-4, respectively). Size markers (in nt) are shown on the left of both panels.

hybridized to total RNA isolated from BW5147.3 cells. After digestion with S1 nuclease, the protected fragment was found to migrate with a mobility of 189 nt (Fig. 6A, lanes 1-3), showing that the primer extension and S1 nuclease protection products correspond to the same nucleotide. No specific product was identified using MEL cell mRNA in primer extension (Fig. 6A, lanes 4-7 and Fig. 6B, lanes 3, 4) or S1 nuclease protection experiments (Fig. 6A, lanes 13, 14).

Transcriptional regulation of *mGATA-3*: transfection into murine cell lines

Transient transfection assays were initially employed to identify *cis*-regulatory elements that might be responsible for the tissue-restricted expression pattern of mGATA-3. Constructs containing segments of the *mGATA-3* locus driving expression of the *chloramphenicol acetyltransferase* (*CAT*) reporter gene were transfected into GATA-3-expressing and -non-expressing cell lines. CAT enzyme activity was determined from extracts normalized for transfection efficiency based on cotransfected pRSV.luciferase activity. A neuroblastoma line, C1300, and a T lymphoma line, BW5147.3, were used as representative examples of cells that express mGATA-3 in neural crest-derived neuronal (neuroblastoma; NB) or T lymphocyte (T) cell lineages, respectively (we were unable to identify a CNS cell line that expresses GATA-3 for these studies). BALB/c 3T3 fibroblasts cells do not express mGATA-3 (Fig. 4B, lane 4) or other GATA-binding activities (Landry et al., 1993), and were used as a negative control.

A segment of the *mGATA-3* gene encompassing the presumptive minimal promoter, from nt −308 to +50 (Fig. 5C), was examined first (Fig. 7A). After transfection into all three cell lines, the transcriptional activity was significantly higher (from 6- to 50-fold) than that of a plasmid in which sequences surrounding the mGATA-3 CAP site were deleted (ΔSac2; compare Fig. 7, rows A and B). Since the transfection efficiency between cell lines varied significantly, changes in transcription levels were normalized to the relative activity of 308pCAT (set equal to 1; Fig. 7A).

We next examined a plasmid containing contiguous DNA sequences from −308 through the mGATA-3 translation initiation codon in exon 2, at nt +1002 to 1004 (Fig. 5C). This plasmid was 7-fold more active in T cells and 4-fold more active in NB cells, but no more active in 3T3 cells, than 308pCAT. Thus a positive transcriptional stimulatory element appears to reside within the region that also contains the *mGATA-3* first intron (I1). Deletion of 308pI1CAT from the 5′ end to nucleotide −50 had no effect on CAT activity (data not shown), suggesting that this positive transcriptional *cis*-effector sequence lies between nt +50 and the translation initiation codon.

We next assayed sequences extending 5′ to −308 of the *mGATA-3* promoter. 1203pI1CAT showed unaltered transcriptional activity in GATA-3-expressing cells, but had significantly lower activity in fibroblasts (compare Fig. 7C,D). The net effect of including additional promoter sequence was to lower non-specific (3T3 cell) transcription, leading to more than 10-fold higher activity of 1203pI1CAT in GATA-3-expressing cells than in fibroblasts. Addition of even further 5′ sequence, including all of the sequence shown in Fig. 5C, had no effect greater than 2-fold (Fig. 7E,F and data not shown).

Since the addition of nt +50 through +1002 to a minimal promoter stimulated transcription in cells expressing GATA-3 (Fig. 7C), we next addressed the question of whether or not I1 alone displayed classical (orientation- and position-independent) tissue-specific enhancer activity. I1 was cloned 5′ to 308pCAT in both the sense and anti-sense orientations (Materials and Methods), and was found to enhance promoter activity very modestly in the sense orientation (Fig. 7G). In the antisense orientation, I1 significantly suppressed promoter activity in all three cell types (Fig. 7H). Thus, I1 does not act as an enhancer of mGATA-3 transcription (Fig. 7G,H).

In summary, these data indicate that at least two distinct *cis*-regulatory regions lie within the 3 kbp surrounding the mGATA-3 CAP site, which together act to confer significant (>10-fold differential; Fig. 7D) tissue specificity controlling mGATA-3 transcription after transient transfection into cell lines. These regionally localized elements include a position-dependent positive sequence within the *mGATA-3* first intron (compare Fig. 7A,C), as well a negative one between nt −308 and −1203 that suppresses mGATA-3 transcription in cells
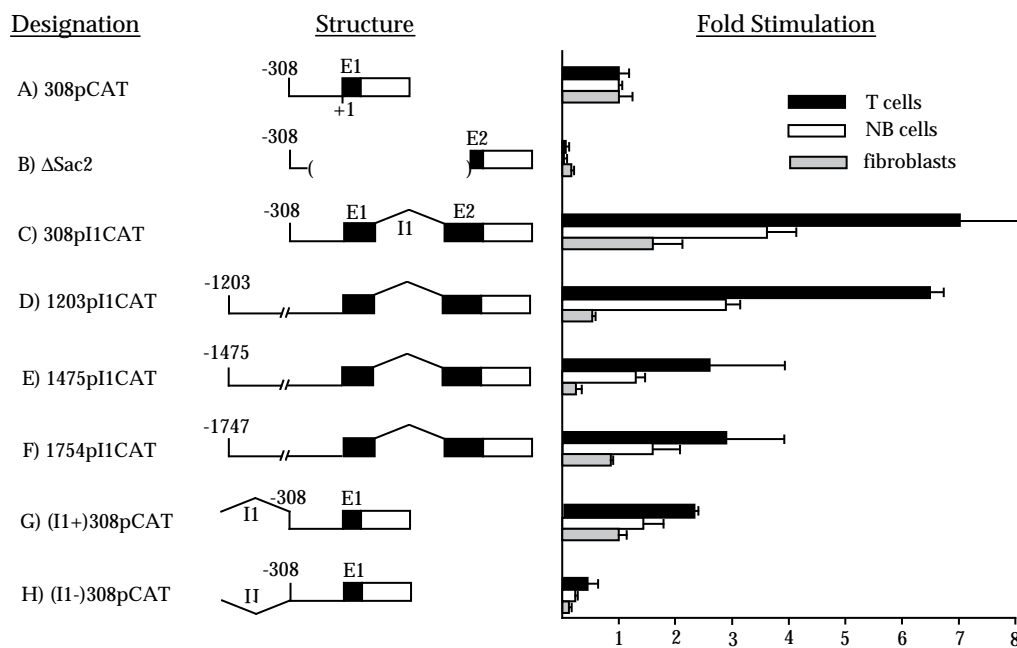
Fig. 7. Expression of mGATA-3-directed reporter genes in transiently transfected murine cell lines. The *CAT* reporter gene, transcriptionally directed by various segments of the *mGATA-3* locus (Fig. 5C), were transfected into cell lines that either express (BW 5147, black bars; and C1300 cells, open bars) or do not express (BALB/c 3T3 cells; shaded bars) mGATA-3 (Fig. 4). Cotransfected pRSV.luciferase was used as the internal control for transfection efficiency (Materials and Methods). Construction details for the diagrammatic representations shown in the second column (Structure) are given in Materials and Methods: the thin line represents mGATA-3 5′ flanking sequence, filled boxes represent *mGATA-3* exons, the first intron (I1) is represented by the inflected line, and the open box represents the position of the *CAT* reporter gene. Each transfection was repeated, in duplicate, on (at least) three separate occasions; the s.e.m. for the transfections was generally less than 25%. The values shown represent the average CAT activity, normalized to that of a minimal promoter element [A) 308pCAT], which was arbitrarily assigned an activity of one in all three cell types.

where the gene is not normally transcribed (compare Fig. 7C,D). In their normal configuration and orientation within the locus, these elements impart significant (>10-fold) cell-type-specificity to mGATA-3 transcription after transient transfection.

Transcriptional regulation of mGATA-3: transgenic mice

We next sought to determine whether the regulatory elements identified by transient transfection analysis would also obey the more stringent transcriptional criterion of directing appropriate mGATA-3 regulation in vivo. Fertilized embryos were injected with presumptive regulatory elements of the *mGATA-3* locus driving expression of the bacterial gene (*lacZ*) encoding β-galactosidase (β-gal). A construct containing contiguous sequences from −2052 through +1004 of *mGATA-3* (the entire sequence shown in Fig. 5C) was fused to the initiation codon of the *lacZ* gene, microinjected into fertilized eggs and then transferred into pseudopregnant foster mothers. Staged embryos were then stained for β-gal activity. Eight of forty-five microinjected embryos were determined to be transgenic by Southern blot analysis, with transgene copy numbers ranging from 4 to greater than 20 (data not shown). Four of the eight transgenic embryos displayed β-gal expression; we assume that the embryos that failed to express the reporter gene reflect suppression due to integration site effects (Dillon and Grosveld, 1993; Hogan et al., 1986). One of these embryos was lost during manipulation, while the three remaining embryos all exhibited β-gal staining.

Whole-mount staining of an e10 embryo displayed *lacZ* expression in trigeminal and facio-acoustic ganglia, the sym-

pathetic trunk, the otic vesicle and somites (Fig. 8A,B). An e11.5 embryo showed lower intensity staining in some cranial nerves, somites and the ribs (Fig. 8C,D). An e12.5 embryo showed strong β-gal activity in many ganglia of the PNS, including the trigeminal and facio-acoustic ganglia, the sympathetic trunk and dorsal root ganglia (Fig. 8E-G). In this embryo, discrete staining is also detected in what appear to be individual cells at the rostral end of the telencephalon, in the eye, the vomeronasal organ and again in the ribs and ear. Transverse sections of this embryo revealed low-level β-gal activity in the ventral half of the spinal cord and in the neural layer of the retina (data not shown).

Clearly many regions of the e12.5 transgenic embryo display a transgene staining pattern coincident with the mRNA expression profile derived from in situ hybridization (Figs 1-3 and data not shown). One location displaying a clear difference is the midbrain: in cross-sections of the e12.5 transgenic embryo, no β-gal staining is observed in the mesencephalon or pons region, while GATA-3 mRNA is clearly present in these regions at this same stage (Fig. 2). Thus a transgene containing the *mGATA-3* sequences shown in Fig. 5C, when linked to a *lacZ* reporter gene, faithfully recapitulates almost the entire normal expression pattern of mGATA-3 in transgenic mice, with the exception that the very prominent CNS accumulation of mGATA-3 is not reproduced. These data therefore confirm and extend the transient transfection studies and indicate that transcriptional control elements required for the tissue- and developmental stage-specific expression of mGATA-3 lie within approximately a 3 kbp segment of the gene (Fig. 5C), but also that the element(s) conferring appropriate CNS regulation likely lie outside these boundaries.
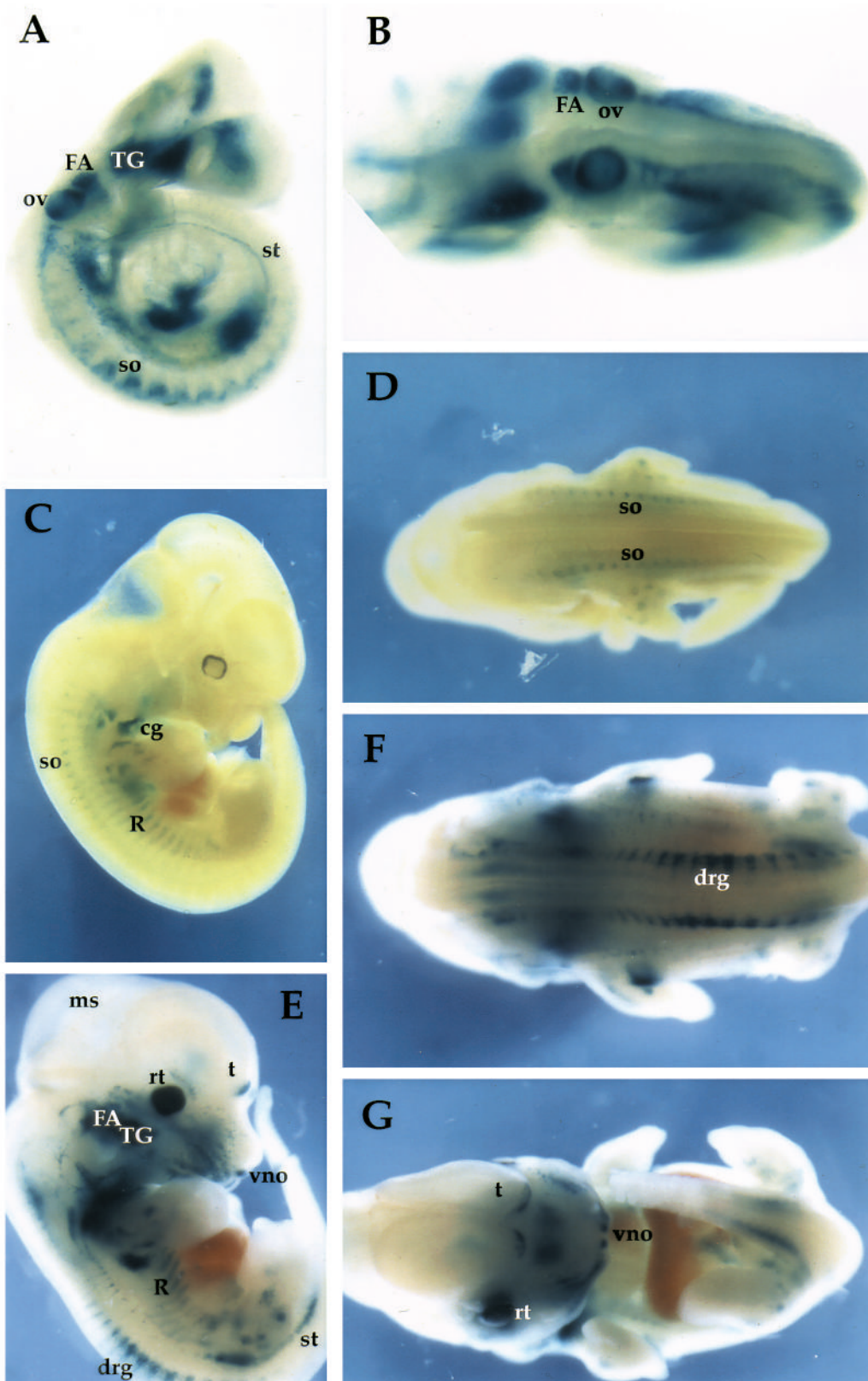
Fig. 8. Expression of mGATA-3-directed β-galactosidase synthesis in transgenic murine embryos. The entire sequence shown in Fig. 5C was fused to the bacterial *lacZ* gene and injected into fertile mouse eggs. At various times after implantation, the embryos were taken and stained for β-gal activity. (A,B) Lateral and dorsal views, respectively, of an e10 transgenic embryo showing β-galactosidase activity in trigeminal ganglia (TG), facio-acoustic ganglia (FA), the sympathetic trunk (st), otic vesicle (ov) and somites. (C,D) Lateral and dorsal views, respectively, of an e11.5 embryo showing weak transgene expression in some cranial ganglia (cg), somites (so) and the ribs (R). (E-G) Lateral, dorsal and ventral views, respectively, of an e12.5 embryo showing transgene activity in trigeminal and facio-acoustic ganglia, the sympathetic trunk, in dorsal root ganglia (drg) and the eye (specifically, in the neural retina; rt, see text), in the vomeronasal organ (vno) and at the rostral end of both lobes of the telencephalon (t). The position of the mesencephalon (ms) is indicated.

DISCUSSION

In the present study, we describe the sites of most abundant mGATA-3 expression from the onset of its first detectable embryonic accumulation to e14.5 of development, as well as the expression during T cell ontogeny and the cloning and physical characterization of the *mGATA-3* locus. In addition, we present a preliminary regulatory analysis, which describes

the overall sequence requirements for transcriptional control of the *mGATA-3* gene using both transient transfection into GATA-3-expressing murine cell lines and analysis in transgenic mouse embryos. The obligatory first step in assigning potential regulatory function(s) to a transcription factor is to first clearly define its expression pattern. A dearth of information presently exists regarding both the expression pattern and regulation of GATA-3. The present studies should significantly illuminate possibilities for when and where GATA-3 might exert transcriptional regulatory effects.

Despite common features in overall GATA gene structure, there may be significant differences in transcriptional initiation and regulation between the *GATA-1* and *GATA-3* genes (Hannon et al., 1991; Tsai et al., 1991). For example in T cells, transcription of the *mGATA-3* gene initiates at a single, distinct site located 189 nucleotides 5′ to the first intron (Fig. 6), in contrast to both *mGATA-1* and *cGATA-1*, for which extensive mRNA 5′ end heterogeneity was found in erythroid cells with multiple, broadly scattered transcription initiation sites (Hannon et al., 1991; Tsai et al., 1991).

Transcriptional regulation of the *mGATA-1* gene appears to be dependent on a simple organization of two promoter elements: one contains duplicated GATA sites, while a second contains a duplication of a CACCC factor recognition motif. Both motifs significantly contribute to expression of the *mGATA-1* gene in transient transfection assays since mutation of either element results in a four-fold reduction in promoter activity (Tsai et al., 1991), activation values similar to those established in the transfection experiments reported here (Fig. 7). Regulation of the *cGATA-1* gene promoter is somewhat more complex, but again appears to involve GATA sites as a critical autoregulatory element (Schwartzbauer et al., 1992). Although GATA-binding sites were found within the *mGATA-3* first intron, site-specific mutation of these sites resulted in no change in mGATA-3 transcriptional activity (K. G., unpublished observations).

The initial observation that GATA-3 is expressed in the embryonic brain (Ko et al., 1991; Yamamoto et al., 1990) and the further finding that it is also expressed in specific cell layers of the chick optic lobes during embryonic development (Kornhauser et al., 1994) prompted a more detailed analysis of the embryonic expression pattern of mGATA-3 in the murine CNS and PNS. We show here that GATA-3 expression in the nervous system is quite low early in embryogenesis (prior to e10), but later becomes localized to specific regions of the nervous system (cranial/sympathetic ganglia, diencephalon, mesencephalon, spinal cord). Several anatomically distinct loci that express mGATA-3 at e12.5 (e.g. the telencephalon, retina and spinal cord) fail to continue to express the factor by e14.5. mGATA-3 expression in the spinal cord at e10 is diffuse (Fig. 1H), becomes restricted to the ventral half by e12.5 (Fig. 2G), and then declines by e14.5 (data not shown) coincident with the time in development when spinal cord neurons no longer undergo mitosis (Nornes and Carry, 1978). Thus, GATA-3 mRNA expression appears to undergo both localized induction and suppression at a number of sites during gestation.

By 12.5 days of gestation, mGATA-3 expression in the CNS condenses within specific cell layers of the mesencephalon, diencephalon and pons. mGATA-3-positive cells reside in the outer layers of the mesencephalon, distal to the ventricle. In

this region of the brain, neuroblasts occupy the cell layers directly bordering the ventricle and, as they cease mitosis, these neurons migrate radially from the ventricular zone (Angevine and Sidman, 1961). This suggests that GATA-3 expression corresponds to a specific stage of maturation during the differentiation of CNS neurons, a view supported by analogous studies of GATA factor expression in the developing optic lobes of the chicken (Kornhauser et al., 1994).

Having determined the embryonic expression pattern of mGATA-3 (Figs 1-3), it became of interest to define the specific *cis*-regulatory elements required for the generation of this highly restricted expression pattern. We therefore characterized the genomic locus and then, by transient transfection into cell lines, identified candidate transcriptional control elements both in the first intron and in the distal part of the promoter. However, given the rather modest differences in activation of these reporter genes in the GATA-expressing and -deficient cell lines (Fig. 7), and given the significant specificity of reporter gene control elicited by these same regulatory sequences in vivo (Fig. 8), we are uncertain as to the continued efficacy of using transient transfections to refine the possible in vivo transcriptional regulatory mechanisms that are employed in controlling this gene.

To determine whether the regulatory elements identified by transfection into cell lines represent functional regulatory elements in vivo, the bacterial *lacZ* gene was used to create an mGATA-3 regulatory sequence-directed transgene. When injected into fertilized mouse eggs, this transgene recapitulates the normal mGATA-3 expression profile almost entirely, including expression in cranial and sympathetic ganglia, dorsal root ganglia, spinal cord, retina, somites and vomeronasal organ. The most notable exception to the in vivo mRNA expression pattern is the absence of β-gal staining in the midbrain, and we therefore conclude that this transgene does not contain the full complement of regulatory sequences required for complete transcriptional control of mGATA-3. We are currently examining other segments of the locus to reconstruct the complete regulatory pattern of mGATA-3 transcription, and are also extending these observations using anti-mGATA-3 antisera to determine whether GATA-3 expression in the CNS might be post-transcriptionally controlled (M. E. R., unpublished observations).

Since mGATA-3 displays a complex pattern of expression during development and since the trangenic embryos represent transiently analyzed animals examined at different days of gestation, these initial transgenic studies must be interpreted with considerable caution. However, each of the transgenic embryos displayed staining patterns that are similar to those resolved by mRNA in situ hybridization analysis at comparable stages, and each also showed an expression pattern that was developmentally consistent from animal to animal. These initial results therefore indicate that the segment of the *mGATA-3* locus from nt −2052 to +1004 is necessary, but not sufficient, to recapitulate accurately the endogenous mGATA-3 transcription pattern. As with all transgenic analysis, position effects and mosaicism in founder embryos must be accounted for, and thus further analysis of transcriptional regulation of mGATA-3 using established lines to refine the activity of identified elements (e.g. in the PNS and kidney), as well as to identify currently undefined elements (e.g. in the CNS), appears to be clearly warranted.

REFERENCES

Angevine, J. B. and Sidman, R. L. (1961). Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature* **192**, 766-768.

Arceci, R. J., King, A. A. J., Simon, M. C., Orkin, S. H. and Wilson, D. B. (1993). Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol. Cell. Biol* **13**, 2235-2246.

Ausubel, F. M., R. Brent, R. E. Kingston, D. M. Moore, J. G. Seidman, J. A. Smith and Struhl, K. (1989). *Current Protocols in Molecular Biology*. New York: John Wiley and Sons.

Choi, O.-R. and Engel, J. D. (1986). A 3′ enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. *Nature* **323**, 731-734.

Dillon, N. and Grosveld, F. (1993). Transcriptional regulation using transgenic mice. In *Gene Transcription: A practical approach* (ed. B. D. Hames and S. J. Higgins). Oxford Univ. Press.

Dorfman, D. M., Wilson, D. B., Bruns, G. A. P. and Orkin, S. H. (1992). Human transcription factor GATA-2; evidence for regulation of preproendothelin-1 gene expression in endothelial cells. *J. Biol. Chem.* **267**, 1279-1285.

Evans, T. and Felsenfeld, G. (1989). The erythroid-specific transcription factor Eryf1: A new finger protein. *Cell* **58**, 877-885.

Fridovich-Keil, J. L., Gudas, J. M., Bryan, I. B. and Pardee, A. B. (1991). Improved expression vectors for eukaryotic promoter/enhancer studies. *BioTechniques* **11**, 572-579.

Graham, F. L. and van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**, 456-467.

Hambor, J. E., Mennone, J., Coon, M. E., Hanke, J. H. and Kavathas, P. (1993). Identification and characterization of an Alu-containing, T-cell-specific enhancer located in the last intron of the human CD8a gene. *Mol. Cell. Biol.* **13**, 7056-7070.

Hannon, R., Evans, T., Felsenfeld, G. and Gould, H. (1991). Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA* **88**, 3004-3008.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359.

Hogan, B., Costantini, F. and Lacy, E. (1986). *Manipulating the Mouse Embryo*. Cold Spring Harbor: Cold Spring Harbor Press.

Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J. D. and Yamamoto, M. (1993). Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature* **362**, 466-469.

Joulin, V., Bories, D., Eleouet, J.-F., Labastie, M.-C., Chretien, S., Mattei, M.-G. and Romeo, P.-H. (1991). A T-cell specific TCR δ DNA binding protein is a member of the human GATA family. *EMBO J.* **10**, 1809-1816.

Kaufman, M. H. (1992). *The atlas of mouse development*. London: Academic Press.

Keller, G., Kennedy, M., Papayannopoulou, T. and Wiles, M. V. (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol. Cell. Biol.* **13**, 473-486.

Kelley, C., Blumberg, H., Zon, L. I. and Evans, T. (1993). GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. *Development* **118**, 817-827.

Ko, L. J. and Engel, J. D. (1993). DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.* **13**, 4011-4022.

Ko, L. J., Yamamoto, M., Leonard, M. W., George, K. M., Ting, P. and Engel, J. D. (1991). Murine and human T lymphocyte GATA-3 factors mediate transcription through a *cis*-regulatory element within the human T cell receptor δ gene enhancer. *Mol. Cell. Biol.* **11**, 2778-2784.

Kornhauser, J. M., Leonard, M. W., Yamamoto, M., LaVail, J. H., Mayo, K. E. and Engel, J. D. (1994). Temporal and spatial changes in GATA transcription factor expression are coincident with development of the chicken optic tectum. *Mol. Brain Res.* **23**, 100-110.

Landry, D. B., Engel, J. D. and Sen, R. (1993). Functional GATA-3 binding sites within murine CD8α upstream regulatory sequences. *J. Exp. Med.* **178**, 941-949.

Lavernere, A. C., MacNeill, C., Mueller, C., Poehmann, R., Burch, J. and Evans, T. (1994). GATA-4, -5, and -6 constitute a new subfamily of transcription factors in developing heart and gut. *Development,* submitted.

Leonard, M. W., Lim, K.-C. and Engel, J. D. (1993). Expression of the GATA transcription factor family during early erythroid development and differentiation. *Development* **119**, 519-531.

Leonard, M. W. and Patient, R. K. (1991). Primer extension analysis of mRNA isolated from transfected cell lines. In *Methods in Molecular Biology: Gene Transfer and Expression Protocols*, vol. **7**. (ed. E. J. Murray). pp 297-306. London: Humana Press.

Lindenbaum, M. H. and Grosveld, F. (1990). An in vitro globin gene switching model based on differentiated embryonic stem cells. *Genes Dev.* **4**, 2075-2085.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Martin, D. I. K. and Orkin, S. H. (1990). Transcriptional activation and DNA-binding by the erythroid factor GF-1/NF-E1/Eryf1. *Genes Dev.* **4**, 1886-1898.

Martin, D. I. K., Zon, L. I., Mutter, G. and Orkin, S. H. (1990). Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature* **344**, 444-447.

Merika, M. and Orkin, S. H. (1993). DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.* **13**, 3999-4010.

Mount, S. M. (1982). A catalog of splice junction sequences. *Nuc. Acids Res.* **10**, 459-472.

Nornes, D. and Carry, M. (1978). Neurogenesis in spinal cord of mouse: an autoradiographic analysis. *Brain Res.* **159**, 1-16.

Omichinski, J. G., Trainor, C., Evans, T., Gronenborn, A. M., Clore, G. M. and Felsenfeld, G. (1993). A small single-'finger' peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. *Proc. Natl. Acad. Sci. USA* **90**, 1676-1680.

Penit, C. and Vasseur, F. (1989). Cell proliferation and differentiation in the fetal and early postnatal mouse thymus. *J. Immunol.* **142**, 3369-3377.

Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S.-F., D'Agati, V., Orkin, S. H. and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257-260.

Pons, G., O'Dea, R. F. and Mirkin, B. L. (1982). Biological characterization of the C1300 murine neuroblastoma: an in vivo neural crest tumor model. *Cancer Res.* **42**, 3719-3723.

Redondo, J. M., Hata, S., Brocklehurst, C. and Krangel, M. S. (1990). A T cell-specific transcriptional enhancer within the human T cell receptor δ locus. *Science* **247**, 1225-1229.

Redondo, J. M., Pfohl, J. L. and Krangel, M. S. (1991). Identification of an essential site for transcriptional activation within the human T-cell receptor δ enhancer. *Mol. Cell. Biol.* **11**, 5671-5680.

Robey, E. A., Fowlkes, B. J., Gordon, J. W., Kioussis, D., von Boehmer, H., Ramsdell, F. and Axel, R. (1991). Thymic selection in CD8 transgenic mice supports an instructive model for commitment to a CD4 or CD8 lineage. *Cell* **64**, 99-107.

Romeo, P.-H., Prandini, M.-H., Joulin, V., Mignotte, V., Prenant, M., Vainchenker, W., Marguerie, G. and Uzan, G. (1990). Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature* **344**, 447-449.

Rosenthal, N. (1987). Identification of regulatory elements of cloned genes with functional assays. *Met. Enz.* **152**, 704.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Schwartzbauer, G., Schlesinger, K. and Evans, T. (1992). Interaction of the erythroid transcription factor cGATA-1 with a critical autoregulatory element. *Nuc. Acids Res.* **20**, 4429-4436.

Simon, M. C., Pevny, L., Wiles, M. V., Keller, G., Costantini, F. and Orkin, S. H. (1992). Rescue of erythroid development in gene targeted GATA-1-mouse embryonic stem cells. *Nature Gen.* **1**, 92-98.

Tamura, S., Wang, X.-H., Maeda, M. and Futai, M. (1993). Gastric DNA-

binding proteins recognize upstream sequence motifs of parietal cell-specific genes. *Proc. Nat. Acad. Sci. USA* **90**, 10876-10880.

Trainor, C. D., Evans, T., Felsenfeld, G. and Boguski, M. S. (1990). Structure and evolution of a human erythroid transcription factor. *Nature* **343**, 92-96.

Tsai, S.-F., Martin, D. I. K., Zon, L. I., D'Andrea, A. D., Wong, G. G. and Orkin, S. H. (1989). Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339**, 446-451.

Tsai, S.-F., Strauss, E. and Orkin, S. H. (1991). Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes Dev.* **5**, 919-931.

Whitelaw, E., Tsai, S.-F., Hogben, P. and Orkin, S. H. (1990). Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. *Mol. Cell. Biol.* **10**, 6596-6606.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W. J., Stott, D. and Allemann, R. K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes Dev.* **5**, 2048-2059.

Whyatt, D. J., deBoer, E. and Grosveld, F. (1993). The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *EMBO J.* **12**, 4993-5005.

Wilkinson, D. G., Bailes, J. A., Champion, J. E. and McMahon, A. P. (1987). A molecular analysis of mouse development from 8 to 10 dpc detects changes only in embryonic globin expression. *Development* **99**, 493-500.

Wilson, D. B., Dorfman, D. M. and Orkin, S. H. (1990). A non-erythroid GATA-binding protein is required for function of the human preproendothelin-1 promoter in endothelial cells. *Mol. Cell. Biol.* **10**, 4854-4862.

Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H. and Engel, J. D. (1990). Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev.* **4**, 1650-1662.

Yang, H.-Y. and Evans, T. (1992). Distinct roles for the two cGATA-1 finger domains. *Mol. Cell. Biol.* **12**, 4562-4570.

Yang, Z. and Engel, J. D. (1993). T lymphocyte transcription factor hGATA-3 regulates HIV-1 expression. *Nucl. Acids Res.* **21**, 2831-2836.

Yang, Z., Gu, L., Romeo, P.-H., Bories, D., Motohashi, H., Yamamoto, M. and Engel, J. D. (1994). Human GATA-3 trans-activation, DNA binding and nuclear localization activities are organized into distinct structural domains. *Mol. Cell. Biol.* **14**, 2201-2212.

Yomogida, K., Ohtani, H., Harigae, H., Ito, E., Nishimune, Y., Engel, J. D. and Yamamoto, M. (1994). Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse sertoli cells. *Development* **120**, 1759-1766.

Zon, L. I., Mather, C., Burgess, S., Bolce, M. E., Harland, R. M. and Orkin, S. H. (1991). Expression of GATA-binding proteins during embryonic development in Xenopus laevis. *Proc. Natl. Acad. Sci. USA* **88**, 10642-10646.

Zon, L. I., Tsai, S.-F., Burgess, S., Matsudaira, P., Bruns, G. A. P. and Orkin, S. H. (1990). The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. *Proc. Natl. Acad. Sci. USA* **87**, 668-672.