J. Biochem. **127**, 105–112 (2000)

Identification of Human *GATA-2* Gene Distal IS Exon and Its Expression in Hematopoietic Stem Cell Fractions[1]

Xiaoqing Pan,* Naoko Minegishi,* Hideo Harigae,‡ Hironori Yamagiwa,* Masayoshi Minegishi,‡ Yasuyuki Akine,† and Masayuki Yamamoto*,[2]

**Institutes of Basic Medical Sciences and Center for Tsukuba Advanced Research Alliance (TARA) and †Institute of Clinical Medicine, University of Tsukuba, Tsukuba 305–8577; ‡Department of Clinical Chemistry, School of Medicine, and §Institute of Aging, Cancer and Developmental Biology, Tohoku University, Sendai 980–7700*

Received September 14, 1999; accepted October 18, 1999

Transcription factor GATA-2 is essential for the proper function of hematopoietic stem cells and progenitors. Two first exons/promoters have been found in the mouse *GATA-2* gene, and a distal IS promoter shows activity specific to hematopoietic progenitors and neural tissues. To ascertain whether the two-promoter system is also utilized in the human *GATA-2* gene, we isolated and analyzed a P1 phage clone containing this gene. The nucleotide sequence of the human *GATA-2* gene 5' flanking region was determined over 10 kbp, and a human IS exon was identified in the locus through sequence comparison analysis with that of the mouse *GATA-2* IS exon. RNA blotting and reverse-transcribed PCR analyses identified a transcript that starts from the IS exon in human leukemia-derived cell lines. The IS-originated transcript was also identified in CD34-positive bone marrow and cord blood mononuclear cells, which are recognized as clinically important hematopoietic stem cell-enriched fractions. Phylogenic comparison of the human and mouse *GATA-2* gene sequences revealed several regions in the locus that exhibit high sequence similarity. These results demonstrate that the *GATA-2* gene regulatory machinery is conserved among vertebrates. The fact that the human IS promoter is active in the hematopoietic stem cell/progenitor fraction may be an important clue for the design of a vector system that can specifically express various genes in hematopoietic stem cells and progenitors.

Key words: GATA-2, hematopoietic stem cell, nucleotide sequence, promoter, transcriptional regulation.

Six members of the GATA family of transcription factors have been identified in vertebrates (*1–15*). Based on their sequence similarities and expression profiles, GATA factors are grouped into two subfamilies: GATA-1/-2/-3 and GATA-4/-5/-6. The former group of GATA factors is mainly expressed in hematopoietic lineage cells. For instance, GATA-1, the founding member of this family, is expressed in erythroid, megakaryocytic, eosinophilic, and mast cells (*1–4, 16–18*). GATA-3 is expressed mainly in T lymphocytes in hematopoietic lineages (*7, 19–20*). The expression of GATA-1 is critical for lineage selection and terminal maturation of erythroid and megakaryocytic cells (*21, 22*).

In contrast to GATA-1, GATA-2 is highly expressed in hematopoietic progenitors, erythroblasts, mast cells, and megakaryocytes (*23–30*). Targeted disruption of the *GATA-2* gene has demonstrated that, without GATA-2, mice die by embryonic day 11 (E11) due to severe disruption of primitive hematopoiesis (*31*). Without GATA-2, differentiation of definitive hematopoietic lineage was also affected in adult chimera mice (*31*). The lethal disruption of hematopoiesis was repaired by a transgenic integration of *GATA-2* yeast artificial chromosome clone into *GATA-2* gene knockout mouse (*30*). These results indicate the necessity of GATA-2 for the proper function of immature hematopoietic cells. In contrast, retroviral expression of GATA-2 in hematopoietic stem cells and progenitors blocked both their differentiation and amplification, suggesting that maintenance of *GATA-2* gene expression level within a particular range is critical for normal hematopoiesis (*32*).

Insights into *GATA-2* gene regulation have been obtained by analyzing the human (*25, 33*), *Xenopus* (*34*), mouse (*35, 36*), and chicken (*37*) *GATA-2* gene. The *GATA-2* gene consists of six exons. We previously found two alternative first exons/promoters of mouse *GATA-2* gene (*35*). The distal first exon (IS) is specific to the hematopoietic stem cell and progenitor fraction and neuron, whereas the proximal first exon (IG) is utilized widely in tissues where the *GATA-2* gene is generally expressed. Mouse IS promoter activity has been demonstrated in a transgenic reporter gene analysis, in which a 7.0-kbp region flanking 5′ to the mouse IS exon could induce reporter gene expression *in vivo* in murine hematopoietic progenitors in the aorta-

[1] This work was supported in part by a research grant from the Ichiro Kanehara Foundation; a grant from Rotary Yoneyama Memorial Foundation; Grants-in-Aid from the Ministry of Education, Science, Sports and Culture; CREST and the Japanese Society for Promotion of Sciences (JSPS Research for the Future Project).

[2] To whom correspondence should be addressed. Phone: +81-298-53-6158, Fax: +81-298-53-7318, E-mail: masi@tara.tsukuba.ac.jp

Abbreviations: G6PD, glucose 6-phosphate dehydrogenase; RT-PCR, reverse transcriptase–polymerase chain reaction.

gonad-mesonephros (AGM) region and para-aortic splanchnopleura (*36*).

Although we previously isolated human *GATA-2* genomic phage clones and determined the *GATA-2* gene structure, we were able to identify only one first exon in the human *GATA-2* gene (*25*), which corresponded to the proximal IG exon of the mouse *GATA-2* gene. Our attempt to isolate human IS exon from the genomic phage clones was unsuccessful. Thus, the presence of the IS exon and associated promoter in the human *GATA-2* gene remains to be elucidated. In this study, to examine whether the human *GATA-2* gene contains a distal first exon that corresponds to the mouse IS exon, we isolated a P1 phage clone containing the human *GATA-2* gene. We analyzed the human and mouse *GATA-2* gene structures, compared their sequences, and found an alternative distal exon in the human *GATA-2* gene. Importantly, the human IS promoter was found to be active in the hematopoietic stem cell and progenitor fraction of bone marrow and cord blood cells.

MATERIALS AND METHODS

Genomic Cloning and Sequence Comparison—PCR-based screening of the human genomic P1 phage library (Genome Systems) was performed. The following two primers were used: 5′-CTGGCGCACAACTACATGGAACC-3′, which spans nucleotides 365–387 of the human GATA-2 cDNA sequence; and 5′-GTGACTGCCTCTCGTACTTC-3′, which is complementary to nucleotides 927–946 of the cDNA sequence (*25*). A single P1 clone bearing the intact *GATA-2* gene was obtained. To determine size and orientation, P1 DNA recovered from this clone was digested with *Not*I and subjected to pulse-field gel electrophoresis with CHEF-DRIII (Bio-Rad). Southern blot analyses were carried out with gene-specific probes and with a probe for the vector. DNA restriction fragments derived from the P1 clone were subcloned into pBluescript II SK (Stratagene). Mouse *GATA-2* genomic DNA fragments in the previously isolated λ phage clones (*35*) were also subcloned to determine sequences. Serially deleted clones containing IS and IG 5′ flanking sequences were made. Two-directional DNA sequences were determined by cycle sequencing with sequence-specific primers or with M13 primers, using a PRISM™ 377 DNA Sequencer (Perkin-Elmer). Sequence comparison was carried out using two programs: DNASIS V 3.0 (Hitachi Software Engineering) and the PipMarker program (pipmaster@bio.cse.psu.edu).

Genomic Southern Blot Analysis—Human genomic DNA was isolated from the normal human leukocytes of two healthy volunteers. Genomic Southern blot analysis was carried out using human genomic DNA and a random-labeled probe as described previously (*35*).

RNase Protection Assay—An RNase protection assay was performed as described previously (*35*). To prepare the RNA probe, a human *GATA-2* genomic fragment was subcloned into pBluescript. A *Hin*d-III–*Not*I fragment (416 bp), including the 5′ flanking sequence and a part of the IS exon, was linearized with *Sty* I, and a labeled 326-bp antisense RNA was synthesized from the T3 promoter of the plasmid using T3 RNA polymerase (Promega). Total RNA (50 μg) from KCL-22 cells was hybridized to the labeled RNA transcript at 50°C for 12 h. Yeast tRNA (Sigma) was used as a control. The RNA samples were digested with an RNase mixture (Ambion), and the size of the hybrids was determined by denaturing polyacrylamide gel electrophoresis.

Cell Culture and RNA Blotting Analysis—The culture conditions of the cell lines used in this study were as described previously (*25, 38*). Total cellular RNA was extracted using RNA zol B (TEL-TEST). Total RNA (20 μg) was separated on a 1% formaldehyde-agarose gel and transferred onto a ZetaProbe GT membrane (Bio-Rad). A 229-bp *Apa*I–*Apa*I fragment containing IS exon was used as a probe for IS-originated mRNA. A 167-bp PCR-generated fragment encompassing the IG exons was used as a probe for IG-originated mRNA. A 900-bp *Bam*HI–*Eco*RI cDNA fragment covering exons 3–6 (*25*) was used as a probe for both mRNAs.

Fractionation of Bone Marrow Cells and Cord Blood Cells and RT-PCR Assay—Bone marrow cells and cord blood cells were isolated from normal healthy donors with informed consent. These cells were separated into CD34-negative and -positive populations using a magnetically activated cell sorting system, MACS (Miltenyi Biotee). RNA extraction and preparation of complementary DNA were performed as described previously (*35*). Both IS and IG cDNAs were amplified by PCR. The following primers were used: primer 1 (IS exon sense, 5′-AAGCGGTCCGCTGAACACC-3′), primer 2 (IG exon sense, 5′-CCGGACCGCGTGTTGATGTACCTTGG-3′), primer 3 (2nd exon antisense, 5′-AACTCCTTCACTCTCAGAGGCC-3′), and primer 4 (2nd exon antisense, 5′-CCACCTGCAGAAGAAGTTAGT-3′). Glucose 6-phosphate dehydrogenase (G6PD) expression was used as an internal standard (*39*).

RESULTS

P1 Phage Clone of Human GATA-2 Gene—Our previous attempt to identify a human counterpart of the mouse *GATA-2* gene IS exon promoter was unsuccessful (*35*). One possible reason for that failure is that the genomic clones were insufficient to include the IS promoter and exon. To overcome this, we isolated one P1 phage clone bearing the human *GATA-2* gene from a P1 phage library. The 85-kbp genomic clone was found to contain approximately 50 kbp of the 5′ flanking region relative to the first code of the translation (Fig. 1A). A series of Southern blot analyses revealed that this clone contains the same restriction enzyme sites as those detected in the human genomic DNA. Each identified fragment of P1 phage (data not shown) coincides with the genomic DNA (Fig. 1B). These results thus show that the P1 phage clone contains intact human *GATA-2* genomic DNA.

Identification of IS Exon in Human GATA-2 Gene—The P1 phage clone enabled us to determine the nucleotide sequences of the human *GATA-2* gene. To identify the human *GATA-2* gene IS exon, we sequenced an approximately 10-kbp genomic region from the IG exon to 5′ in the human *GATA-2* gene. Both strand sequences were determined in the region containing the 5′-flanking and IS exon. We found a segment in the human *GATA-2* gene which shows 61.1% similarity to the mouse IS exon sequence (Fig. 2). This segment was located at approximately 5 kbp 5′ to the previously identified human *GATA-2* first exon, which location is similar to that of the mouse IS exon (*35*).

We have also identified the human IG exon in the P1

phage clone and determined the human IG exon and promoter sequences. We found that the current sequence, which matches that of the overlapping λ phage library clones (*25*), does not match the previously reported sequence of the human *GATA-2* gene 5′ region (*33*). We propose that, in the latter case, an unrelated DNA fragment may have been inserted into the human *GATA-2* genomic phage clone.

Expression of IS Promoter in Human Leukemia-Derived Cell Lines—To determine whether the IS exon is utilized in the human *GATA-2* gene, we conducted a Northern blot hybridization analysis of GATA-2 mRNAs in human leukemia-derived cell lines. As shown in Fig. 3, we confirmed that human *GATA-2* mRNA was abundantly expressed in six human leukemia-derived cell lines using a cDNA fragment containing exons 3 to 6 as a probe. Showing very close agreement with our previous analysis (*25*), *GATA-2* mRNA was undetectable in lymphoid (THP-6, THP-9, Raji, and Jurkat) and monocytic (THP-1 and U-937) cell lines (data not shown). This expression profile was reproducible by an IG exon–specific probe. In contrast, when we used an IS exon–specific probe, specific bands were visible only in a YN-1 erythroleukemia cell line, a KCL-22 early myeloid cell line, and an M-TAT cytokine-dependent cell line (*38*)

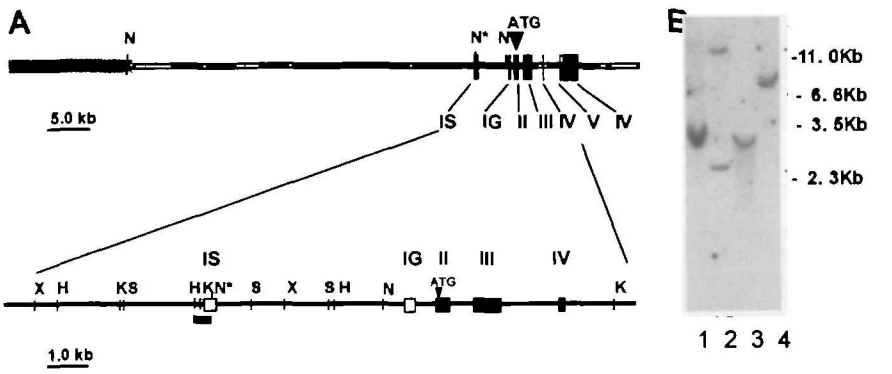

Fig. 1. **Structure of the human *GATA-2* gene.** A: structure and organization of the human GATA-2 P1 phage clone. The vector (cross-hatched), insert (open bar), and *Not*I sites (N) are shown. N* indicates a *Not*I site in the IS exon. Solid boxes represent exons. Abbreviations: H = *Hin*dIII; K = *Kpn*I; N = *Not*I; S = *Sac*I; X = *Xba*I. Open boxes in the lower panel indicate the two first exons. B: Genomic Southern blot analysis of the human *GATA-2* gene. Genomic DNA was digested with *Hin*dIII (lane 1), *Kpn*I (lane 2), *Sac*I (lane 3), and *Xba*I (lane 4), and hybridized with a probe (black bar underneath) specific for a 416-bp *Hin*dIII–*Not*I fragment encoding the 5′ flanking region and a part of the IS exon.

```
-170 CCTGAGTGCCTAGCGGCCCTTGGCCGGCCAGGCGGCTGGACCTGGTAGGCCGGGTGCAGA  h
     **** *      *    * ***** *** ********** ** ** * *** *****
-150 ACTGACTAGTGGCCCAAGCCTGGCCAGCCCGGCGGCTGGACCAGGCAGACAGGGCGCAGA  m

-110 GCCCCGGGTACCCTCCTG-CCCCCTGCGGCGCTCCCGCCCCCGTCGCCACCCCCTGCCCC  h
     * ****** ********* *********** ****       **             **
 -90 G-CCCGGGCACCCTCCTGCCCCCCTGCGGCGTTCCC-------TCC------------CC  m

 -51 CCCGCAAAGTGATGTCGAAATAAAAA---GCCGCGTTGCGGCCGGTGGGGGAGAGAGAGG  h
     *** * ******************   *****  *********          *****
 -50 CCCTCGAAGTGATGTCGAAATAAAAAGCTGCCGCTCTGCGGCCGGGAACC----GAGAGA  m

  +7 CAGCGTGAGCGCCAGGAAGGTAGCGAGGCCAGCGTCGCCCCGGGACTCGCTGCTCAAGTC  h
     ****  ** *   **  **** ** *********  *  * *   *** ******* ***
  +7 CAGCCAGAACC-AAGAGAGGTGGCCAGGCCAGCG--GGGCGGTAGCTCACTGCTCAGGTC  m

 +67 TGTCTATTGCCTGCCGCCACATCCATCCTAGCAGGGCCCCGTCGCCCACCAGGCGGACAA  h
       **  * ****      *** ** * * **          * **     *  ** ******
 +64 --TCCCTGGCCTAAGATCACCTCAACCATA----------GCCG--GTCTCGGGGGACAA  m

                    NotI
+127 AAGCGGTCCGCTGAACACCATGCGGCCGCTCGGCGTGCCGCCCAGGCTCTGCTGGTGAGC  h
     ******    *** *  *  **   ** **  ** *****   *** *  ******* *
+110 AAGCGGCTGCCTGCGCGACGAGCTTGCGTTC--CGCGCCGCGGGGGCCCCACTGGTGACC  m

+187 GCCGCCACCCCGCGCCCAGGTCCCGCGAGCCCGCCTGCCGCGCACCTCGCCCTGCT----  h
     ****     ******  *  ******* ***   ** *** ***   *** * * ***
+168 GCCG--CGGCCGCGCTGA-CTCCCGCGGGCCGACCAGCCACGCCGCTCACACCGCTGCCT  m

+243 CCCAGCTCTACTCCAGGCCCCGTCCGCCCGGGGGCGCCGCCCACCGCGCCTCGCTCG     h
     *** ***  *  ** ****  * ******* ******************* *  ***
+225 CCCTGCT-GAGCCCTGGCCGTGGCCGCCCGAGGGCGCCGCCCACCGCGCCGCAGTCG     m
```

Fig. 2. **Nucleotide sequence of the IS exon and promoter regions of human *GATA-2* gene.** Human (h) and mouse (m) IS promoter/exon sequences. Nucleotides are numbered at the left end of each row. The transcription initiation site is indicated by « and is numbered +1. The box encloses the IS exon. The *Not*I restriction site inside of the IS exon is underlined.

(Fig. 3), and the expression level varied among the cell lines. These results thus indicate that the IS promoter is indeed utilized for transcription of the human *GATA-2* gene.

Structure of Human IS Exon—We employed an RNase protection assay to determine the transcription start site of the human IS exon. Total RNA from KCL22 cells was used for this purpose, as the cell line contains IS-derived transcript abundantly (see above). The sizes of the protected products were estimated to be 148 and 141 nt relative to the *Not*I restriction site within the exon IS (Fig. 4). While these protected products were found in the KCL-22 cell RNA lane, the bands were not seen in the control yeast

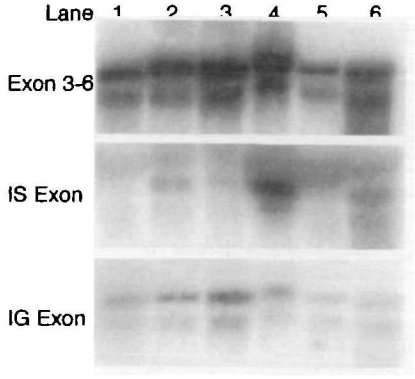

Fig. 3. **RNA blot analysis of human leukemia-derived cell lines.** Total cellular RNA (20 μg) from human leukemia-derived cell lines was loaded into formaldehyde-agarose gel. Cell lines used are KG-1a (lane 1), MTAT/GM-CSF (lane 2), MTAT/Epo (lane 3), KCL-22 (lane 4), Kyo-1 (lane 5), and YN-1 (lane 6). Specific probes for human IS and IG exons and for exons 3–6, respectively, were prepared and used as described in the text.

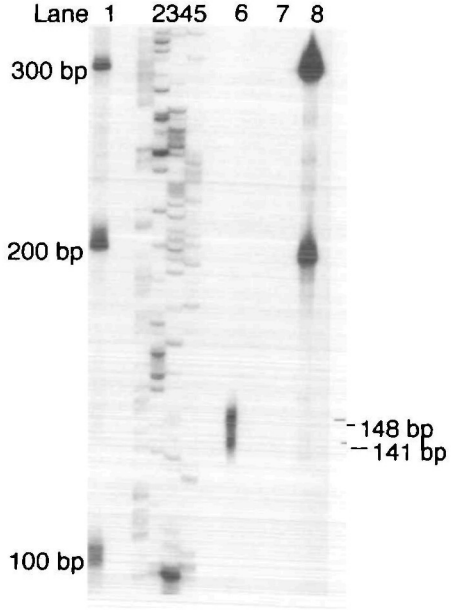

Fig. 4. **Transcription start site of the human GATA-2 IS promoter.** The transcription start site was determined by RNase protection assay. Total cellular RNA (50 μg) from KCL-22 leukemia cells (lane 6) and yeast tRNA (20 μg; lane 7) was hybridized with a 326-nucleotide RNA probe (lane 8 shows the probe alone). After treatment with RNase cocktail, protected products were separated. Positions of the protected products are depicted. Sizes of the bands were estimated by an RNA marker (lane 1) and by the DNA sequence ladder (lanes 2–5).

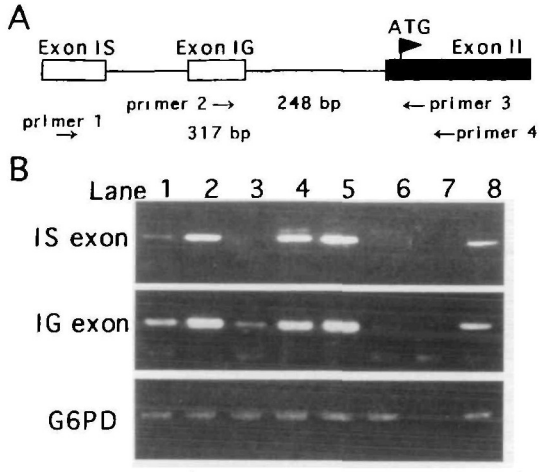

Fig. 5. **Expression of IS exon in human hematopoietic stem cells and progenitors.** A: Two primer sets for RT-PCR analysis were designed. For IS-derived mRNA, primers 1 and 4 (amplicon, 317 bp) were used, whereas primers 2 and 3 were used for IG-derived mRNA (amplicon, 248 bp). B: RT-PCR analysis of the GATA-2 gene expression in human hematopoietic stem cell and progenitor fractions. Bone marrow cells (lanes 1–3) and cord blood cells (lanes 4–6) from healthy donors were separated using magnetic beads. Three populations were examined by RT-PCR, unsorted (lanes 1 and 4), CD 34$^+$ (lanes 2 and 5), and CD 34$^-$ (lanes 3 and 6). Lane 7 is without RNA control, and lane 8 is a positive control using cDNA synthesized from total RNA of UT-7 leukemia cells. The top two panels show human IS and IG cDNA amplification, and the bottom panel shows G6PD amplification used to normalize RT-RNAs. Two representative experiments out of five performed with cells from different donors are shown.

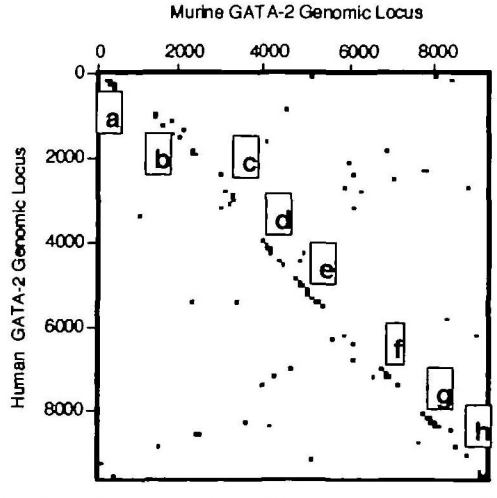

Fig. 6. **Phylogenic comparison of human and mouse *GATA-2* genes.** Mouse and human GATA-2 sequences were compared in a dot-matrix analysis using the DNASIS 3.0 program. The horizontal axis refers to the mouse sequence, and the vertical axis refers to the human sequence. Setting both check size and matching at 11 bp, eight regions of extensive homology were observed (a–h).

tRNA lane. The 3′ boundary of the IS exon was determined by comparison of the sequence of RT-PCR products with the genomic sequence (Fig. 2). The determined boundary conforms to the GT-AG rule (*40*). These results indicate that transcription of the human *GATA-2* gene is initiated primarily at 299 bp upstream relative to the 3' end of the IS exon (see arrow in Fig. 2). This assignment shows very close agreement with that for the murine *GATA-2* gene IS exon (*35*). The comparison of human and mouse IS promoter and IS exon sequences showed a high degree of sequence similarity (see Fig. 2).

IS-Derived GATA-2 mRNA Is Abundantly Expressed in Normal Human Hematopoietic Stem/Progenitor Cell Fractions—As the murine IS promoter was highly utilized in the c-Kit$^+$/Sca-1$^+$ hematopoietic stem/progenitor cell fraction, we next examined the transcriptional activity of the human IS promoter in the hematopoietic progenitor cell fraction. Bone marrow cells from healthy donors and cord blood cells were obtained with informed consent. Hematopoietic stem/progenitor cells were sorted with CD34 antibody and magnetic beads, according to the method developed for clinical hematopoietic stem cell transplantation (*41, 42*). Expression of GATA-2 mRNA in the total mononuclear cells and CD34-positive and CD34-negative cell populations were then examined by RT-PCR. Two sets of primers were used to detect IS- and IG-originated GATA-2 transcripts (Fig. 5A).

The IS exon and IG exon were both utilized much more actively in the CD34–positive cell fraction than in the CD34–negative cell fraction (Fig. 5B). Specificity of IS promoter usage was more restricted to the hematopoietic progenitor fraction than was that of the IG promoter. Indeed, we could even detect weak expression of the IG-derived GATA-2 mRNA in the CD34–negative cell fraction. This analysis was repeated with two more donors, and the results were found to be reproducible (data not shown). These results thus demonstrate that the IS promoter of the human *GATA-2* gene is highly active in hematopoietic stem/progenitor cells.

Sequence Similarity between Mouse and Human GATA-2 Genes—The results thus far show phylogenic conservation of the promoter structure of the *GATA-2* gene in vertebrates. This fact further suggests the conservation of gene regulatory sequences in the *GATA-2* locus. By identifying regions of high sequence similarity, we also determined the mouse *GATA-2* genomic DNA sequence over 10 kbp. The mouse sequence was compared to the human *GATA-2* genomic sequence. Both mouse and human *GATA-2* gene

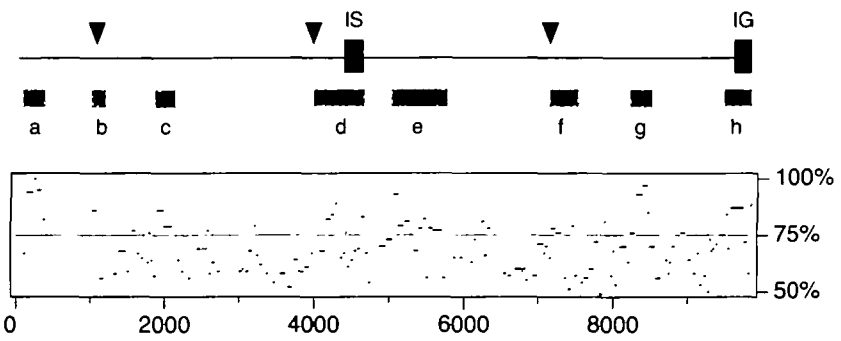

Fig. 7. **Percent identity plot of *GATA-2* gene regulatory region.** The human and mouse *GATA-2* genomic sequences were processed using the PipMarker program (pipmaster@bio.cse.psu.edu). Percent identity (vertical axis) was plotted against length in base pairs (horizontal axis), and percent identity in each gap-free aligning segment was plotted using the coordinates of the human sequence. The percent identity is only plotted between 50 and 100%, limiting the output to a range of moderately to strongly conserved sequences. Extensively homologous regions with percent identity over 75% are denoted a–h. The corresponding locations of these regions (hatched blocks a–h) are depicted in the map on the top lane. Arrowheads show the HS sites of the murine hematopoietic cell *GATA-2* locus (*36*).

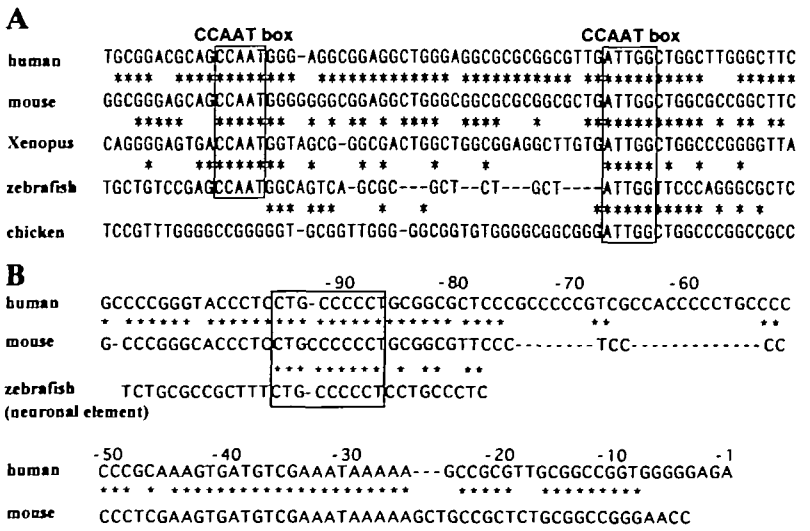

Fig. 8. **Conserved structure of vertebrate *GATA-2* gene promoters.** A: The human IG promoter sequence is compared with previously reported mouse (*35*), chicken (*37*) and *Xenopus* (*34*) *GATA-2* IG promoter sequences. Core CCAAT elements (boxed) in these promoters are well conserved. Identical sequences were also found in zebrafish *GATA-2* genomic sequence (*46*). B: The human IS promoter sequence is compared with mouse IS promoter sequence. A highly conserved sequence was found in the –90 bp region (boxed). This motif is also conserved in the zebrafish gene neuronal element.

sequences have been deposited in DDBJ/GenBank/EMBL databases.

We compared the mouse and human *GATA-2* genomic DNA sequences to search for DNA regions conserved between mouse and human. Stretches of about 10 kbp of the sequences were aligned, and similarity of the two sequences was analyzed by a Homology Plot program in DNASIS computer software. Figure 6 shows the results of these sequence comparison analyses. Eight regions with high sequence homology were identified between the two species and are depicted in the figure.

In an alternative approach, we employed a PipMarker program to compute alignments of similar regions in the two DNA sequences. This program is especially useful for generating pairwise alignments of long sequences (*43–45*). The resulting alignments are summarized as a percent identity plot in Fig. 7. Eight extensive regions were found in this analysis with percent identities larger than 75%, and these included both IS and IG promoter/exon regions. Of the remaining six, three were located upstream of the IS exon and the remaining three were between the IS and IG exons. The high homology regions show a fairly good coincidence with the previous matrix analysis (see the top panel of Fig. 7). In addition, most of these regions were mapped as DNase I hypersensitive sites (*36*). Thus some or all of the genomic DNA regions identified in this analysis are suggested to be important for *GATA-2* gene regulation.

Conserved Cis-Elements in the IS and IG Promoter—We also carried out phylogenic comparison of the IS and IG promoter sequences. As shown in Fig. 8A, two CCAAT boxes in the IG promoters were highly conserved among species. The sequences adjacent to these elements are also fairly well conserved, although the sequence of the chicken *GATA-2* gene is more divergent than those of other *GATA-2* genes (Fig. 8A). Similarly, we found that IS promoter sequences around –90 are conserved in mouse, human, and zebrafish (Fig. 8B). In the zebrafish, this region has been reported as a neuronal enhancer (*46*). Like the murine IS promoter, a TATA box–like AAATAAAAA element exists at locations –34 to –25 in the human IS promoter. These observations suggest that the transcriptional regulatory machinery of the two promoters of the *GATA-2* gene has been conserved during molecular evolution.

DISCUSSION

We identified in this study an alternative distal first exon and associated promoter of the human *GATA-2* gene that corresponds to the IS exon/promoter in the mouse *GATA-2* gene. The human IS promoter is active in human hematopoietic progenitor fractions prepared from bone marrow and cord blood cells. Extensive structural analysis of the human and mouse *GATA-2* gene revealed that the organization of the *GATA-2* gene is conserved between the two species. Particularly, we found eight regions in this locus, which show high sequence similarity. As three of these overlap with the DNase I HS sites, important *GATA-2* gene regulatory activity may reside in these regions.

RNA blot analyses with RNA samples from mouse and chicken culture cell lines suggested that IS promoter activity is weaker than that of the IG promoter (*35–37*). This was reproducible in the human system (see Fig. 3). It should be noted, however, that both IS and IG exon–originated GATA-2 cDNA clones were equally isolated in the initial λ phage library screenings. For instance, clones derived from chicken reticulocyte (*2*) and mouse fetal liver (*25*) cDNA libraries both originated from the distal IS exon. The two libraries represent hematopoietic tissues. In contrast, clones isolated from two non-hematopoietic tissues, *i.e.*, human umbilical vein endothelial cell (*5*) and stage 17 *Xenopus* embryo (*47*) cDNA libraries, originated from the proximal IG exon. These results thus imply that the IS promoter is actually utilized in hematopoietic cells. Since IS promoter activity is strong in the hematopoietic stem cell/progenitor fraction, one plausible possibility is that the IS promoter may be a participant in some signaling cascades, although the specific contribution of IS promoter activity to hematopoietic cell function remains to be clarified.

Reporter gene analysis in a transgenic mouse system *in vivo* has become an authentic way to study the transcriptional activity of vertebrate genes. Although the technique is time-consuming, the results reflect more physiological aspects than does the cell transfection assay. In support of the *in vivo* transgenic mouse assay, we believe that two approaches were important: DNase I footprinting analysis and comparative sequencing with genes of multiple species. Both approaches allow searching for candidate transcriptional regulatory regions. Indeed, we found eight genomic regions in the human *GATA-2* locus that show high sequence similarity with the mouse *GATA-2* gene. We also found three strong and two weak DNase I hypersensitive (HS) sites in the mouse hematopoietic cell *GATA-2* locus (*36*). The three strong HS sites are consistent with the highly homologous regions (b, d, and f). Importantly, a previous transgenic mouse reporter study and reporter transfection assay demonstrated the regulatory activity of the regions (*35, 36*).

Cell-surface marker CD34 is well known to be a hallmark of human hematopoietic stem cells/progenitors. This surface marker has been widely used in the isolation and characterization of hematopoietic stem cells from bone marrow and cord blood, as well as from circulating blood cells (*41, 48–50*). Previously, GATA-2 was found to be expressed in CD34-positive leukemic cell lines and primary myelogenous leukemia cells (*25, 51, 52*). We have shown in this study that, of the two GATA-2 transcripts, IS promoter–originated mRNA is restricted to the CD34-positive cell fraction of normal bone marrow and cord blood cells. In leukemic cell lines, IS mRNA expression seems to be limited to early erythroid, early myeloid, and cytokine-dependent hematopoietic progenitor cell lines, and is not present in pro-lymphoid cell lines. From this fact and the differing expression levels of IS-originated mRNA among cell lines, one may assume that the IS promoter is active in even more primitive cell fractions such as the $CD34^+/Lin^-$ subpopulation. In our recent studies of the mouse *GATA-2* gene, the gene was found to be expressed in the $Lin^-/c\text{-}kit^+/Sca\text{-}1^+$ fraction of adult bone marrow cells (*35*). In addition, we have demonstrated, using the transgenic mouse approach, that the expression of murine *GATA-2* gene in the early hematopoietic compartment of embryos is in part conducted by the IS promoter (*36*). All these data suggest an essential contribution of the IS promoter function to the growth and/or differentiation of human hematopoietic progenitors. One important approach for ascertaining the functional contribution of the *GATA-2* IS promoter is to ex-

amine the expression level of the IS promoter in human leukemia cells, and this is now underway in our laboratories.

Elucidation of the common regulatory mechanisms of vertebrate *GATA-2* gene expression is essential for our understanding of the transcriptional regulation operating during hematopoiesis. Full evaluation of the expression and role of the *GATA-2* gene in human hematopoiesis may enable us to control the purification of hematopoietic stem/progenitor cells, which would be of practical use in clinical medicine.

We thank Ms. Itsumi Tanaka, Ayako Taienncho, and Naóko Narita for help. We also thank Dr. Ed Coan for critical reading of the manuscript.

REFERENCES

1. Evans, T. and Felsenfeld, G. (1989) The erythroid-specific transcription factor Eryf1: a new finger. *Cell* **58**, 95–124
2. 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 (GATA) multigene family. *Genes Dev.* **4**, 1650–1662
3. Zon, L.I., Tsai, S., Burgess, S., Matsudaira, P., Burns, 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-chromatin. *Proc. Natl. Acad. Sci. USA* **87**, 668–672
4. Engel, J.D., George, K.M., Ko, L.J., Kornhauser, J.M., Leonard, M.W., Ting, P., and Yamamoto, M. (1991) Transcription factor regulation of hematopoietic cells. *Semin. Hematol.* **28**, 158–169
5. Lee, M.E., Temizer, D.H., Clifford, J.A., and Quertermous, T. (1991) Cloning of the GATA binding protein that regulates endothelin-1 gene expression in endothelial cells. *J. Biol. Chem.* **266**, 16188–16192
6. Ho, I.C., Vorhees, P., Marin, N., Oakley, B.K., Tsai, S.F., Orkin, S.H., and Leiden, J.M. (1991) Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. *EMBO J.* **10**, 1187–1192
7. Ko, L.J., Yamamoto, M., Leonard, M.W., George, K.M., Ting, P., and Engel, J.D. (1991) Murine and T-lymphocyte GATA-3 factors mediate transcription through a *cis*-regulatory element within the human T-cell δ gene enhancer. *Mol. Cell Biol.* **11**, 2778–2784
8. Dorfman, D.M., Wilson, D.B., Bruns, G.A.P., and Orkin, S.H. (1992) Human transcription factor GATA-2. *J. Biol. Chem.* **267**, 1279–1285
9. 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
10. Laveriere, A.C., MacNeill, C., Mueller, C., Poelmann, R.E., Burch, J.B., and Evans, T. (1994) GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J. Biol. Chem.* **269**, 23177–23184
11. Orkin, S.H. (1995) Transcription factors and hematopoietic development. *J. Biol. Chem.* **270**, 4955–4958
12. Soudais, C., Bielinska, M., Heikinheimo, M., MacArthur, C.A., Narita, N., Saffitz, J.E., Simon, M.C., Leiden, J.M., and Wilson, D.B. (1995) Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation *in vitro*. *Development* **121**, 3877–3888
13. Molkentin, J.D., Lin, Q., Duncan, S.A., and Olson, E.N. (1997) Requirement of the transcription factor GATA-4 for heart formation and ventral morphogenesis. *Genes Dev.* **11**, 1061–1072
14. Morrisey, E.E., Ip, H.S., Tang, Z., Lu, M.M., and Parmacek, M.S. (1997) GATA-5: a transcription activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev. Biol.* **183**, 21–36
15. Gao, X., Sedgwick, T., Shi, Y.B., and Evans, T. (1998) Distinct functions are implicated for the GATA-4, -5, -6 transcription factors in the regulation of intestine epithelial cell differentiation. *Mol. Cell. Biol.* **18**, 2901–2911
16. Martin, F., Prandini, M.H., Thevenon, D., Marguerie, G., and Uzan, G. (1993) The transcription factor GATA-1 regulates the promoter activity of the platelet glycoprotein IIb gene. *J. Biol. Chem.* **268**, 21606–21612
17. Zon, L., Yamaguchi, Y., Yee, K., Albee, E.A., Kimura, A., Bennett, J.C., Orkin, S.H., and Ackerman, S.J. (1993) Expression of mRNA for the GATA-binding proteins in human eosinophils and basophils: potential role in gene transcription. *Blood* **81**, 3234–3241
18. Harigae, H., Takahashi, S., Suwabe, N., Ohtsu, H., Gu, L., Yang, Z., Tsai, F.-Y., Kitamura, Y., Engel, J.D., and Yamamoto, M. (1998) Differential roles of GATA-1 and GATA-2 in growth and differentiation of mast cells. *Genes Cells* **3**, 39–50.
19. Ting, C.-N., Olson, M.C., Barton, K.P., and Leiden, J.M. (1996) Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* **384**, 474–478
20. Hendriks, R.W., Nawijn, M.C., Engel, J.D., van Doorninck, H., Grosveld, F., and Karis, A. (1999) Expression of the transcription factor GATA-3 is required for the development of the earliest T cell progenitors and correlates with stages of cellular proliferation in the thymus. *Eur. J. Immunol.* **29**, 1912–1918
21. Takahashi, S., Onodera, K., Motohashi, H., Suwabe, N., Hayashi, N., Yanai, N., Nabeshima, Y., and Yamamoto, M. (1997) Arrest in erythroid cell development by promoter-specific disruption of GATA-1 gene. *J. Biol. Chem.* **272**, 12611–12615
22. Suwabe, N., Takahashi, S., Nakano, T., and Yamamoto, M. (1998) GATA-1 regulates growth and differentiation of definitive erythroid lineage cells during *in vitro* ES cell differentiation. *Blood* **92**, 4108–4118
23. Leonard, M., Brice, M., Engel, J.D., and Papayannopoulou, T. (1993) Dynamics of GATA transcription factor expression during erythroid differentiation. *Blood* **82**, 1071–1079
24. Mouthon, M.A., Bernard, O., Mitjavila, M.T., Romeo, P.H., Vainchenker, W., and Mathieu, M.D. (1993) Expression of tal-1 and GATA-binding proteins during human hematopoiesis. *Blood* **81**, 647–655
25. Nagai, T., Harigae, H., Ishihara, H., Motohashi, H., Minegishi, N., Tsuchiya, S., Hayashi, N., Gu, L., Andres, B., Engel, J.D., and Yamamoto, M. (1994) Transcription factor GATA-2 is expressed in erythroid, early myeloid, CD34 human leukemia-derived cell lines. *Blood* **84**, 1074–1084
26. Orlic, D., Anderson, S., Biesecker, L.G., Sorrentino, B.P., and Bodine, D.M. (1995) Pluripotent hematopoietic stem cells contain high levels of mRNA for c-kit, GATA-2, p45NF-E2, and c-myb and low levels or no mRNA for c-fms and receptors for granulocyte colony-stimulating factor and interleukin 5 and 7. *Proc. Natl. Acad. Sci. USA* **92**, 4601–4605
27. Labbaye, C., Valtieri, M., Barberi, T., Meccia, E., Masella, B., Pelosi, E., Condorelli, G.L., Testa, U., and Peschle, C. (1995) Differential expression and functional role of GATA-2, NF-E2 and GATA-1 in normal adult. *J. Clin. Invest.* **95**, 2346–2358
28. Jippo, T., Mizuno, H., Xu, Z., Nomura, S., Yamamoto, M., and Kitamura, Y. (1996) Abundant expression of transcription factor GATA-2 in proliferating but not in differentiated mast cells in tissues of mice: demonstration by *in situ* hybridization. *Blood* **87**, 993–998
29. Tsai, F.-Y. and Orkin, S.H. (1997) Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **89**, 3636–3643
30. Zhou, Y., Lim, K.-C., Onodera, K., Takahashi, S., Ohta, J., Minegishi, N., Tsai, F.-Y., Orkin, S.H., Yamamoto, M., and Engel, J.D. (1998) Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development. *EMBO J.* **17**, 6689–6700
31. Tsai, F.-Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W., and Orkin, S.H. (1994) An early hematopoietic defect in mice lacking the transcription factor GATA-2.

Nature **371**, 221–226
32. Persons, D.A., Allay, J.A., Allay, E.R., Ashmun, R.A., Orlic, D., Jane, S.M., Cunningham, J.M., and Niehuis, A.W. (1999) Enforced expression of the GATA-2 transcription factor blocks normal hematopoietic stem cell function and fate. *Blood* **93**, 488–499
33. Fleenor, D.E., Langdon, S.D., deCastro, C.M., and Kaufman, R.E. (1996) Comparison of human and Xenopus GATA-2 promoters. *Gene* **179**, 219–223
34. Brewer, A.C., Guille, M.J., Fear, D.J., Partington, G.A., and Patient, R.K. (1995) Nuclear translocation of a maternal CCAAT factor at the start of gastrulation activates Xenopus GATA-2 transcription. *EMBO J.* **14**, 757–766
35. Minegishi, N., Ohta, J., Suwabe, N., Nakauchi, H., Ishihara, H., Hayashi, N., and Yamamoto, M. (1998) Alternative promoters regulate transcription of the mouse GATA-2 gene. *J. Biol. Chem.* **273**, 3625–3634
36. Minegishi, N., Ohta, J., Yamagiwa, H., Suzuki, N., Kawauchi, S., Zhou, Y., Takahashi, S., Hayashi, N., Engel, J.D., and Yamamoto, M. (1999) The mouse GATA-2 gene is expressed in the para-aortic splanchnopleura and aorta-gonads and mesonephros region. *Blood* **93**, 4196–4270
37. Nony, P., Hannon, R., Gould, H., and Felsenfeld, G. (1998) Alternative promoters and developmental modulation of expression of the chicken GATA-2 gene in hematopoietic progenitor cells. *J. Biol. Chem.* **273**, 32910–32919
38. Minegishi, N., Minegishi, M., Tsuchiya, S., Fujie, H., Nagai, T., Hayashi, N., Yamamoto, M., and Konno, T. (1994) Erythropoietin-dependent induction of hemoglobin synthesis in cytokine-dependent cell lines M-TAT. *J. Biol. Chem.* **269**, 27700–27704
39. Boehm, T., Spillantini, M.-G., Sofroniew, M.V., Surani, M.A., and Rabbitts, T.H. (1991) Developmentally regulated and tissue-specific expression of mRNA encoding the two alternative forms of the LIM domain oncogene rhombotin: evidence for thymus expression. *Oncogene* **6**, 695–703
40. Reed, R. and Maniatis, T. (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell* **46**, 681–690
41. de Wynter, E.A., Coutinho, L.H., Pei, X., Marsh, J.C., Hows, J., Luft, T., and Testa, N.G. (1995) Comparison of purity and enrichment of CD34 cells from bone marrow, umbilical cord and peripheral blood (primed for apheresis) using five separation systems. *Stem Cells* **13**, 524–532
42. Cairo, M.S. and Wagner, J.E. (1997) Placental and/or umbilical cord blood: an alternative source of hematopoietic stem cells for transplantation. *Blood* **90**, 4665–4678
43. Hardison, R. and Miller, W. (1993) Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol. Biol. Evol.* **10**, 73–102
44. Hardison, R., Chao, K.-M., Schwartz, S., Stojanovic, N., Ganetski, M., and Miller, W. (1994) Globin gene server: A prototype E-mail database server featuring extensive multiple alignments and data compilation. *Genomics* **21**, 344–353
45. Slightom, J.L., Bock, J.H., Tagle, D.A., Gumucio, D.L., Goodman, M., Stojanovic, N., Jackson, J., Miller, W., and Hardison, R. (1997) The complete sequences of the galago and rabbit β-globin locus control regions: Extended sequence and functional conservation outside the cores of DNase hypersensitive sites. *Genomics* **39**, 90–94
46. Meng, A., Tang, H., Ong, B.A., Farrell, M.J., and Lin, S. (1997) Promoter analysis in living zebrafish embryos identifies a *cis*-acting motif required for normal expression of GATA-2. *Proc. Natl. Acad. Sci. USA* **94**, 6267–6272
47. 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
48. Civin, C.I., Strauss, L.C., Brovall, C., Fackler, M.J., Schwartz, J.F., and Shaper, J.H. (1984) Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J. Immunol.* **133**, 157–165
49. Berenson, R.J., Andrews, R.G., Bensinger, W.I., Kalamasz, D., Knitter, G., Buckner, C.D., and Bernstein, I.D. (1988) Antigen CD34 marrow cells engraft lethally irradiated baboons. *J. Clin. Invest.* **81**, 951–955
50. Berenson, R.J., Bensinger, W.I., Hill, R.S., Andrews, R.G., Garcia-Lopez, J., Kalamasz, D.F., Still, B.J., Spitzer, G., Buckner, C.D., and Bernstein, I.D. (1991) Engraftment after infusion of $CD34^+$ marrow cells in patients with breast cancer or neuroblastoma. *Blood* **77**, 1717–1722
51. Shimamoto, T., Ohyashiki, J.H., Ohyashiki, K., Kimura, N., Nakazawa, S., and Toyama, K. (1994) GATA-1, GATA-2 and stem cell leukemia gene expression in acute myeloid leukemia. *Leukemia* **8**, 1176–1180
52. Minegishi, N., Morita, S., Minegishi, M., Tsuchiya, S., Konno, T., Hayashi, N., and Yamamoto, M. (1997) Expression of GATA transcription factors in myelogenous and lymphoblastic leukemia cells. *Int. J. Hematol.* **65**, 239–249