THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 31, Issue of August 2, pp. 27629–27635, 2002

Printed in U.S.A.

Variant Forms of α-Fetoprotein Transcripts Expressed in Human Hematopoietic Progenitors

IMPLICATIONS FOR THEIR DEVELOPMENTAL POTENTIAL TOWARDS ENDODERM*

Received for publication, March 4, 2002, and in revised form, May 6, 2002
Published, JBC Papers in Press, May 10, 2002, DOI 10.1074/jbc.M202117200

Hiroshi Kubota‡§, Robert W. Storms¶, and Lola M. Reid‡‖

From the ‡Department of Cell and Molecular Physiology and Program in Molecular Biology and Biotechnology, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, North Carolina 27599-7038 and ¶Medical Oncology and Transplantation, Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710-2630

Hematopoietic stem cells have been identified as multipotent cells that give rise to all adult hematopoietic lineages. Although the hematopoietic lineage is derived from the mesodermal germ layer in the embryo, recent data suggest that bone marrow cells with an antigenic profile consistent with that of hematopoietic stem cells can also differentiate to cell types of the endodermal lineages, such as hepatocytes. However, the molecular mechanisms associated with these events are entirely unknown. For decades, α-fetoprotein (AFP) has been used as a differentiation marker for endodermal cells, because it was thought that the transcription of AFP mRNA is tightly regulated in a developmental and tissue-specific process. In this report we describe two new variant forms of AFP transcripts in human hematopoietic progenitors that are not expressed in mature cells. The variant AFP (vAFP) cDNA sequences isolated from a multipotent hematopoietic cell line, K562, revealed that the vAFP differed from the authentic transcript, consisting of 15 exons, by replacing exon 1 of AFP with one or two exons located in the 5′-untranslated region of the *AFP* gene. In addition to the K562 cell line, vAFP transcripts were detected in normal bone marrow, thymus, and brain but were not detected in normal spleen, intestine, liver, or the hepatocellular carcinoma cell line, HepG2. This suggests expression in normal hematopoietic progenitors. This hypothesis was confirmed by the finding that $CD34^+Lin^-$ hematopoietic progenitor cells purified from cord blood by flow cytometric sorting also expressed the variant transcripts. These results suggest that some hematopoietic progenitors are in a state that permits them to express certain types of transcripts that have been considered unique to endoderm.

Stem cells, pluripotent cells capable of extensive growth without losing their multipotentiality, are becoming the centerpieces of new strategies for cell therapies, gene therapies, and tissue engineering, all contributing to a new era in medicine (1–3). The molecular machinery of self-renewal and differentiation is one of most intriguing subjects of stem cell biology. Analyses of gene expression patterns using subtracted cDNA library or cDNA microarray techniques have revealed several candidates of new or known transcripts that are expressed exclusively in stem cell populations, but not in terminally differentiated cells (4, 5). These studies provide enormous information facilitating the development of approaches for understanding stem cells. However, the applied techniques have an intrinsic weakness in that they cannot detect the stem/progenitor cell-specific isoforms of transcripts having relatively long domains in common between the isoforms. Therefore, an alternative strategy is to analyze individual genes with expression that is regulated in developmentally specific ways in particular cell lineages.

Cell differentiation in the developing embryo is regulated by extrinsic inductive signals and an intrinsic programmed genetic code. Differentiation into the three germ layers (ectoderm, endoderm, and mesoderm) from primitive ectoderm (epiblast) is a crucial step during development and thought to be an irreversible process. This leads to the development of unique cell types from each germ layer, *e.g.* epidermal or neuronal cells from ectoderm; epithelial cells in internal organs or digestive tract from endoderm; and hematopoietic and mesenchymal cells from mesoderm. Recent studies (6, 7) of cell transplantation, however, suggest that somatic stem or progenitor cells from adult tissues are able to generate cells with fates different from those normally associated with the germ layer that gave rise to the stem cells. A typical example of so-called "transdifferentiation" is that of murine bone marrow-derived $ckit^{high}Thy^{low}Lin^-Sca\text{-}1^+$ cells giving rise to mature hepatocytes (8). This antigenic profile is used to define hematopoietic stem cells. Although reports about these examples of possible transdifferentiation in mammals are accumulating (9), it remains to be determined whether the studies observed transdifferentiation or identified more primitive progenitors than hematopoietic stem cells in bone marrow. In addition, recent reports proposed that spontaneous cell fusion between transplanted donor progenitors and recipient cells might cause misinterpretation in previous studies reported as possible transdifferentiation (10, 11). In any event, biochemical analyses are necessary for understanding multipotency of the cells not only in those experimental setting but in normal physiological environment.

* This work was supported by Grant RO1-DK52851 from the National Institutes of Health and grants from Incara Cell Technologies (formerly Renaissance Cell Technologies), Research Triangle Park, NC. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF484525 and AF484526.

§ To whom correspondence may be addressed: Dept. of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, 3850 Baltimore Ave., Philadelphia, PA 19104-6009. Tel.: 215-898-8805; Fax: 215-898-0667; E-mail: kubota@vet.upenn.edu.

‖ To whom correspondence may be addressed: Glaxo Bldg., Campus Box 7038, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC 27599-7038. Tel.: 919-966-0346; Fax: 919-966-6112; E-mail: stemcell@med.unc.edu.

Adult liver parenchymal cells consist of hepatocytes and biliary epithelial cells. They are derived from common precursors, hepatoblasts, during embryonic development (12). The bipotent progenitors originate from multipotent endodermal cells in the foregut by inductive signals from the septum transversum surrounding the outgrowth of the endoderm (13). Although it is not known whether the bone marrow-derived cells can differentiate into hepatoblasts or can convert directly to hepatocytes, the cells possessing potential for hepatic differentiation should express some early endodermal markers before full differentiation into hepatocytes.

One such marker is α-fetoprotein (AFP),[1] which is a major serum protein produced primarily by visceral endoderm of the yolk sac and by hepatoblasts as well as more differentiated fetal hepatic cells (14, 15). AFP is one of the earliest markers for endodermal differentiation; the transcriptional expression starts with the appearance of visceral endoderm in the early embryo and is regulated tightly by developmental and tissue-specific cues (14–16). Therefore, in most studies that assess endodermal differentiation of human embryonic stem cells or embryonic germ cells, the expression of AFP transcripts or the products have been used as a marker of endoderm (17–19).

In this study, we have identified two variant forms of human AFP transcripts from a multipotent hematopoietic cell line, K562. The cDNA sequences revealed that the differences in the variant AFP (vAFP) mRNAs from that of the authentic transcript, which consists of 15 exons, is the presence of one or two unique exons, named exon A and exon B, replacing exon 1 of *AFP*. The variant forms were detected in normal $CD34^+Lin^-$ hematopoietic progenitor cells but not in mature blood cells. The expression of the variant AFP transcripts suggests that hematopoietic progenitors are in a state that permits them to express certain types of transcripts that have been considered unique to the endoderm.

EXPERIMENTAL PROCEDURES

Cell Culture—The human hepatoblastoma cell line, HepG2, was maintained in Eagle's MEM supplemented with 1 mM sodium pyruvate, 2 mM L-glutamine, 50 units/ml penicillin, 50 μg/ml streptomycin, 0.1 mM MEM non-essential amino acids solution, 5 μg/ml insulin, and 10% fetal bovine serum. The human erythroleukemia cell line, K562 was maintained in DMEM/F12 supplemented with 2 mM L-glutamine, 50 units/ml penicillin, 50 μg/ml streptomycin, 5×10^{-5} M 2-mercaptoethanol and 10% fetal bovine serum. HepG2 and K562 were obtained from the Tissue Culture Core Facility of the Cancer Center at the University of North Carolina at Chapel Hill. Normal human fetal lung fibroblasts, MRC5 (American Type Culture Collection), were cultured in the same supplemented DMEM/F12 in the presence of 10% horse serum (Sigma) and 1 ng/ml of bovine FGF-2 (Collaborative Biomedical Research).

Polymerase Chain Reaction—The primer sequences used for PCR are shown in Table I. For reverse transcriptase-PCR (RT-PCR), total RNAs were extracted from various cell lines and subjected to cDNA synthesis. Briefly, cDNAs were synthesized from 10 μg of total RNAs by oligo(dT) priming and AMV reverse transcriptase (Seikagaku) in a reaction volume of 20 μl at 42 °C. PCR was performed in a total volume of 50 μl consisting of 1 μM each primer, 200 μM each dNTP, 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM TrisHCl, pH 8.3, and 1.25 units of Amplitaq polymerase gold (PerkinElmer Life Sciences) with synthesized cDNA. Samples were heated to 94 °C for 3 min followed by amplification for 30 cycles for AFP or 35 cycles for vAFP of 2 min at 94 °C, 2 min 62 °C, and 3 min at 72 °C. After the last cycle, a final extension step was done at 72 °C for 6 min. Then 5 μl of each PCR reaction were analyzed by 1% agarose gel electrophoresis. For nested RT-PCR of AFP or vAFP in various human tissues, one-fourth of the cDNA synthesized from 10 μg of total RNAs (CLONTECH) were used for the first PCR. One-hundredth of the volume of the first PCR product was applied for nested PCR using an internal primer set. The PCR consisted of 1 cycle of 3 min at 95 °C; 30 cycles of 1.7 min at 95 °C, 2 min at 62 °C, 1.5 min at 72 °C followed by a 4.5 min extension at 72 °C. Then, 5 μl of each sample were analyzed in 2% agarose gels. For RT-PCR of sorted cells from human umbilical cord blood cells, total RNA was isolated with RNeasy kit (Qiagen). cDNAs synthesized from total RNAs of sorted cells were normalized by the cell number. Nested PCR for vAFP was performed with cDNA equivalent to 48,000 cells. Single step RT-PCR for β-actin or CD34 was done with cDNA equivalent to that of 1000 cells. For analysis of the genomic sequence of exon A, genomic DNAs were isolated from HepG2, K562, and MRC5 with QIAamp Tissue Kit (Qiagen). Amplified PCR products, using g-ex-AS and g-ex-AA primers (Table I), were cloned into pCRII-TOPO (Invitrogen), and the DNA sequences were determined by the DNA sequencing core facility of the University of North Carolina at Chapel Hill (Dr. Laura Livingston, director).

TABLE I
Primer sequences used in this study

RT-PCR	
ex-1S	5'-ACCATGAAGTGGGTGGAATC-3'
ex-2S	5'-CTTCCATATTGGATTCTTACCAATG-3'
ex-3S	5'-GGCTACCATATTTTTTGCCCAG-3'
ex-4S	5'-CTACCTGCCTTTCTGGAAGAAC-3'
ex-5S	5'-GAGATAGCAAGAAGGCATCCC-3'
ex-6S	5'-AAAGAATTAAGAGAAAGCAGCTTG-3'
ex-12S	5'-TAAACCCTGGTGTTGGCCAG-3'
ex-3A	5'-CCTGAAGACTGTTCATCTCC-3'
ex-14A	5'-ATTTAAACTCCCAAAGCAGCAC-3'
ex-AS	5'-AGAATTAAGGGACAGACTATGGGC-3'
ex-A1S	5'-GATGTCTGTCTCATAACACTTGGG-3'
ex-A2S	5'-TAAGCTTGGCAACTTGCAACAGGG-3'
ex-1capS	5'-ATATTGTGCTTCCACCACTGCC-3'
beta-actin S	5'-TGCAAGGCCGGCTTCGCGGGC-3'
beta-actin A	5'-TCCTTCTGCATCCTGTCGGCA-3'
CD34S	5'-TCATGAGTCTTGACAACAACGG-3'
CD34A	5'-CAGCCACCACGTGTTGTCTTGC-3'
Anchored PCR	
SN-poly(C)	5'-ACTAGTTAGCGGCCGCACTGGGC^{14}-3'
SN-primer	5'-ACTAGTTAGCGGCCGCACTGGG-3'
ex-1A	5'-CTCACCTATTCCATATTCATTTC-3'
ex-2A	5'-GGTCAGCTAAACTTATCTCTGC-3'
ex-3A2	5'-GGCTTCTTGAACAAACTGGGC-3'
ex-4A	5'-GCTGCAGCAGTCTGAATGTCC-3'
Genomic cloning of exonA	
g-ex-AS	5'-ATTCTGTTTTCACCCCATAGGTG-3'
g-ex-AA	5'-TTTCTCAGATATTCAAGCCCCAG-3'

Identification of the 5′-End of Variant AFP Transcripts—The rapid amplification of cDNA ends (RACE) method (14) was performed with modifications to identify the 5′-end of the vAFP transcripts. Briefly, with terminal transferase (Roche Molecular Biochemicals), a poly(dG) tail sequence was introduced into K562 cDNA derived from 10 μg of total RNA. One-tenth volume of the cDNA was amplified for the first 5 cycles with 94 °C for 1 min, 45 °C for 1.5 min, and 72 °C for 2.5 min, followed by 15 cycles with 94 °C for 1 min, 60 °C for 1.5 min, and 72 °C for 2.5 min using SN-poly(C) primer and ex-2A, ex-3A, or ex-4A. PCR products purified with QIAquick PCR purification kit (Qiagen) were reamplified with SN primer and an internal AFP primer (ex-1A, ex-2A, or ex-3A) at 94 °C for 1 min, 62 °C for 1.5 min, and 72 °C for 2.5 min for 30 cycles. The final PCR products were cloned into pCRII-TOPO and sequenced.

Umbilical Cord Blood (UCB) Cell Preparation and Fluorescence-activated Cell Sorting—Human UCB was obtained from the Carolina Cord Blood Bank following institutional review board-approved informed consent. UCB was processed as described (27). Unfractionated nucleated UCB cells were prepared by deleting red blood cells with the use of Hespan (DuPont Pharma, Wilmington, CE) and hemolysis at 37 °C in 0.17 M NH_4Cl, 20 mM Tris-HCl, pH 7.2, and 200 μM EDTA. Lineage-committed cells were removed from the unfractionated cells using the StemSep enrichment system (progenitor mixture, StemCell Technologies Inc, Vancouver, British Columbia). The recovered cells were termed Lin^- UCB cells. Cells were purified using a FACStar Plus cell sorter (Becton Dickinson Immunocytometry Systems (BDIS), San Jose, CA). For cell sorting, the Lin^- UCB cells were stained with a mixture of three phycoerythrin (PE) conjugated CD34-specific antibodies (clones Qbend10, Immu133 and Immu409, Beckman Coulter,

[1] The abbreviations used are: AFP, α-fetoprotein; vAFP, variant α-fetoprotein; RACE, rapid amplification of cDNA ends; UCB, umbilical cord blood; MEM, minimal essential medium; DMEM, Dulbecco's MEM; RT-PCR, reverse transcriptase-PCR; PE, phycoerythrin; 7AAD, 7-aminoactinomycin D.

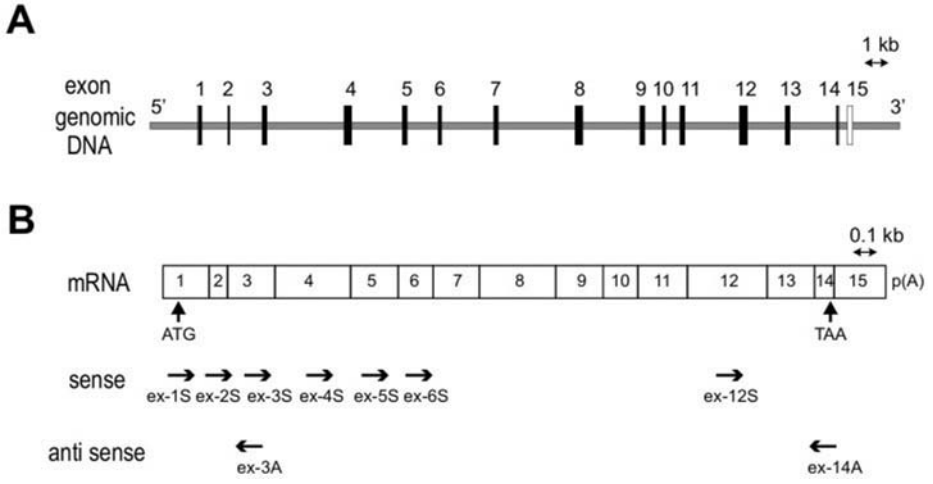

FIG. 1. **Schematic representation of the human *AFP* gene, AFP transcript, and position of primers used for RT-PCR.** *A*, human *AFP* gene consists of 15 exons and 14 introns and spans ~20,000 base pairs. The initiation methionine codon (ATG) and the termination codon (TAA) of the AFP transcript are indicated in exon 1 and exon 14, respectively. *B*, approximate position of sense primers and antisense primers are shown. The nucleotide sequences of primers were described in Table I.

Miami, FL). For all cell sorting, (7-aminoactinomycin D (Molecular Probes; 7AAD) was used to exclude dying cells. This was added to a final concentration of 10 μg/ml during the antibody incubations. In some instances, the expression of CD38 was also monitored (clone HB7, BDIS).

RESULTS

Variant AFP mRNA Expressed in K562, a Human Erythroleukemia Cell Line—The human *AFP* gene consists of 15 exons (Fig. 1*A*). The coding sequence is from exon 1 to exon 14 (20). To investigate the expression of possible variant forms of AFP mRNA in hematopoietic cells, two different portions of the AFP cDNA sequence were selected as target sequences of RT-PCR. The primer combination of ex-1S and ex-3A was used to amplify exon 1 containing the initiation methionine to exon 3, and the combination of ex-12S and ex-14A was used to amplify exon 12 to exon 14 containing the termination codon (Fig. 1*B*). The results of the PCR are shown in Fig. 2*A*. Both combinations of the primers resulted in intense amplification bands in the cDNA from HepG2, a hepatocellular carcinoma cell line (*lanes 1* and *2*). By contrast, cDNA from the erythroleukemia cell line, K562, showed an amplification band only for the C-terminal portion, identified by the ex-12S and ex-14A (*lane 5*) whereas there was no band by PCR of the primer combination of ex-1S and ex-3A (*lane 4*). This result suggests that K562 expresses a short form of authentic AFP transcript without the N terminus. In fact, the result of the PCR for the whole coding region of AFP using ex-1S and ex-14A primers showed that the single band of 1.8 kb (*lane 3*) was amplified from the HepG2 cDNA, whereas there was no detectable band in K562 (*lane 6*).

The *AFP* genes in rats, mice, and humans are organized similarly into 15 coding exons interrupted by 14 introns (20–22). Although there has been no report about any variant form of the AFP mRNA in humans, several short forms of the AFP transcript have been characterized in rats (23, 24). The rat vAFP lacks the first six or seven exons, but shares common 3′-sequences containing the seventh or eighth exon to the fifteenth exon, respectively (23). Detailed analysis of one variant form of rat showed that the rat vAFP used an additional exon, designated as exon V that is located in the seventh intron of the rat gene, as the first exon of the variant form. However, it turned out that human vAFP transcripts in K562 had different structures from the rat vAFP because RT-PCR of the primer combination exon 7 and ex-14A could detect the variant transcripts (data not show). Therefore, a series of 5′-primers from exon 2 to exon 6 were constructed (Fig. 1*B*) to see the difference between authentic and variant forms of AFP transcripts in K562. With each of these, ex-14A primer was used as a 3′-primer. The results of the RT-PCR showed, intriguingly, that

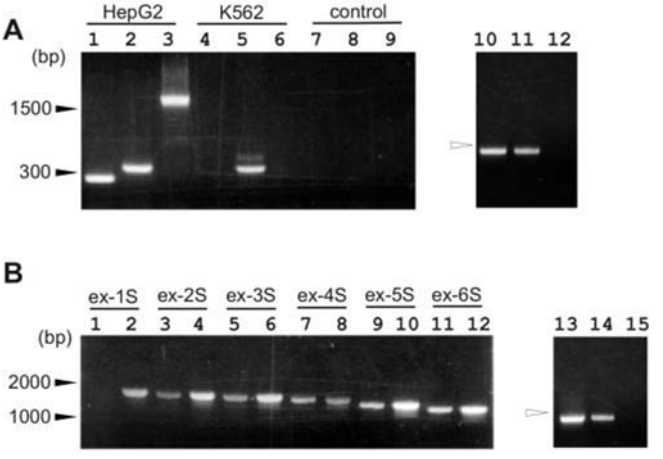

FIG. 2. **Expression of AFP mRNAs in K562 and HepG2 was analyzed by RT-PCR.** *A*, primer combination of ex-1S and ex-3A (*lanes 1*, *4*, and *7*), ex-12S and ex-14A (*lanes 2*, *5*, and *8*), or ex-1S and ex-14S (*lanes 3, 6,* and *9*) was used to amplify exon 1 to exon 3, exon 12 to exon 14, or exon 1 to exon 14 of AFP cDNA, respectively. *B*, series of 5′-primers from exon 1 to exon 6 (ex-1S to ex-6S) and ex-14A as the 3′-primer were used to detect AFP transcripts. *Lanes 1*, *3*, *5*, *7*, *9*, and *11* for K562. *Lanes 2*, *4*, *6*, *8*, 10, and *12* for HepG2. The amount of HepG2 cDNA was one-hundredth of that from K562. The *right panel* in *A* and *B* shows expression of β-actin. *A*, *lane 10*, HepG2; *lane 11*, K562; and *lane 12*, control. *B*, *lane 13*, HepG2; *lane 14*, K562; and *lane 15*, control. *Open triangle* indicates 1 kbp.

human variant AFP transcripts in hematopoietic cells share the entire coding exons, except for exon 1, in the authentic form that is expressed in HepG2 (Fig. 2*B*). Recently, it has been reported that an AFP isoform using a new exon located in intron 1 is expressed preferentially in the yolk sac and fetal liver in the mouse (25). However, no detectable amplified band in the PCR using an intron 1-specific 5′-primer (data not shown) suggested that the human vAFP transcripts in hematopoietic cells are novel isoforms of AFP.

Molecular Cloning of Variant AFP cDNA from K562—To determine the sequences of the 5′-end of the vAFP transcript in K562, the RACE method was applied as described under "Experimental Procedures." As a result, two types of variant transcripts were identified. The sequences are shown in Fig. 3*A*. Comparing the sequences of the variant transcripts to the GenBank™ data base, two regions of genomic sequence of the human *AFP* gene were identified. One type (type A) of the variant forms used an additional exon, designated as exon A, located about 5 kb upstream from exon 1 (Fig. 3*B*). The other type (type AB) used exon A in conjunction with another exon, designated as exon B, located about 1.6 kb upstream from exon

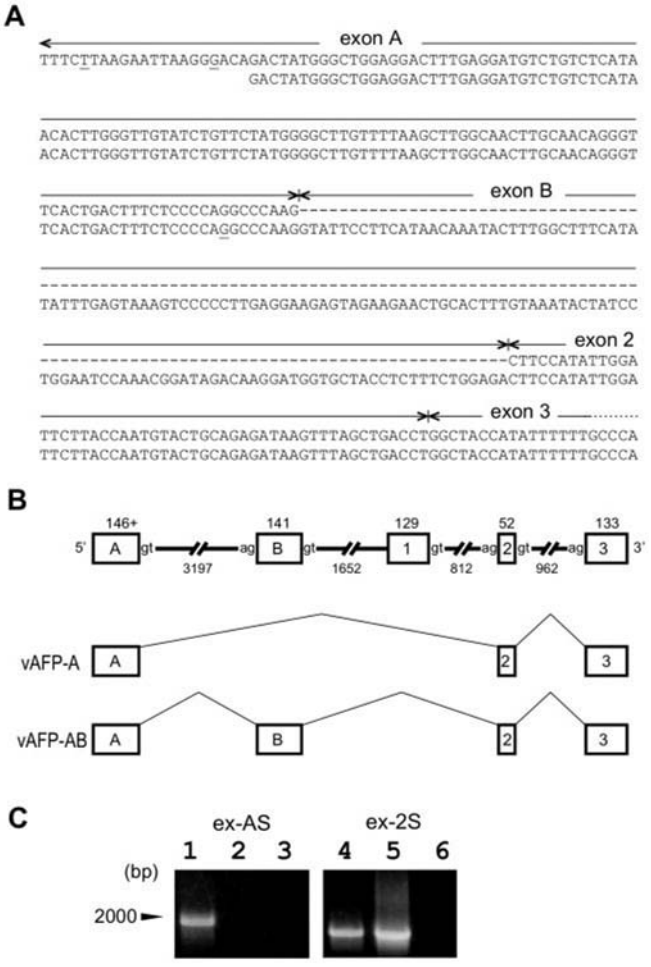

FIG. 3. **Molecular structure and expression of vAFP.** *A*, DNA sequences of two variant forms of AFP transcripts isolated from K562 cDNA. *B*, schematic illustration of genomic structure of variant AFP. *Open boxes* and *horizontal lines* indicate exons and introns, respectively. Their lengths in nucleotides are indicated by numbers *above* or *below* these elements. One type of the variant forms used one additional exon (*exon A*) located ~5-kb upstream from exon 1. The other type used exon A and another exon (exon B) located ~1.6 kb from exon 1. Genomic organization of the variant forms was deduced from genomic sequence of the *AFP* gene. *C*, selective expression of variant AFP transcripts in K562. RT-PCR of primer combination of exon A primer (ex-AS) and ex-14A primer were performed with cDNA from K562 (*lane 1*) and HepG2 (*lane 2*). K562, but not HepG2, expressed variant forms of AFP transcripts. The strong signal of HepG2 by RT-PCR using ex-2S and ex-14A indicates that HepG2 expresses the authentic form only. Equal amounts of cDNA were used for the reaction. *Lane 3*, control reaction.

1 (Fig. 3*B*). Among 19 clones analyzed in these experiments, 15 clones were type A; 3 clones were type AB; and one clone was the germ line transcript of intron 1. Although the genomic sequences containing exon A were reported originally from Watanabe *et al.* (26), some nucleotides in the exon A of K562 differed from the genomic sequence (Fig. 3*A*, *underline*). First, to determine whether this difference arose from artifacts associated with the anchored PCR technique, the genomic sequence of exon A in K562 cells was compared with the cloned cDNA sequence. This showed that the cDNA and genomic DNA of K562 cells were identical (data not shown). Next, to determine whether the genomic sequence of exon A of K562 is specific for the cell line, the genomic sequence of HepG2 and normal human fibroblasts, MRC5, was compared with that of K562. The results indicated no difference in the genomic sequence from all three sources, K562, HepG2, and MRC5 (data not shown). In addition, the genomic DNA sequence on the AFP locus from the data base of the human genome project (GenBank™; accession no. AP002825) corresponded with the sequences of K562. These results suggest that the DNA sequence of exon A cloned from K562 does not include artifacts from the PCR cloning.

Selective Expression of Variant Transcripts in K562—The expression pattern of the variant and authentic transcripts of AFP was studied in K562 and HepG2. A sense primer of exon A for RT-PCR was designed. The combination of the exon A primer, ex-AS, and exon 14 primer, ex-14A, successfully detected the transcripts in K562, but not HepG2 (Fig. 3*C*). The amplified products from K562 were cloned and sequenced. As expected, the two types of variant AFP transcripts either using both exons A and B or exon A alone were followed by exon 2 to exon 14. The number of clones with a type A sequence was 7 of 12, while that of type AB clone was 5 of 12 (data not shown). This suggests that the two types of vAFP are expressed equally in K562. The far stronger signal of HepG2 by RT-PCR using ex-2S and ex-14A indicated that HepG2 expressed the authentic form only. This result demonstrated clearly that the expression pattern of variant and authentic forms of AFP rigorously discriminates between K562 and HepG2 cells.

Expression of Variant Forms of AFP in Normal Human Tissues—We next examined whether the variant transcripts of AFP are expressed in normal cells. Since we anticipated the expression in normal cells would be at very low levels, a nested (two step) RT-PCR was performed. The first PCR was performed with primers, ex-A1S and ex-14A, to amplify the whole coding sequence of variant forms. Subsequently, nested PCR was performed by an internal primer combination, ex-A2S and ex-3A, which can distinguish between the presence of type A and type AB by the molecular size (Fig. 4*A*). The PCR results indicate that the tissue distribution of the vAFP transcripts is fairly restricted. Only primary hematopoietic organs, bone marrow (*lane 1*) and thymus (*lane 2*), and brain (*lane 7*) expressed the transcripts. Other tissues, including liver, lung, trachea, kidney, stomach, small intestine, colon, heart, and spleen, did not have cells expressing vAFP. These results strongly suggest that hematopoietic progenitors, but not mature blood cells, express vAFP, because the spleen is a secondary hematopoietic organ in which hemato-lymphopoiesis does not take place under normal physiological conditions. In addition, cDNA from peripheral blood cells (CLONTECH) never showed the expression of vAFP (data not shown). On the other hand, authentic AFP was detected in brain, kidney, small intestine, and thymus, but not in bone marrow. No expression in normal human fetal fibroblasts (*lane 13*) or small intestine (*lane 4*) suggests that the vAFP expression is not associated simply with cell proliferation.

Variant Transcripts of AFP Expressed in Hematopoietic Progenitors—The RT-PCR of tissue RNAs cannot define which cells express the variant forms of AFP transcripts. Bone marrow and thymus consist of a very heterogeneous cell population. Therefore, we studied purified hematopoietic progenitors from UCB by flow cytometric sorting to learn whether or not they express vAFP. Unfractionated nucleated cells in UCB contain only small numbers of $CD34^+$ cells (Fig. 5*A*) in which hematopoietic progenitors are highly enriched (27). The $CD34^+$ UCB cells could be enriched after removing Lin-positive cells (Fig. 5*A*). Both live ($7AAD^-$), unfractionated cells and $CD34^+Lin^-$ cells were isolated by flow cytometric sorting, and the RNAs extracted were subjected to nested PCR for vAFP. A total of four different cord bloods were tested. As shown in Fig. 5*B*, the expression of vAFP transcripts was detected in $CD34^+Lin^-$ cells, enriched hematopoietic progenitors (*lanes 2, 4, 6*, and *8*; 4 of 4), but not in unfractionated cord blood cells

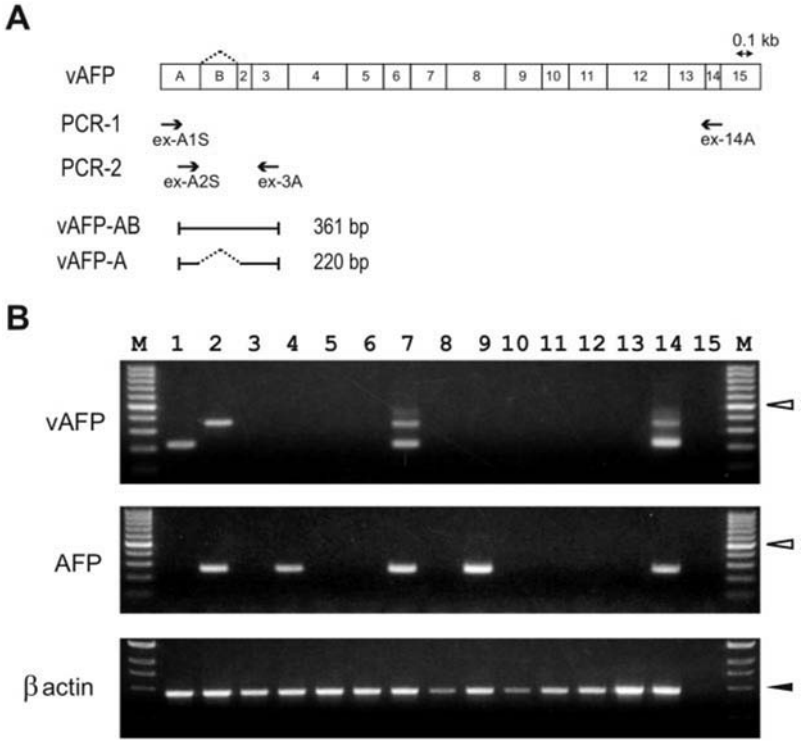

FIG. 4. **RT-PCR analyses of variant AFP transcripts in normal human tissues.** *A*, nested PCR strategy. First PCR (PCR-1) of primer combination of ex-A1S/ex-14A and second PCR (PCR-2) of that of ex-A2/ex-3A are illustrated. The amplified DNA fragments of two AFP variant forms were distinguishable in agarose electrophoresis. *B*, expression of variant AFP mRNAs was analyzed in cDNAs from various human tissues. *Lane 1*, bone marrow; *lane 2*, thymus; *lane 3*, spleen; *lane 4*, small intestine; *lane 5*, colon; *lane 6*, stomach; *lane 7*, brain; *lane 8*, heart; *lane 9*, kidney; *lane 10*, liver; *lane 11*, lung; *lane 12*, trachea; *lane 13*, MRC5; *lane 14*, K562; and *lane 15*, no template. Nested RT-PCR of the authentic AFP transcript was performed with the same human tissues cDNAs. Primer combination of ex-1capS/ex-14A for first PCR and that of ex-1S/ex-3A for second PCR were used. The result from single step RT-PCR of β-actin is shown at the *bottom*. *Open* and *closed triangles* indicate 500 or 1000 bp of the ladder marker, respectively.

(*lanes 1*, *3*, and *5*; 3 of 4). Cloning and DNA sequencing experiments showed that the amplified bands corresponded to cDNA sequences of type A and type AB from K562 cells (data not shown). CD34 expression in the sorted $CD34^+Lin^-$ cells was confirmed by RT-PCR (Fig. 5*B*).

DISCUSSION

The present study identified two variant forms of AFP expressed preferentially in hematopoietic progenitors, but not in mature blood cells or in hepatic cells. These results suggest that the AFP gene locus of chromatin in hematopoietic progenitors is open and accessible to transcription factors for the mRNA expression. In other words, chromatin-related repression of authentic AFP, which is a mechanism that will block inappropriate expression of authentic AFP in non-hepatic cells (28), is incomplete in hematopoietic progenitors and allows vAFP to be transcribed. In the case of K562, strictly speaking, very small amounts of authentic AFP transcripts could be detected when cycles of RT-PCR were increased (Fig. 4*B*, *lane 14*). This indicates that K562 express vAFP at a much higher level than authentic AFP. Conversely, vAFP could be detected in HepG2 by increasing cycles of RT-PCR (data not shown). Therefore, our data derived from the cell lines suggest that three types of patterns are present with respect to the authentic and variant transcripts of AFP. Hepatic cells (HepG2) express authentic AFP dominantly, while hematopoietic cells (K562) express vAFP dominantly. Fibroblasts (MRC5) express neither authentic AFP nor vAFP. This prompts us to ask two questions with respect to AFP expression. 1) What is the mechanism of the opening of the AFP locus in the chromatin, giving rise to whichever of the transcripts is expressed? 2) What is the transcriptional machinery associated with the different transcripts?

The mouse has proved an excellent model system for studying tissue-specific and developmentally regulated transcriptional control of authentic AFP *in vivo* as well as *in vitro* (14–16). Extensive studies have established that the transcriptional control of the *AFP* gene is mediated by five cis-acting regulatory domains, including the *AFP* promoter, three distinct

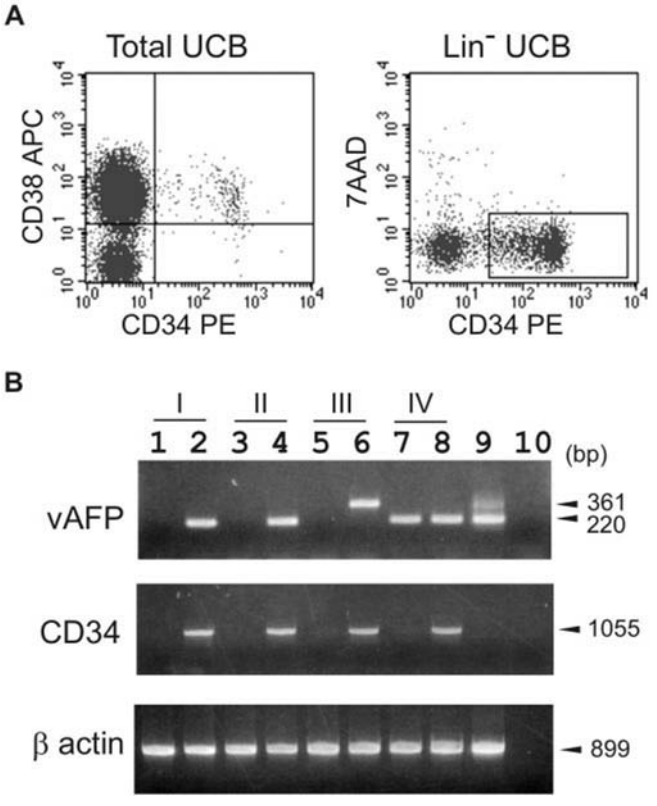

FIG. 5. **Expression of variant AFP transcripts in normal human hematopoietic prgenitors.** *A*, representative dot plot of flow cytometry for CD34 and CD38 expression in UCB cells after ficoll centrifugation are shown at the *left*. $CD34^+$ cell population was 2.3% on the average. $CD34^+$ cell population was enriched after removing Lin^+ cells. Live CD34 cell fraction (*inside of open box*) was sorted for RT-PCR analysis. *B*, RT-PCR analyses of variant AFP transcripts of sorted UCB were performed as described in the legend to Fig 3. *Roman numerals* indicate individual UCB samples. *Lanes 1*, *3*, *5*, and *7*, unfractionated UCB cells; *lanes 2*, *4*, *6*, and *8*, $CD34^+Lin^-$ UCB cells; *lane 9*, K562; *lane 10*, control reaction.

enhancer elements, and one repressor region located between the *AFP* promoter and the upstream enhancers (reviewed in Ref. 29). There are a number of transcription factors binding to the promotor region to control the expression of the authentic AFP form (reviewed in Ref. 29). It has been proposed that GATA factors, especially GATA4, could be a master gene for endodermal differentiation (30) and for the expression of fetoprotein transcription factor, which activates AFP expression in early liver development (31). On the other hand, other GATA factors are key regulators for hematopoiesis (32). Interestingly, several possible GATA family binding sites were identified in the exon A genomic sequence as well as other transcription factor binding sites associated with hematopoiesis such as AML-1 (33) and MZF-1 (34). Because exon A was identified as only one 5′-end of two variant forms by our RACE method in all independent clones, exon A is possibly the first exon of vAFP in K562. Thus, it is noteworthy to see whether these transcription factors expressed in hematopoietic progenitors are involved in the transcription of vAFP. However, the transcriptional start site of the variant forms of AFP in hematopoietic cells remains to be determined.

Although extensive research has been performed to elucidate *AFP* gene regulation in various species (15, 35), there has been no report about human variant AFP transcripts nor identification of exon A and exon B 5′-upstream of exon 1 in any species. Variant AFP transcripts identified in rodents were expressed preferentially in hepatic or yolk sac cells (24, 25). These suggest that the vAFP transcripts in the human hematopoietic cells are different from ones identified in rat or in mouse not only in their molecular structures, but also in their expression pattern. Interestingly, all variant forms of AFP reported in the previous studies and this study use unique first exons. Because the different usage of the first exon suggests that the vAFP expression is dependent on a unique promotor (24, 25), further studies are required to elucidate the hematopoietic cell-specific promotor of vAFP.

At present, apart from the hematopoietic organs, the brain is the only tissue where vAFP expression is observed. It has been shown that cultured neurospheres, which are considered as neural stem cells, isolated from the anterior portion of the lateral wall of the lateral ventricles in adult mice were able to differentiate to endodermal cells (36). Therefore, it is important to determine the identity of $vAFP^+$ cells in the brain and isolate them from brain tissue in order to ask whether they have a capability to form neurospheres *in vitro*. The expression of vAFP in non-endodermal cells may represent a unique immature state, which has the developmental potential of endodermal cells.

The function of the vAFP is unclear because the amount of mRNA was extremely low, and no protein products were ever found in assays using immunohistochemistry with anti-AFP antibodies on hematopoietic progenitors.[2] In addition, there is no long open reading frame starts in exon A or B to connect with exon 2 of authentic AFP if only ATG is considered as an initiation codon. In that case, a possible initiation codon at exon 3 would be used for the translation so that the translated product from vAFP transcripts would be a truncated form. However, because there is TTG and CTG in exon B to connect the open reading frame of exon 2, it might be possible for one of them to work as an alternative initiation codon (37, 38). In any event, further studies are required to elucidate the possible functions of vAFP transcripts.

As a final point, this study raises a warning about the use of RT-PCR techniques for the detection of AFP mRNA. AFP is used commonly as a molecular marker for endodermal differentiation (17–19) or certain carcinomas such as hepatomas or germ cell tumors (15, 35). Various RT-PCR techniques have been developed for the detection of circulating carcinoma cells in peripheral blood (39). Nevertheless, none of them, to the best of our knowledge, used exon 1-specific primers to detect the authentic form of AFP. Thus, the use of not only the exon 1-specific primer but the exon A-specific primer for RT-PCR to analyze AFP mRNA expression will be necessary to consider which types of transcripts are expressed in the cells examined.

In conclusion, this is the first study in which variant forms of an endodermal marker, AFP, have been found, with a selective pattern of expression in human hematopoietic progenitors. Our results suggest that the expression of vAFP transcripts may reflect the ability of progenitors for hematopoietic cells to differentiate to endodermal cells and provides a unique opportunity for studying the mechanism.

Acknowledgments—We thank N. Moss for critical evaluation of the manuscript, M. Cook for operating the cell sorter, and C. Lodestro for laboratory management. We also thank members of the laboratory for helpful discussion and encouragement.

[2] N. Moss and H. Kubota, unpublished observations.

REFERENCES

1. Fuchs, E., and Segre, J. A. (2000) *Cell* **100,** 143–155
2. Xu, A. S. L., Luntz, T. L., Macdonald, J. M., Kubota, H., Hsu, E., London, R. E., and Reid, L. M. (2000) in *Principles of Tissue Engineering* (Lanza, R. P., Langer, R., and Chick, W. L., eds), pp. 559–598, Academic Press, San Diego, CA
3. Lagasse, E., Shizuru, J. A., Uchida, N., Tsukamoto, A., and Weissman, I. L. (2001) *Immunity* **14,** 425–436
4. Phillips, R. L., Ernst, R. E., Brunk, B., Ivanova, N., Mahan, M. A., Deanehan, J. K., Moore, K. A., Overton, G. C., and Lemischka, I. R. (2000) *Science* **288,** 1635–1640
5. Terskikh, A. V., Easterday, M. C., Li, L., Hood, L., Kornblum, H. I., Geschwind, D. H., and Weissman, I. L. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7934–7939
6. Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C., and Vescovi, A. L. (1999) *Science* **283,** 534–537
7. Krause, D. S., Theise, N. D., Collector, M. I., Henegariu, O., Hwang, S., Gardner, R., Neutzel, S., and Sharkis, S. J. (2001) *Cell* **105,** 369–377
8. Lagasse, E., Connors, H., Al-Dhalimy, M., Reitsma, M., Dohse, M., Osborne, L., Wang, X., Finegold, M., Weissman, I. L., and Grompe, M. (2000) *Nat. Med.* **6,** 1229–1234
9. Morrison, S. J. (2001) *Curr. Biol.* **11,** R7–9
10. Terada, N., Hamazaki, T., Oka, M., Hoki, M., Mastalerz, D. M., Nakano, Y., Meyer, E. M., Morel, L., Petersen, B. E., and Scott, E. W. (2002) *Nature* **416,** 542–545
11. Ying, Q. L., Nichols, J., Evans, E. P., and Smith, A. G. (2002) *Nature* **416,** 545–548
12. Kubota, H., and Reid, L. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 12132–12137
13. Zaret, K. (1998) *Curr. Opin. Genes Dev.* **8,** 526–531
14. Tilghman, S. M. (1985) *Oxf. Surv. Eukaryotic Genes* **2,** 160–206
15. Chen, H., Egan, J. O., and Chiu, J. F. (1997) *Crit. Rev. Eukaryotic Gene Expression* **7,** 11–41
16. Hammer, R. E., Krumlauf, R., Camper, S. A., Brinster, R. L., and Tilghman, S. M. (1987) *Science* **235,** 53–58
17. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998) *Science* **282,** 1145–1147
18. Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D. A., and Benvenisty, N. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 11307–11312
19. Shamblott, M. J., Axelman, J., Littlefield, J. W., Blumenthal, P. D., Huggins, G. R., Cui, Y., Cheng, L., and Gearhart, J. D. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 113–118
20. Gibbs, P. E., Zielinski, R., Boyd, C., and Dugaiczyk, A. (1987) *Biochemistry* **26,** 1332–1343
21. Eiferman, F. A., Young, P. R., Scott, R. W., and Tilghman, S. M. (1981) *Nature* **294,** 713–718
22. Nahon, J. L., Danan, J. L., Poiret, M., Tratner, I., Jose-Estanyol, M., and Sala-Trepat, J. M. (1987) *J. Biol. Chem.* **262,** 12479–12487
23. Lemire, J. M., and Fausto, N. (1991) *Cancer Res.* **51,** 4656–4664
24. Watanabe, T., Jimenez-Molina, J. L., and Chou, J. Y. (1992) *Biochem. Biophys. Res. Commun.* **185,** 648–656
25. Scohy, S., Gabant, P., Szpirer, C., and Szpirer, J. (2000) *Nucleic Acids Res.* **28,** 3743–3751
26. Watanabe, K., Saito, A., and Tamaoki, T. (1987) *J. Biol. Chem.* **262,** 4812–4818
27. Storms, R. W., Goodell, M. A., Fisher, A., Mulligan, R. C., and Smith, C. (2000) *Blood* **96,** 2125–2133
28. Crowe, A. J., Sang, L., Li, K. K., Lee, K. C., Spear, B. T., and Barton, M. C. (1999) *J. Biol. Chem.* **274,** 25113–25120
29. Spear, B. T. (1999) *Semin. Cancer Biol.* **9,** 109–116
30. Zaret, K. (1999) *Dev. Biol.* **209,** 1–10

31. Pare, J. F., Roy, S., Galarneau, L., and Belanger, L. (2001) *J. Biol. Chem.* **276,** 13136–13144
32. Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., and Orkin, S. H. (1994) *Nature* **371,** 221–226
33. Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G., and Downing, J. R. (1996) *Cell* **84,** 321–330
34. Morris, J. F., Rauscher, F. J., 3rd, Davis, B., Klemsz, M., Xu, D., Tenen, D., and Hromas, R. (1995) *Blood* **86,** 3640–3647
35. Lazarevich, N. L. (2000) *Biochemistry* (*Mosc.*) **65,** 117–133
36. Clarke, D. L., Johansson, C. B., Wilbertz, J., Veress, B., Nilsson, E., Karlstrom, H., Lendahl, U., and Frisen, J. (2000) *Science* **288,** 1660–1663
37. Hann, S. R., King, M. W., Bentley, D. L., Anderson, C. W., and Eisenman, R. N. (1988) *Cell* **52,** 185–195
38. Prats, H., Kaghad, M., Prats, A. C., Klagsbrun, M., Lelias, J. M., Liauzun, P., Chalon, P., Tauber, J. P., Amalric, F., Smith, J. A., *et al.* (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1836–1840
39. Paterlini-Brechot, P., Vona, G., and Brechot, C. (2000) *Semin. Cancer Biol.* **10,** 241–249