

MOLECULAR GENETICS

SZF1: A novel KRAB-zinc finger
gene expressed in CD34⁺ stem/progenitor cells

Cheng Liu^a, Mark Levenstein^a, Joseph Chen^a, Elina Tsifrina^a,
Raluc Yonescu^a, Constance Griffin^{a,b}, Curt I. Civin^{a,c}, and Donald Small^{a,c}

^aThe Johns Hopkins Oncology Center, and the Departments of

^bPathology and ^cPediatrics, The Johns Hopkins University School of Medicine, Baltimore, MD

(Received 12 February 1998; revised 3 August 1998; accepted 8 August 1998)

The identification and study of genes expressed in hematopoietic stem/progenitor cells should further our understanding of hematopoiesis. Transcription factors in particular are likely to play important roles in maintaining the set of genes that define the stem/progenitor cell. We report here the identification of a putative KRAB-zinc finger gene (SZF1) from a cDNA library prepared from human bone marrow CD34⁺ cells. Characterization of SZF1 implicates its role in hematopoiesis. The predicted protein contains a highly conserved KRAB domain at the NH₂ terminus and four zinc fingers of the C₂H₂ type at the COOH terminus. Two alternatively spliced products of SZF1 were isolated, which predict proteins of 421 (SZF1-1) and 361 (SZF1-2) amino acids, differing from each other only at the carboxy terminus. The two transcripts of SZF1 have different expression patterns. SZF1-2 is ubiquitously expressed, as indicated by Northern blot, RNase protection, and reverse transcriptase polymerase chain reaction. SZF1-1 expression, in contrast, was detected only in CD34⁺ cells. We recently isolated the promoter region for the stem/progenitor cell expressed FLT3/FLK-2/STK-1 gene and used this region to generate a reporter construct to test the effect of SZF1 expression. Cotransfection of the reporter construct with SZF1 constructs showed that SZF1-2 repressed transcription three- to fourfold, whereas SZF1-1 showed a lower level of repression. The expression pattern of SZF1 transcripts and the transcriptional repression of a CD34⁺-specific promoter demonstrate a possible role for SZF1 in hematopoietic stem/progenitor cell differentiation. © 1999 International Society for Experimental Hematology. Published by Elsevier Science Inc.

Keywords: Hematopoietic stem/progenitor cells—Zinc finger gene—KRAB domain—Transcriptional repression

Introduction

Transcription factors bind to sets of genes and stimulate or repress the transcription of these genes, thus mediating cell fate decisions and regulating differentiation. Hematopoietic development involves many transcription factors and the phenotypes of the differentiated lineages reflect the sets of genes activated or repressed by these factors [1,2]. The processes of self-renewal and multilineage differentiation of hematopoietic stem cells occur continuously during the lifetime of the organism and can be regulated in response to need. By generating knockout mice, a number of transcription factors have been implicated in hematopoietic stem/progenitor cell development. For example, homozygous disruption of either AML1, c-myb, or scl/tal1 results in defects in fetal hematopoiesis in mice that die during early embryonic development [3–6]. These genes appear to be essential for the development of multiple hematopoietic lineages. Transcription factors that appear to be lineage specific have also been identified. These genes include E2A and Pax-5, which are required for B-cell development [7–9], C/EBP α , which affects granulocytes [10], PU.1, which affects myeloid and lymphoid development [11], and NF-E2, which is essential for the development of megakaryocytes [12].

Several zinc finger genes are among the transcription factors that play important roles in hematopoiesis. Among them, the GATA2 and GATA3 genes appear to be involved in the early development of multiple lineages [13,14]. GATA2- and GATA3-deficient embryos die in early development as a result of impaired fetal liver hematopoiesis. Homologous knockout of another zinc finger gene, Ikaros, blocked lymphoid development [15]. Disruption of GATA1 in mice led to lineage-selective deficits in erythropoiesis [16]. Egr1, a myeloid differentiation primary response gene, has been shown to be essential for differentiation of hematopoietic cells along the macrophage lineage [17]. MZF1 inhibits myeloid differentiation, possibly by negative regu-

Offprint requests to: Donald Small, The Johns Hopkins Oncology Center, 600 North Wolfe Street, Rm. 3-109 Baltimore, MD 21287-5001

lation of hematopoietic genes such as CD34 and c-myc [18,19].

Hematopoiesis is a complex process in which the critical stage- and lineage-specific regulators establish lineage commitment through their combined action. A new transcription factor expressed in hematopoietic stem/progenitor cells is of interest, as it may control the transcription of one or more genes involved in the processes of self-renewal or differentiation of these cells. We report here the identification of SZF1 (stem cell zinc finger protein 1) from a cDNA library prepared from CD34⁺ human bone marrow cells. SZF1 codes for a protein containing C₂H₂-type zinc fingers and a KRAB domain. Two alternatively spliced transcripts were isolated. SZF1-2 was expressed in most cell types and tissues, whereas SZF1-1 appears limited to expression in CD34⁺ cells. SZF1-2 represses the FLT3/FLK-2/STK-1 promoter three- to four-fold, whereas SZF1-1 shows a minimal level of repression.

Materials and methods

cDNA library screening and DNA sequencing

Several hundred clones from a CD34⁺ cDNA library [20] were randomly sequenced and searched against available databases. One of the unique fragments obtained contained zinc fingers of the C₂H₂ type. This 1.2-kb fragment was used to screen cDNA libraries from K562 cells (Clontech, Palo Alto, CA), human lung (Clontech), and CD34⁺ cells. At least 10⁶ recombinants from each library were screened by standard procedures [20,21]. Positive clones were subcloned into pBluescript II KS⁻ (Stratagene, La Jolla, CA) and sequenced using the Sequenase Version 2.0 DNA sequencing Kit (USB, Cleveland, OH).

Sequence analysis

Computer searches of available databases were performed using the BLAST program [22]. Protein alignments were generated with the Genetics Computer Group Pileup program (Genetics Computer Group, Madison, WI).

Northern analysis

Northern hybridization was performed according to standard protocols [20,21]. In brief, 5 µg of each polyA⁺ RNA sample was incubated with 50% formamide, 6.5% formaldehyde, and 1× MOPS (pH 7.0) at 55°C for 15 minutes. The samples were electrophoresed in 1.2% agarose/formaldehyde gels and transferred to nylon membrane (Hybond, Amersham, UK) by the capillary method [23]. Blots containing RNA samples from cell lines as well as multitissue Northern blots obtained commercially (Clontech) were hybridized for 2 hours in 50% formamide, 5× SSPE, 10× Denhardt's, 2% SDS, and 100 µg/mL denatured salmon sperm DNA (Clontech) with randomly primed ³²P-dCTP-labeled probe and exposed to film after washing. The blots were then stripped and rehybridized with other probes [24].

RNAse protection assays

The RNAse protection assay was performed using the MAXIscript T3 in vitro transcription kit (Ambion, Austin, TX) [20]. The antisense RNA probe of SZF1 (bp1256-1421) was synthesized by run-

off transcription using bacteriophage T3 RNA polymerase on a pBluescript II KS⁻ plasmid containing a fragment of SZF1 linearized at the SacI site. It results in a probe of 230 bases containing 165 bases of SZF1 that would be protected by hybridization with SZF1 message. Total RNA 5 µg from each sample was hybridized with the ³²P-UTP-labeled antisense RNA probe and then treated with RNAse A and T1. A β-actin probe was included in the hybridization reactions as an internal control for RNA loading. Protected fragments were resolved on 8M urea, 6% acrylamide gel, and exposed to film (Kodak X-OMAT).

Reverse transcriptase polymerase chain reaction

One microgram of mRNA from each sample was reverse transcribed with Moloney murine leukemia virus reverse transcriptase (M-MLV-RT) (Gibco/BRL, Gaithersburg, MD) using random hexamers or oligo(dT)₁₅ (Boehringer Mannheim, Indianapolis, IN) as primers. Polymerase chain reaction (PCR) was performed with cycles of 95°C for 1 minute, 45°C for 1 minute, and 72°C for 2 minutes, repeated for 35 cycles. Primer pairs used for PCR were as follows: 1) 5'-CGGGATCCTAATACGACTCACTATAGGGAGACCACCATG-ATTGATTTTCAAATG-3'; 2) 5'-GGAATTCCTCAAGG-CTCAGTC-3'; 3) 5'-AAGAGACTGAGCCTTGAG-3'; 4) 5'-ATTTCTGCATGGAAACC-3'; 5) 5'-GGGGAGGCAACAGAATAT-3'; 6) 5'-ATAAGGTTTCTCCCCGGA-3'; 7) 5'-GAGGCTTTCGTGAAAAGT-3'; and 8) 5'-CAGGAGAGTGT-CATGGAA-3'.

Fluorescence in situ hybridization

Three independent P1 plasmids (1629, 1630 and 1631) with approximately 85-kb SZF1 genomic fragments were obtained from a human genomic DNA P1 library (DMPC-HFF1; Genome Systems, St. Louis, MO) by PCR screening using SZF1 oligonucleotides (5' primer CTGTGTTCTTCCATTAGC; 3' primer GGCCTTAGC-CATTTGTCT). P1 plasmids 1629, 1630, and 1631 containing the entire SZF1 gene were nick-translated with biotin-14 dATP (BRL, Gaithersburg, MD). Slides with chromosome spreads were made from normal male lymphocytes cultured with BrdU [25]. Fluorescence in situ hybridization (FISH) was performed as described with modifications [26]. Twenty microliters of hybridization mix (2× SSCP, 60% formamide, 10% dextran sulfate, 4 ng/µL biotinylated probe that had been coprecipitated with Cot-1 DNA [BRL] and herring sperm DNA) was denatured at 70°C for 5 minutes, pre-annealed at 37°C for 60 minutes, placed on slides, and hybridized at 37°C overnight. Slides were washed in 70% formamide/2× SSC at 43°C for 20 minutes, and two changes of 2× SSC at 37°C for 5 minutes each. Biotinylated probe was detected with an in situ hybridization kit (Oncor Inc., Gaithersburg, MD). In situ hybridization also was performed using metaphases that had been Giemsa banded and photographed prior to hybridization.

Cell culture, isolation, and RNA extraction

Most of the cell lines were grown in RPMI 1640 (Gibco RBL) with 10% HI-FCS (Hyclonr, Logan, UT), and Molm1 cells were maintained in RPMI 1640 medium with 20% HI-FCS. The normal lung epithelial primary cell line (HBE) came from a normal lung tissue sample and was cultured in keratinocyte growth medium with bovine pituitary extract (BPE) at 30 ng/mL (Clonetics, Walkersville, MD). CD34⁺ and CD34⁻ cells were isolated from normal bone marrow cells by immunomagnetic separation [20,27]. To induce differentiation, ML-1 cells and HL60 cells were incubated with TPA (12-0-tetradecanoylphorbol-13-acetate) at a concentration of

33 nM for 24 hours [28–30]. Total RNA was isolated from cell lines and the primary tumor samples by the guanidium thiocyanate method [31] and polyadenylated RNA was prepared using a mRNA isolation kit (Becton Dickinson Labware, Bedford, MA).

Plasmids and transfection

To create the pSTK-140 reporter, a PCR fragment that encompassed the sequence from the translation start codon (+61 bp) to 140-bp upstream of the transcription start site was cloned into the polylinker site of the promoterless pXP2 vector. It contains a Kozak sequence along with the luciferase coding cDNA downstream of the polylinker site. pNeoSZF1-1 and pNeoSZF1-2 were constructed by cloning the entire coding sequences of SZF1-1 and SZF1-2 into PCI-neo (Promega, Madison, WI). All of the constructs were confirmed by sequencing. Molm1 (a megakaryocytic cell line) and Laz 221 (a pro-B cell line) cells were transiently transfected by electroporation. For each transfection, 10 million cells for Molm1 or 20 million cells for Laz 221 were cotransfected by 20 µg of luciferase reporter plasmid and 20 µg of each SZF1 expression construct, empty vector or dH₂O control. Two micrograms of an internal control plasmid (pEQ176-β-galactosidase driven by the cytomegalovirus [CMV] promoter) also was included in each transfection to normalize luciferase activity for differences in electroporation efficiency. Luciferase activity for each transfection is expressed as percentage relative to the dH₂O control. For optimal transfection efficiency, 960 µFD and 290 V were used for Molm1 cells and 360 V for Laz 221 cells. Following electroporation, cells were incubated at 37°C for 8 hours in a 5% CO₂ incubator before harvest. Cells were washed and the luciferase assay was performed as described [32] and β-galactosidase assay as per the manufacturer (Galacto-light™, Bedford, MA). All transfections were performed in duplicate and repeated for a total of three separate experiments. The results for each set of transfection conditions were averaged and the error bars show the standard deviation. The krab domain (pdKRAB) and zinc finger domain (pdZINC) deletions were generated by PCR using the SZF1-2 cDNA.

In vitro transcription and translation and immunoprecipitation

In vitro transcription and translation were performed according to the manufacturer's instructions (TNT T7/T3 coupled Reticulocyte Lysate System; Promega). [³⁵S]methionine-labeled SZF1-1 and SZF1-2 proteins were produced from constructs that were generated by cloning the entire coding sequences of SZF1-1 and SZF1-2 into PCI-neo (Promega). Transcripts were generated with T7 RNA polymerase. In vitro transcripts from pNeoSZF1-1 and pNeoSZF1-2 were added to the reticulocyte lysate translation mixture containing 60 µCi of [³⁵S]methionine and incubated for 1 hour at 30°C. The lysates were diluted into phosphate-buffered saline with 1% NP-40. Proteins were immunoprecipitated 1 hour under these conditions with either the preimmune sera (10 µL), polyclonal antisera (10 µL) that was raised against a synthesized peptide (RQKAVTAEKSSDKRQ) located upstream of the zinc fingers (shown in Fig. 1 in italics), or the same antisera (10 µL) preincubated with the peptide (25 mg) used to generate it. The polyclonal antisera was prepared by HRP (Denver, PA) from a New Zealand white female rabbit with three boosts of the KLH-conjugated peptide. Antigen-antibody complexes were isolated on protein A-sepharose beads (Sigma, St. Louis, MO), and the pellets were washed five times with phosphate-buffered saline containing 0.5% Tween-20. The [³⁵S]methionine-labeled samples were analyzed by SDS-PAGE and autoradiography.

Results

Cloning of SZF1

The initial isolation of SZF1 was made through the random sequencing of a portion of several hundred clones from a cDNA library prepared from human bone marrow CD34⁺ cells. One of the fragments was unique but showed homology to the Kruppel family of zinc finger proteins. We therefore sequenced the entire 1200-bp fragment. Then, cDNA libraries from CD34⁺ cells, as well as libraries made from K562 cells and human lung (which in preliminary experiments showed expression of the transcript), were screened with the 1.2-kb fragment. Overlapping cDNA fragments that hybridized to the probe gave rise to two alternatively spliced cDNA products. Most regions were isolated several times in more than one cDNA library. The complete SZF1 nucleotide and predicted amino acid sequences are shown in Figure 1.

Sequence analysis of SZF1

Two predicted protein products of 421 and 361 amino acids are predicted from the open reading frame (ORF) of the two transcripts, termed SZF1-1 and SZF1-2. The size of the proteins is based on the usage of the 5'-proximal ATG codon, which follows the consensus Kozak site (underlined in Fig. 1) downstream of an in-frame stop codon. To test the prediction that this ATG is the most frequently used start site for translation initiation (at least in rabbit reticulocyte lysate), the in vitro transcription and translation assay was performed with full-length SZF1-1 and SZF1-2 constructs. As shown in Figure 2, left panel, the longest and most intense bands correspond to 48 kDa for SZF1-1 and 41 kDa for SZF1-2. This corresponds to the molecular weight predicted by the full-length ORF of each transcript if the predicted initiating methionine is used. The smaller bands seen in both lanes could represent degradation products or initiation at internal methionines. To confirm that the full-length products represent translation from the predicted ORF, immunoprecipitation (IP) was performed with polyclonal antisera generated to a peptide (shown in italics in Fig. 1) that is shared by the predicted ORF of both SZF1-1 and SZF1-2. As seen in Figure 2, right panel, the antisera recognizes the full-length translated products of both the SZF1-1 and SZF1-2 transcripts and detected a doublet in SZF1-2 that could result from use of both the 5' proximal ATG and the ATG located five amino acids downstream (Fig. 1).

Both transcripts share the same amino-terminal sequence upstream of the zinc fingers and through the first three zinc fingers, but diverge after amino acid 349 toward the end of the fourth zinc finger. In a region close to the amino terminal end, a Kruppel-associated box (KRAB) domain is present with highly conserved A (amino acids 31-73) and B (amino acids 74-93) elements. This domain was initially noted in zinc finger genes from *Xenopus* [33]. The KRAB domain has been suggested to form an α-helix and may be

SZF1-1

```

1   GTA GCA GCA CCG TGC GTG CGT GCG CAG ATG TGG GCC CCG CGG GAG CAG CTA CTG GGC TGG
61  GCT GCG GAA GCT CTG CCT GCC AAG GAT TCT GCC TGG CCC TGG GAA GAG AAG CCT AGA TAT
121 GGG^GTT CCA GTG GCA GAT GGG GCA CTG GCC AGA GGG GTT AGG AGG ATT TGG ACT CTC CTT
181 CCG AGT CGC TGT GAT ACG GCC TTC AAT TAT AGT TGT CTT TCC TTC TTG GAT TTC CAC CAC
241 TTC CCT GAA ACC ACG AGC TGG AAG CTG AGA CCA GCT GGT CAC TGC TGG GCC TGC TAG GAG
301 CCC ACA CGG AAG CTG CCC ACA GCC AGA CAC CAG AGC CTC AAA GGC CGC CAT CTT GAA AAA

361  CCA ACC CTG ACC TCC ACA ATG ATT GAT TTT CAA ATG TTG AAC CAG CTT TGC AGA ACT ATA
      met ile asp phe gln met leu asn gln leu cys arg thr ile
      <----->
421  ATA AAC CCA AGT GTG^ATA CCC TGT CTC AAG TAT TGC GGT GAT CAA ATA^GGA CCA GTG ACT
15  ile asn pro ser val ile pro cys leu lys tyr cys gly asp gln ile gly pro val thr
      ----- KRAB-A -----
481  TTC GAG GAT GTG GCT GTG CTT TTC ACT GAG GCA GAG TGG AAG AGA CTG AGC CTT GAG CAG
35  phe glu asp val ala val leu phe thr glu ala glu trp lys arg leu ser leu glu gln
      -----> <-----> <----->
541  AGG AAC CTA TAC AAA GAA GTG ATG CTG GAA AAT CTC AGG AAT CTG GTC TCA TTG GAA TCA
55  arg asn leu tyr lys glu val met leu glu asn leu arg asn leu val ser leu glu ser
      ----- KRAB-B ----->
601  AAG CCA GAA GTC CAT ACC TGC CCT TCT TGC CCT CTG GCC TTT GGC AGT CAG CAG TTC CTC
75  lys pro glu val his thr cys pro ser cys pro leu ala phe gly ser gln gln leu
      <=== ck2 =====>
661  AGC CAA GAT GAG CTA CAC AAT CAT CCT ATT CCA GGT TTC CAT GCA GGA AAT CAA CTC CAC
95  ser gln asp glu leu his asn his pro ile pro gly phe his ala gly asn gln leu his
      ----- PEST ----->
721  CCA GGA AAT CCC TGC CCA GAG GAT CAG CCA CAG TCA CAA CAT CCT TCT GAT AAA AAT CAC
115 pro gly asn pro cys pro glu asp gln pro gln ser gln his pro ser asp lys asn his
      <===
781  AGG GGG GCT GAA GCA GAA GAT CAA CGA GTG GAA GGA GGC GTC AGA CCC TTG TTT TGG AGT
135 arg gly ala glu ala glu asp gln arg val glu gly gly val arg pro leu phe trp ser
      == ck2 ==>
841  ACA AAT GAA AGG GGG GCT TTA GTG GGT TTC TCT AGC CTG TTC CAG AGA CCA CCA ATA AGC
155 thr asn glu arg gly ala leu val gly phe ser ser leu phe gln arg pro pro ile ser
      <=== ck2
901  TCT TGG GGA GGC AAC AGA ATA TTA GAG ATA CAG CTC AGT CCA GCC CAG AAT GCA AGC TCT
175 ser trp gly gly asn arg ile leu glu ile gln leu ser pro ala gln asn ala ser ser
      =====>
961  GAG GAA GTA GAC AGA ATT TCC AAG AGG GCA GAA ACC CCA GGG TTT GGA GCA GTC AGG TTT
195 glu glu val asp arg ile ser lys arg ala glu thr pro gly phe gly ala val arg phe

1021 GGG GAG TGT GCA CTA GCT TTT AAC CAG AAG TCA AAC CTG TTC AGA CAG AAG GCA GTC ACA
215 gly glu cys ala leu ala phe asn gln lys ser asn leu phe arg gln lys ala val thr
      <----- pkc ----->
1081 GCA GAA AAA TCT TCA GAC AAA AGG CAG TCA CAG GTG TGC AGG GAG TGT GGG CGA GGC TTT
235 ala glu lys ser ser asp lys arg gln ser gln val cys arg glu cys gly arg gly phe

1141 AGC AGG AAG TCA CAG CTC ATC ATA CAC CAG AGG ACA CAC ACA GGA GAA AAG CCT TAT GTC
255 ser arg lys ser gln leu ile ile his gln arg thr his thr gly glu lys pro tyr val

1201 TGC GGA GAG TGT GGG CGA GGC TTT ATA GTT GAG TCA GTC CTC CGC AAC CAC CTG AGT ACA
275 cys gly glu cys gly arg gly phe ile val glu ser val leu arg asn his leu ser thr

1261 CAC TCC GGG GAG AAA CCT TAT GTG TGC AGC CAT TGT GGG CGA GGC TTT AGC TGC AAG CCA
295 his ser gly glu lys pro tyr val cys ser his cys gly arg gly phe ser cys lys pro
      <---- pkc ---->
1321 TAC CTC ATC AGA CAT CAG AGG ACA CAC ACA AGG GAG AAA TCG TTT ATG TGC ACA GTG TGT
315 tyr leu ile arg his gln arg thr his thr arg glu lys ser phe met cys thr val cys
      <----- pkc -----> PEST ----->
1381 GGG CGA GGC TTT CGT GAA AAG TCA GAG CTC ATT AAG CAC CAG^AGG TGT CAA GTG ACG GTC
335 gly arg gly phe arg glu lys ser glu leu ile lys his gln arg cys gln val thr val
      ----->
1441 CCC TTG GAG GAA TGG TCT TTG CAT CTG ACT ACT TCC TTC TGC AAC TGT GTT CTT CCA TTA
355 pro leu glu glu trp ser leu his leu thr thr ser phe cys asn cys val leu pro leu

1501 GCT TCC ATG ACA CTC TCC TGC TTT ATT TTT TTC TAC ATC TCT AGC CTT TGC TGT TTC CTC
375 ala ser met thr leu ser cys phe ile phe phe tyr ile ser ser leu cys cys phe leu

1561 TCC TAC CCC ACC TTT AGA TTT TAC TCA GAG TTC AGT CTC CAG CCC TAC AAT CTG AGG GAC
395 ser tyr pro thr phe arg phe tyr ser glu phe ser leu gln pro tyr asn leu arg asp
      <----- pkc ----->
1621 ACC TTT ACC AGG TCC CCT TCC TAA CCC TCC AGT CCC AAA TCC AAG ATT CTT TAA CCA CAC
415 thr phe thr arg ser pro ser OCH

1681 TCT AAA AGT TCT TCA GAC TCA GGA CTT AAA CAT AGC CAC GCC ACC TTG GCC TTC AAT GAC
1741 AGG GAT CTA GCA ATG CTG CAT CAT CAG CCT TCC AAT ACC AGG TTT AAG GGT ATT TTA AAC
1801 ACA GCT CCT CTT AAA TCC TCC AAT CTC AGT ACC CAG TGT TTT AGC CAT GCT CGG GTG GCT
1861 AAA TTA CAT CCA GGA ATG GTG CCA GGG CCT TTA GCC ATT TGT CTC TCC TCA CAC TCC AGC
1921 CCA TAT GGC CCA GGT TCT GAC AGT TTG CCT TAC TCC CTT GGG CTG GGG CTA GCC CTA CCT
1981 GAT ACC CTG TGT CAA TGA GTG TAC CTT GGA GAG CTA TCC ACT CAG GCC CCA GTG CCT CTA
2041 TTT GCT AAG GGA CTC TGC CAC AGA AAA GAA GGG GAG AGA TGT TCA TGT AAC CTC AAA ATA
2101 CTT AGG CTT GGT TTT GAT GCT AGA GAG GAA AAA GGA CTT GGA GAG AGA GAA GGA ATG GCT
2161 GGT CCA GAG GCT TTT GTC CAC TCC CTC TCA CTG GAA GTG GTT GAT CTC CAG GAA ATC CCC
2221 AAG GTT AGC CTG CTT AGG GGA AGG GCT AGG GGT ACC TGG AAT GTA GGA TCT CCC CCA TGC
2281 CTG GCC TAC CAC CCT AAT GTG TCT GGA ATT GGT GGG TTC TTG GTC TTG CTG ACT TCA AGA
2341 ATG AAG CCG TGG ACC CTC ACG GTG AGT GTT ACA ATT CTT AAA

```

Figure 1. Continued.

SZF1-2

```

          <---- pkc ---->
1381   GGG CGA GGC TTT CGT GAA AAG TCA GAG CTC ATT AAG CAC CAG AGA ATT CAC ACG GGG GAT
335   gly arg gly phe arg glu lys ser glu leu ile lys his gln arg ile his thr gly asp

1441   AAG CCT TAT GTG TGC AGA GAT TGA GGC CGA GGC TTT GTA AGG AGA TCA TGT CTC AAC ACA
355   lys pro tyr val cys arg asp OPA

1501   CAC CAG AGG ATA CAT TCA GAT GAG AAG CCT TTT GTT TGC AGA GAG TGT GGG CGA GGC TTT
1561   CGT GCT AAA TCA ACT CTC CTC CTA CAC CAG TGG ACA CAT TCA GAG GTG AAA CCT CAC GTG
1621   TGT GAG GAG TGT GGG CAT GGA TTT AGC CAG AAG TCG TCG CTC AAA TCA CAT CGG AGA ACA
1681   CAC TCA GGG GAG AAG CCT TAT GTG TGT GGG GAA TGT GGG CGG GGA TTT AGC CGG AGG ATA
1741   GTC CTC AAT GGA CAC TGG AGG ACA CAC ACG GGA GAG AAG CCT TAC ACG TGC TTT GAG TGT
1801   GGG CGA AAC TTT AGC CTC AAG TCC GCT CTT AGT GTA CAT CAG AGG ATA CAC TCT GGG GAG
1861   AAG CCT TAT GCA TGC ACG GAG TGT GGG CAA GGC TTT ATC ACG AAA TCA CAG CTC ATC AGA
1921   CAC CAG AGG ACA CAC ACA GGA GAA AAG CCT TAT GTC TGC GGA GAG TGT GGG CGA GGC TTT
1981   ATA GCT CAG TCA ACC CTC CAC TAC CAC CGG AGT ACA CAC TCC AAG GAA AAA CCT TAT GTG
2041   TGC AGC CAG TGT GGG CGA GGC TTT TGT GAT AAA TCA ACT CTC CTC GCA CAC GAG CAG ACA
2101   CAT TCA GGG GAG AAG CCT TAT GTG TGT GGG GAA TGT GGG CGG GGA TTT GGC CGG AAG ATA
2161   CTC CTC AAC AGA CAC TGG AGG ACA CAC ACA GGA GAG AAA CCT TAC GCA TGC ATC GAG TGT
2221   GGG CGA AAC TTT AGC CAC AAG TCC ACT CTC AGC TTA CAT CAG AGG ATA CAC TCG GGG GAG
2281   AAG CCT TAT GCA TGC GTG GAG TGT GGG CAA AGC TTT AGG AGA AAG TCA CAG CTC ATC ATA
2341   CAC CAG AAG ATA CAC TCG GGG AAA AGC TTT AGA GGT GCA AGG AGT GAG GAT GTG ATT TTA
2401   GCA ACA AGT CAG CCA TCA GCC ACA CCA GCG GAA ATG CTT AGG GAG AAG CCT TGT TTG TAA
2461   GGT AAT GTG GAC AGA GCT GTA CGT GGA CAT CAT TAC TTG TCA CGT GTC AGA GGA CAC ACT
2521   CGG GAG AAA CCT TCA TGG AGT GAG AGT AAG GTG TTG GCT GGA AGT GGC CCC TTA AGA GAT
2581   ACT TGG AGT CAA ATC TAT CCA CTG TAC GCC CAC CCC ACT CTT GTT CTA AGA GGT TTG GGG
2641   ACA GTC TTT TGA CCC CTT ACA TTC CTT TAG ATG TGA AGA TGA CAG AGA TCT AAC TTC TGA
2701   GAG CAG AGG TGT CAA GTG ACG GTC CCC TTG GAG GAA TGG TCT TTG CAT CTG ACT TCC
2761   TTC TGC AAC TGT GTT CTT CCA TTA GCT TCC ATG ACA CTC TCC TGC TTT ATT TTT TTC TAC
2821   ATC TCT AGC CTT TGC TGT TTC CTC TCC TAC CCC ACC TTT AGA TTT TAC TCA GAG TTC AGT
2881   CTC CAG CCC TAC AAT CTG AGG GAC ACC TTT ACC AGG TCC CCT TCC TAA CCC TCC AGT CCC
2941   AAA TCC AAG ATT CTT TAA CCA CAC TCT AAA AGT TCT TCA GAC TCA GGA CTT AAA CAT AGC
3001   CAC GCC ACC TTG GCC TTC AAT GAC AGG GAT CTA GCA ATG CTG CAT CAT CAG CCT TCC AAT
3061   ACC AGG TTT AAG GGT ATT TTA AAC ACA GCT CCT CTT AAA

```

Figure 1. Nucleotide and predicted amino acid sequence of SZF1 cDNAs. The nucleotide and predicted amino acid sequence of SZF1-1 and SZF1-2 are shown with numbered nucleotides and amino acids listed to the left of each product. RNA splicing sites as determined by sequence comparison of cDNAs with genomic DNA are shown (ˆ) in the SZF1-1 sequence at nucleotides 123, 435, 469, 578, 594, and 1422. The splicing site at 1422 is absent in SZF1-2; thus the sequence between nucleotides 1422 and 2706, which is present in SZF1-2, is spliced out of SZF1-1. The deduced amino acid sequence starts from the conserved Kozak translation start site (underlined). Zinc fingers are underlined from the initial conserved cysteine to the final conserved histidine beginning at amino acid 247. KRAB-A and -B domains are indicated above each domain. Phosphorylation consensus sites for potential casein kinase II (ck2) and protein kinase C (pkc), and for PEST sequences are labeled above each element. Two instability motifs (ATTTA) in the 3' untranslated region of SZF1-2 are underlined. The peptide sequence used to generate polyclonal serum is shown in italics.

involved in protein–protein interactions [34,35]. Toward the carboxy end, four predicted zinc fingers of the C_2H_2 type are present in SZF1-2, followed by 10 amino acids before the stop codon. In SZF1-1, the fourth zinc finger is incomplete, with the final histidine replaced by a glutamine residue. The lack of a second histidine in the final zinc finger has been observed in other zinc finger proteins [36]. An additional 74 amino acids follow the zinc fingers at the carboxy terminus of SZF1-1.

Several regions of the deduced proteins suggest that they may be substrates for phosphorylation. There are three potential casein kinase II phosphorylation sites upstream of the zinc fingers. In addition, there are several regions in and near the zinc fingers of both SZF1-1 and SZF1-2 that match the consensus motif for cyclic AMP (cAMP)-dependent protein kinase and protein kinase C phosphorylation sites [37,38].

Some features suggest that both the mRNA and protein products of SZF1 may have short half-lives. Two AUUUA sequences appear beginning at nucleotides 1641 and 1725 in the 3' untranslated region of SZF1-2. Those sequences confer mRNA instability and lead to a short half-life in other messages that express this feature [39]. There is a PEST consensus site present in both predicted proteins at

amino acids 94–108 and at amino acids 329–342 in SZF1-2. PEST sequences, rich in proline, acidic, serine, and threonine residues, are often present in proteins with short half-lives and may signal for their rapid degradation by the proteolytic machinery of the cell [40].

The cDNA sequences of SZF1-1 and SZF1-2 were used to search the available nucleotide databases to determine the most highly related genes. The result showed that ZNF133, Kid1, and ZNF85 were the most highly related genes with homologies of 65%, 55%, and 45% at the nucleotide level, respectively. ZNF133 [41], Kid1 [42], and ZNF85 (D.A. Poncelet, personal communication) are all zinc finger proteins that have been suggested to play roles in transcriptional repression. The region of SZF1 that encompasses the KRAB and zinc finger domains has the highest degree of homology to these transcription factors (Fig. 3). The identity of the KRAB domain of SZF1 with that of ZNF133 is 68% for KRAB-A and 22% for KRAB-B; with Kid1 it is 74% for KRAB-A and 12% for KRAB-B; and with ZNF85 it is 70% for KRAB-A and 26% for KRAB-B (Fig. 3A). The zinc finger sequence of SZF1 also shows a high degree of homology to these transcription factors, with an identity of about 50% (Fig. 3B). A much lower degree of homology was found in the remainder of the SZF1 coding sequence.

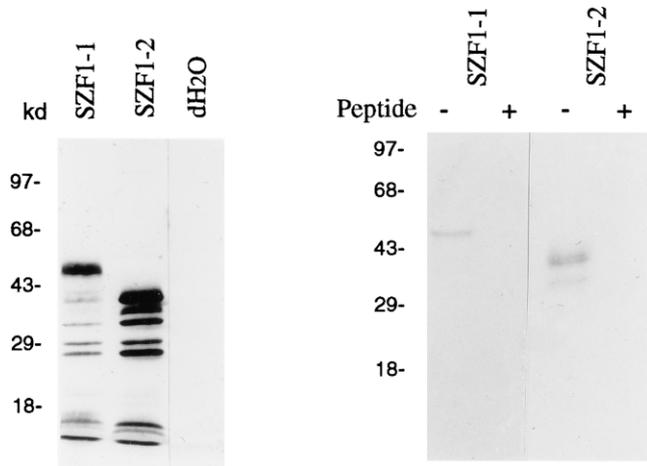


Figure 2. Labelling and immunoprecipitation of in vitro transcribed and translated (TNT) SZF1 proteins. (Left) ³⁵S-methionine was used to label the proteins generated from TNT of full-length SZF1-1 and SZF1-2 constructs. These products, together with the products of a TNT reaction in which no template was added (dH₂O), were resolved on a 10% SDS-polyacrylamide gel, which was then exposed to film. Molecular weight markers are shown to the left. (Right) The TNT products of SZF1-1 and SZF1-2 were immunoprecipitated (IP) with a polyclonal antisera generated to a peptide present in the predicted open reading frame of both transcripts. IP reactions also were conducted after preincubation of the antisera with 25 μg of the peptide to demonstrate specificity of the antiserum (+ peptide lanes). The IP products were then processed as above.

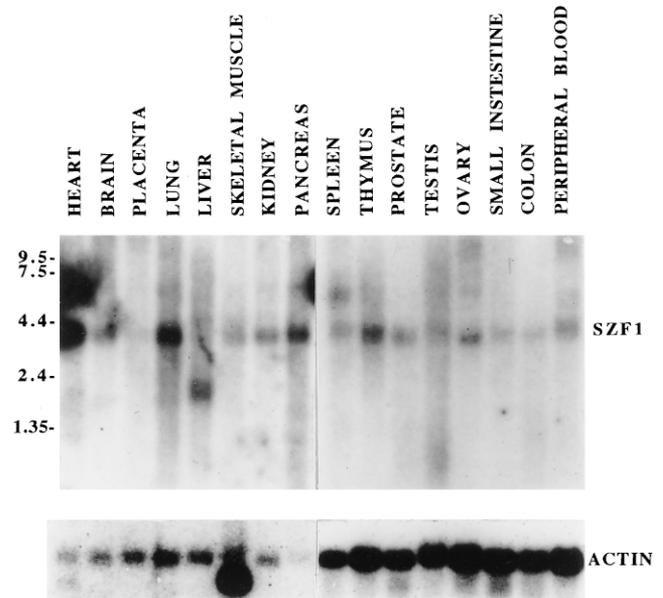


Figure 4. SZF1 expression in normal human tissue. Northern blots containing 2 μg of polyA⁺ RNA from multiple normal human tissues were hybridized with a SZF1 3' probe that contained the 3' region (nucleotide 1414-2389) of SZF1-1 downstream of the zinc fingers (top). The blots were then stripped and reprobed with β actin (bottom). The source of RNA is noted above each lane. Molecular weight size markers are noted to the left.

Tissue and cell line expression of SZF1

To examine the expression of SZF1 in normal human tissues, hybridization of SZF1 probes with multitissue Northern blots was performed. Both 3' and 5' probes were generated, which omit the zinc finger region to decrease the possibility of cross hybridization with other zinc finger-containing genes. Both probes would hybridize with both SZF1-1 and SZF1-2 transcripts. An example of hybridiza-

tion with the 3' probe is shown in Figure 4. A 4.2-kb band was seen in most human tissues, including heart, brain, placenta, lung, liver, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, and blood. The lane containing liver RNA showed a 2-kb band instead. Some samples also showed a larger transcript of approximately 6 kb. The 5' probe also hybridized to the 4.2-kb band in all these tissues and showed a faint band of 1.1 kb (data not shown).

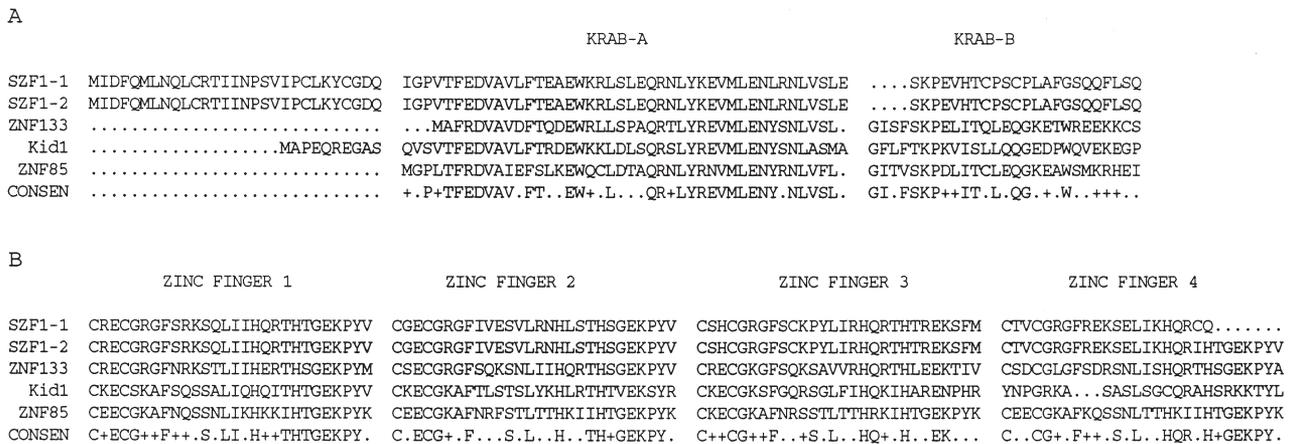


Figure 3. KRAB and zinc finger homologies. Alignment of the KRAB domains (A) and zinc finger domains (B) of SZF1 with the transcriptional repressors ZNF133, Kid1, and ZNF85. Protein alignments were generated with the GCG Pileup program. The consensus sequence is noted only if three of the four proteins have the identical amino acid at this position. (+) indicates the amino acids in this position are conservative substitutions.

To determine if SZF1 was expressed in hematopoietic cells, Northern blotting was performed with samples from a number of hematopoietic cell lines. A discrete 4.2-kb band was seen from several samples, including the RL, HEL, and K562 cell lines (data not shown). The more sensitive RNase protection assay was used to further screen for expression (Fig. 5A). Although the level of expression varies, protected species are seen from every hematopoietic cell line tested, including Jurkat, K422, Raji, Molt-16, RL, HEL, RPMI-8402, K562, ML-1, KG1a, Molt-3, and REH. These represent myeloid, lymphoid, and erythroid lineages. A faint band was seen for the bone marrow sample in a longer exposure. The appearance of the doublet probably results from variable digestion of the ends of the protected RNA and varies from experiment to experiment (e.g., see Fig. 8). As is evident, the probe is not protected from digestion by no RNA or 20 μ g of tRNA. We also isolated RNA from bone marrow mononuclear cells, and CD34-depleted (CD34⁻) and CD34-enriched (CD34⁺) cells and used this RNA for the RNase protection assay to assess expression in these fractions. Figure 5B shows that the gene is expressed in all three of these bone marrow fractions and appears highest in the CD34⁺ fraction.

The Northern blotting and RNase protection experiments conducted do not distinguish between the two SZF1 transcripts. We therefore used reverse transcriptase PCR (RT-PCR) to test for the presence of the two SZF1 transcripts in the polyA⁺ RNA from hematopoietic cell lines, lung cancer cell lines, fetal lung tissue, and a normal lung epithelial cell primary culture. The primers used in the assay cross the SZF1-2

exon that is spliced out of the SZF1-1 transcript and therefore generate a 133-bp RT-PCR fragment from the SZF1-1 transcript and a 1.4-kb fragment from the SZF1-2 transcript (Fig. 6C, primer pair 7-8). Figure 6A shows that RNA from unfractionated bone marrow gives rise to both of the bands, representing SZF1-1 and SZF1-2 transcripts (the SZF1-2 band is fainter and probably results from the decreased efficiency of PCR amplification because of the relative sizes of the products). Fractionation of the bone marrow into the stem and progenitor enriched CD34⁺ fraction (97.5% CD34⁺) with subsequent PCR using the same primers results in only the band representing the SZF1-1 transcript. Southern blotting of RT-PCR amplified products did show a faint band for the SZF1-2 transcript in CD34⁺ cells. In contrast, the RNA isolated from CD34⁻ cells (<0.01% CD34⁺) gives rise to only the larger product, representing SZF1-2. None of the mRNA samples tested from hematopoietic cell lines or other tissues gave rise to the 133-bp product expected from SZF1-1. Instead, most of the mRNA preparations showed the 1.4-kb product expected from the SZF1-2 transcript. Controls in which reverse transcriptase was not added to the RT reactions (“-” lanes) confirmed that the signal was not amplified from contaminating DNA. Southern analysis confirms that the 133-bp and 1.4-kb RT-PCR fragments hybridize to specific internal oligonucleotides and the hybridization of the 1.4-kb fragment to its internal oligonucleotide was also detected in the RT-PCR products from the CD34⁺ fraction (data not shown). Thus, the bands seen in the Northern blots in Fig. 4 likely represent transcripts of SZF1-2.

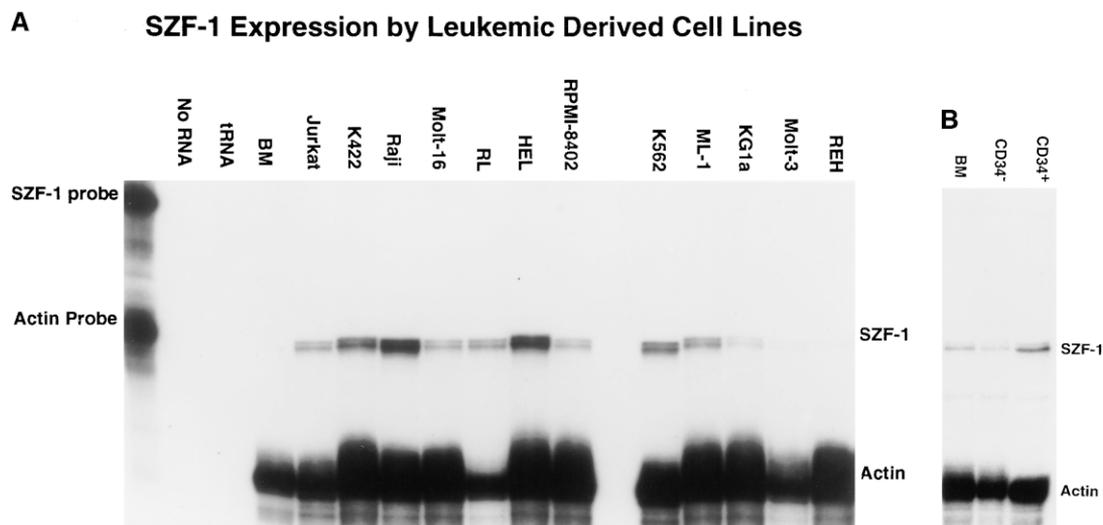


Figure 5. RNase protection analysis of the expression of SZF1 in hematopoietic cell lines and bone marrow fractions. (A) Total RNA 5 μ g from each sample was hybridized with a ³²P-UTP-labeled 230-base antisense RNA probe that contains 165 bp of SZF1 (bases 1256-1421) attached to 65 bp of vector sequence. A β -actin probe was included in the hybridization reactions as an internal control for RNA loading. After hybridization, the samples were treated with RNase A and T1 followed by electrophoresis on a 6% polyacrylamide gel and autoradiography. The SZF1 and β -actin probes are shown on the left. The source of the RNA added is noted above each lane: BM = total bone marrow cells; Jurkat, Molt-3, Molt-16, RPMI-8402 = T-lineage acute lymphocytic leukemia; K422, RL, REH = B-lineage acute lymphocytic leukemia; Raji = Burkitt's lymphoma; HEL = erythroleukemia; K562 = chronic myelogenous leukemia; ML-1, KG1a = acute myelocytic leukemia. The protected SZF1 and actin fragments are indicated to the right. (B) Approximately 5 μ g of total RNA from bone marrow mononuclear cells (BM), and bone marrow cells depleted for CD34 (CD34⁻) or enriched for CD34 (CD34⁺) were treated as described in (A).

To determine if complete transcripts coding for SZF1 were present in normal human bone marrow CD34⁺ cells, several pairs of primers spanning different regions of the transcript were used for RT-PCR (Fig. 6B). All of the fragments amplified by the primer pairs upstream of the alternative splice site are present in both the CD34⁺ and CD34⁻ fractions. Confirming the earlier result, the 133-bp SZF1-1 fragment appears only in normal human marrow CD34⁺ cells and not in CD34⁻ cells after amplification with primer pair 7-8. Controls in which reverse transcriptase was left out of the reaction demonstrated that the signals were not from contaminating DNA (data not shown).

Transcriptional repression of SZF1 on the FLT3/FLK-2/STK-1 promoter

The restricted expression of SZF1-1 in CD34⁺ cells and the ubiquitous expression of SZF1-2 might reflect a role for these transcripts in the regulation of the other genes in-

involved in hematopoiesis. One of these genes could be FLT3/FLK-2/STK-1, which encodes a transmembrane tyrosine kinase receptor that is primarily expressed in human CD34⁺ stem/progenitor cells [20,43–45]. Its ligand (FL) has been shown to stimulate the proliferation of one of the most primitive hematopoietic fractions (CD34⁺/CD38⁻ cells) and to synergize with other factors to expand both bone marrow and umbilical cord blood cells [46–50]. We have recently finished a preliminary characterization of the FLT3/FLK-2/STK-1 promoter and found that the first 140 bp upstream from the transcription start site constitutes the minimal region necessary for promoter activity (Liu et al, submitted). We were interested in investigating the possible involvement of both SZF1 products in the regulation of the FLT3/FLK-2/STK-1 promoter. We performed transient transfection assays with a reporter construct in two FLT3/FLK-2/STK-1 positive cell lines, Molm1, a megakaryocytic cell line, and Laz 221, a pro-B cell line. The reporter construct (shown at

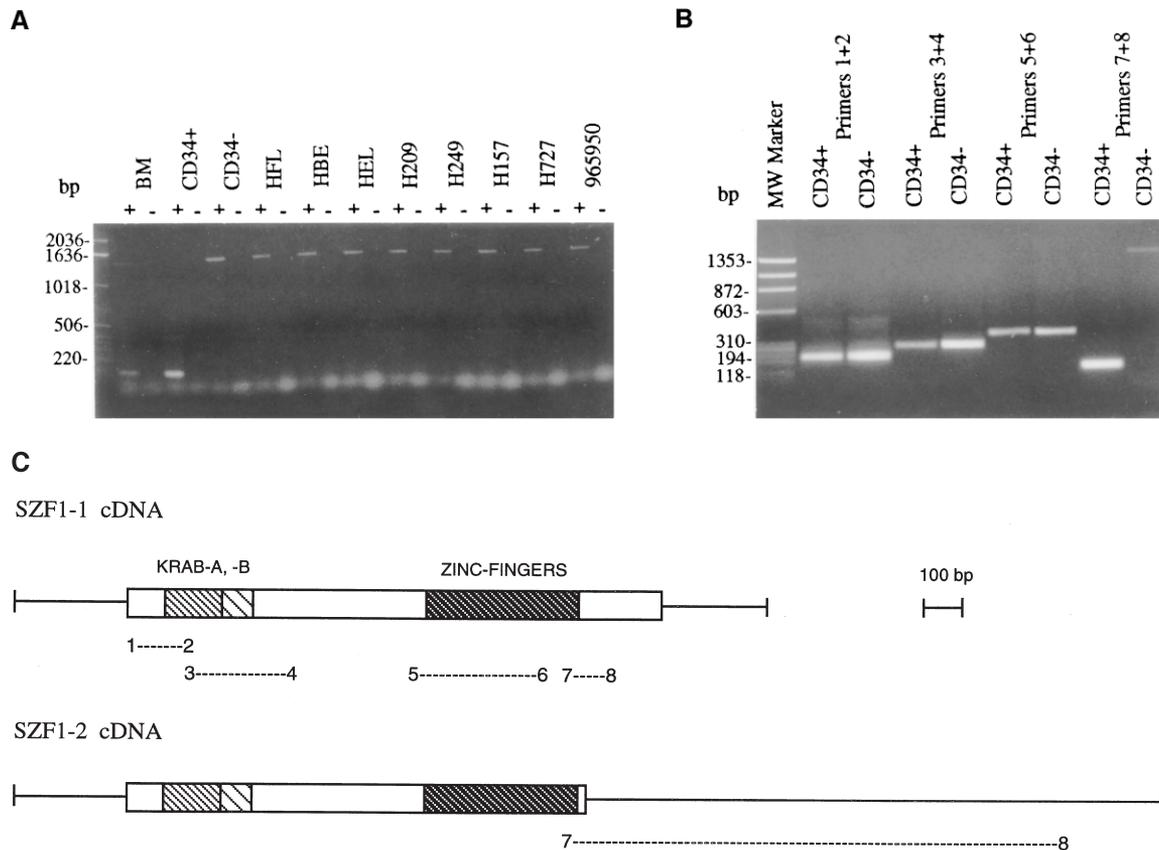


Figure 6. RT-PCR analysis of SZF1 expression in bone marrow, CD34⁺ and CD34⁻ cells, and other tissues. (A) CD34⁺ and CD34⁻ cells were purified from human bone marrow mononuclear cells by immunomagnetic separation, and RNA was isolated by the guanidium thiocyanate method. One microgram of each mRNA sample was reverse transcribed using random hexamer priming. The RT-PCR reactions with the primer pair 7-8 (133-bp product for SZF1-1 or 1417-bp product for SZF1-2) were resolved in a 1% agarose gel, stained with ethidium bromide and photographed. (+) = RT⁺ reaction; (-) = RT⁻ reaction; BM = human bone marrow mononuclear cells; HFL = normal human fetal lung; HBE = primary human lung epithelial cell culture; HEL = hematopoietic cell line; H209, H249 = SCLC cell lines; H727, H157 = HSLC cell lines; 965950 = primary small cell lung cancer tissues. (B) RT-PCR fragments generated by primer pairs 1-2 (210 bp), 3-4 (324 bp), 5-6 (392 bp), and 7-8 were resolved in a 2% agarose gel. The source of each mRNA and the primer pairs used are shown above each lane. DNA size markers are shown on the left. (C) Map indicating location of the KRAB and zinc finger domains and the primer pairs used for RT-PCR on SZF1-1 and SZF1-2 cDNAs.

the bottom of Fig. 7), in which the luciferase cDNA is under the control of the 140-bp STK1 promoter fragment, was co-transfected with each of the SZF1 constructs (shown to the left in Fig. 7). The percentage of the luciferase activity relative to the dH₂O control for each construct is shown in Fig. 7. In both cell lines, overexpression of SZF1-2 resulted in a strong repression of the STK1 promoter (three- to fourfold), whereas SZF1-1 resulted in only weak repression (less than twofold). Cotransfection of constructs truncating either the KRAB or zinc finger domains greatly restored promoter activity, suggesting that both domains are required for repression.

Several of the hematopoietic-derived cell lines expressing SZF1 differentiate in response to various agents. We decided to measure SZF1 expression as a function of differentiation induced in HL60 and ML-1 cell lines by TPA. RNA, isolated from control cells and cells exposed to TPA for 24 hours, was used in an RNase protection assay with actin as an internal control to allow for normalization (Fig. 8). Quantitation by phosphorimager scanning of the gel showed that SZF1 expression was decreased by 70% by 24 hours of induction of the HL60 cells. In ML-1 cells, expression of SZF1 was decreased by 84% of control levels by 24 hours of differentiation. Thus, SZF1 expression is greatly repressed upon differentiation, at least in these two cell lines.

Genomic organization of SZF1

Three overlapping P1 clones that each contain approximately 85-kb genomic DNA fragments were obtained by PCR screening of a P1 library. Subclones were generated

from the P1 plasmids to facilitate sequencing and mapping. RNA splicing sites as determined by sequence comparison of cDNAs with genomic DNA revealed that the SZF1-1 and SZF1-2 cDNAs share the first five splice sites at nucleotides 123, 435, 469, 578, and 594 (as shown by ○ in Fig. 1). However, SZF1-1 results from an internal splicing event within exon 6 at nucleotide 1422. Thus, the sequence between nucleotides 1422 and 2706, which is present in SZF1-2, is spliced out of SZF1-1 (Fig. 9). This alternative splicing results in a different carboxy terminus for the two translated protein products. All of the other exons are present in both cDNAs. Sequencing of the corresponding genomic DNA subclones confirmed the 5' upstream and 3' stop codons, which are present in the cDNAs. Southern blot analysis of DNA isolated from human cells demonstrated that SZF1 is a single copy gene (data not shown).

Chromosomal localization of SZF1

FISH was performed to determine the chromosomal localization of SZF1. P1 plasmids with inserts that encompass the entire coding sequence of SZF1 ORFs were used as the probe. An example of the FISH results is shown in Figs. 10A–B. Clear paired signals were observed only on chromosome 3. Analysis of 34 metaphase cells showed 23 cells (67%) had at least one pair of signals. There were 41 of 44 signals located on chromosome 3, as indicated by G-banding [25], with all of these signals on band p21. In addition, 9 of 10 signals analyzed from the pre-G-banded metaphases had the same 3p21 localization. The ideogram representing

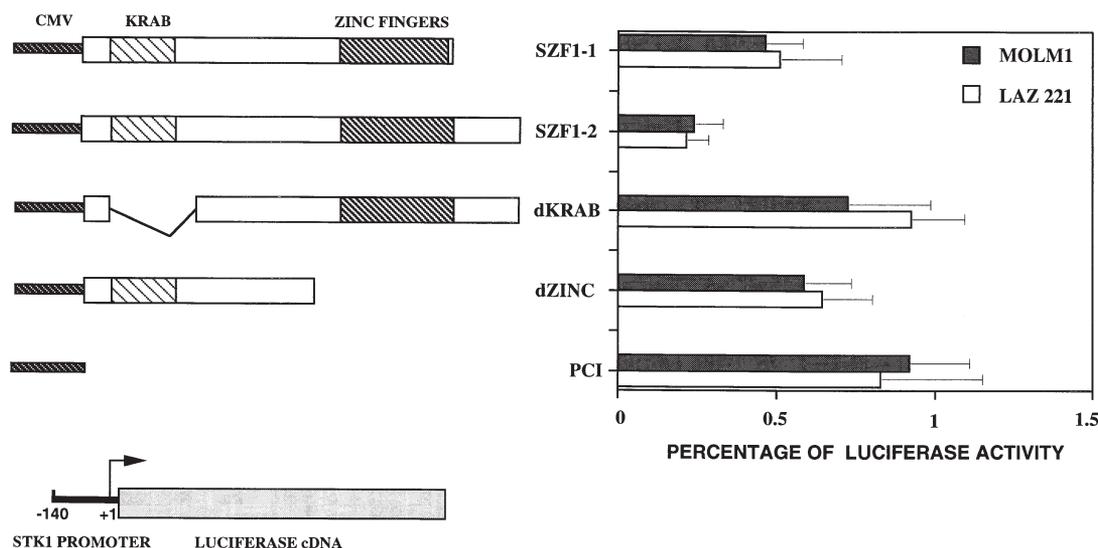


Figure 7. Functional characterization of SZF1 proteins. Twenty micrograms of each SZF1 construct (as indicated schematically on the left) or dH₂O together with 20 μg pSTK-140 reporter (shown at the bottom) and 2 μg pEQ176-β-galactosidase plasmid were transfected into 10 million Molm1 or 20 million Laz 221 cells by electroporation. Following electroporation, cells were incubated at 37°C for 8 hours in a 5% CO₂ incubator before harvest. Cells were washed and extracts were then used for the luciferase and β-galactosidase assays. All transfections were done in duplicate for each experiment, and at least three independent experiments were performed. The activity of β-galactosidase was used to normalize the luciferase values for differences in transfection efficiency. The luciferase activities for each reporter construct in Molm1 cells (dark bar) and Laz 221 cells (open bar) are expressed as percentage relative to the dH₂O control. The length of the error bars show the standard deviation.

SZF-1 Expression with Differentiation

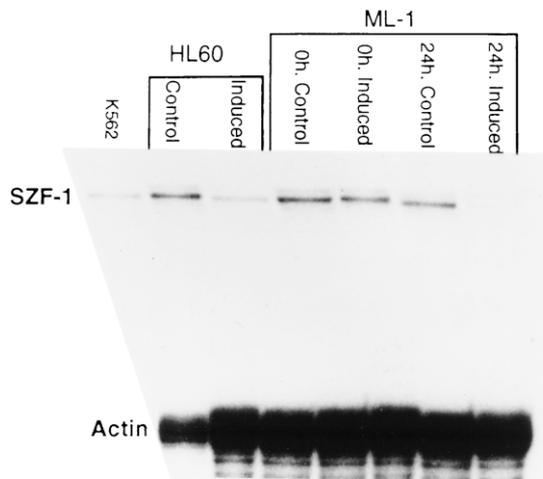


Figure 8. SZF-1 expression with differentiation. HL60 cells and ML-1 cells were treated with TPA for 24 hours. Five micrograms of total RNA isolated from control cells or cells exposed to TPA was added to the RNase protection assay with the same ^{32}P -UTP-labeled 230-base anti-sense RNA probe used in Figure 5. The source of each RNA sample is noted above each lane. The SZF1 and β -actin protected bands are indicated to the left.

the results of chromosomes with paired signals is shown in Fig. 10C. The results indicate that SZF1 maps to human chromosome 3p21, a region that has been implicated in a number of tumors [51–56] and in hematologic malignancies including MDS, AML, CML, and hairy cell leukemias [57–61].

Discussion

In this study, we cloned a novel zinc finger gene from cDNA library prepared from CD34⁺ human bone marrow cells. SZF1 contains the zinc finger motif of the C₂H₂ type,

which is the most commonly used motif for DNA binding by transcription factors [62,63]. There are many cases in hematopoietic development where zinc finger genes play an important role (e.g., Ikaros, GATA gene family, Egr1). The KRAB domains present in SZF1, and found in approximately one-third of zinc finger proteins of the C₂H₂ type, have been proposed to form helical structures that are involved in protein–protein interactions [34,35]. Several lines of evidence suggest that KRAB domains play roles in transcriptional repression. KRAB domains from several different proteins serve as strong transcriptional repressors when fused to the GAL4 DNA binding domain [64–67]. The repression function of KRAB domains may be mediated through the interaction with other factors [68,69]. The KRAB domain in SZF1 displays a high degree of homology with a number of transcriptional repression factors: ZNF133, Kid1, and ZNF85 [41,42] (DA Poncelet, personal communication). Consensus sites in SZF1 for casein kinase II, protein kinase C, and cAMP-dependent protein kinase phosphorylation suggest that phosphorylation may play a role in its function. Regulation by phosphorylation has been demonstrated in a number of other zinc finger genes [70–72]. It appears that the expression of both SZF1 products are under tight control. Both contain PEST sequences likely to result in a short half-life for the proteins, and SZF1-2 contains AUUUA sequences at its 3' end, which are likely to result in a short half-life for this transcript.

The two alternatively spliced transcripts of SZF1 are expressed differently. SZF1-2 is expressed in almost all tissues and cell types, although at widely varying levels. Of the many tissues and cell lines examined, the SZF1-1 transcript appears to be restricted to CD34⁺ hematopoietic stem/progenitor cells. This is supported by our isolation of cDNA clones for SZF1-1 only from the CD34⁺ cell cDNA library. SZF1-1 expression was not detected by RT-PCR from a number of hematopoietic cell lines representing a va-

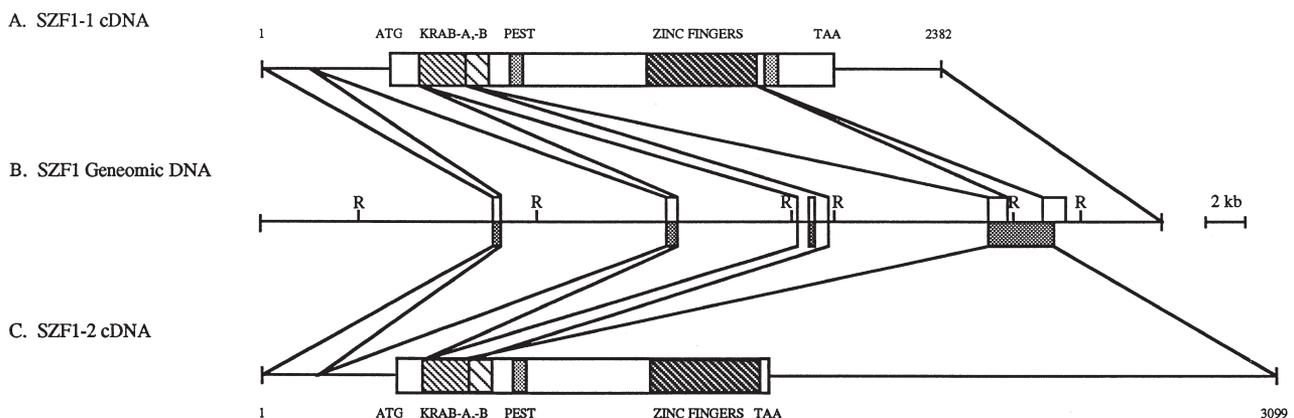


Figure 9. Genomic organization of SZF1 cDNAs and predicted motif structures. The SZF1-1 cDNA (open boxes) and SZF1-2 cDNA (hatched boxes) as well as the boundary of introns and exons are schematically presented in (B), with the approximate location of EcoRI sites (R) indicated. cDNA motifs and sequences that encode predicted zinc fingers, KRAB-A, and KRAB-B domains and the PEST sequences are indicated in (A) for SZF1-1 and in (C) for SZF1-2. The lines indicate the splicing events that occur from SZF1 to result in the observed cDNAs.

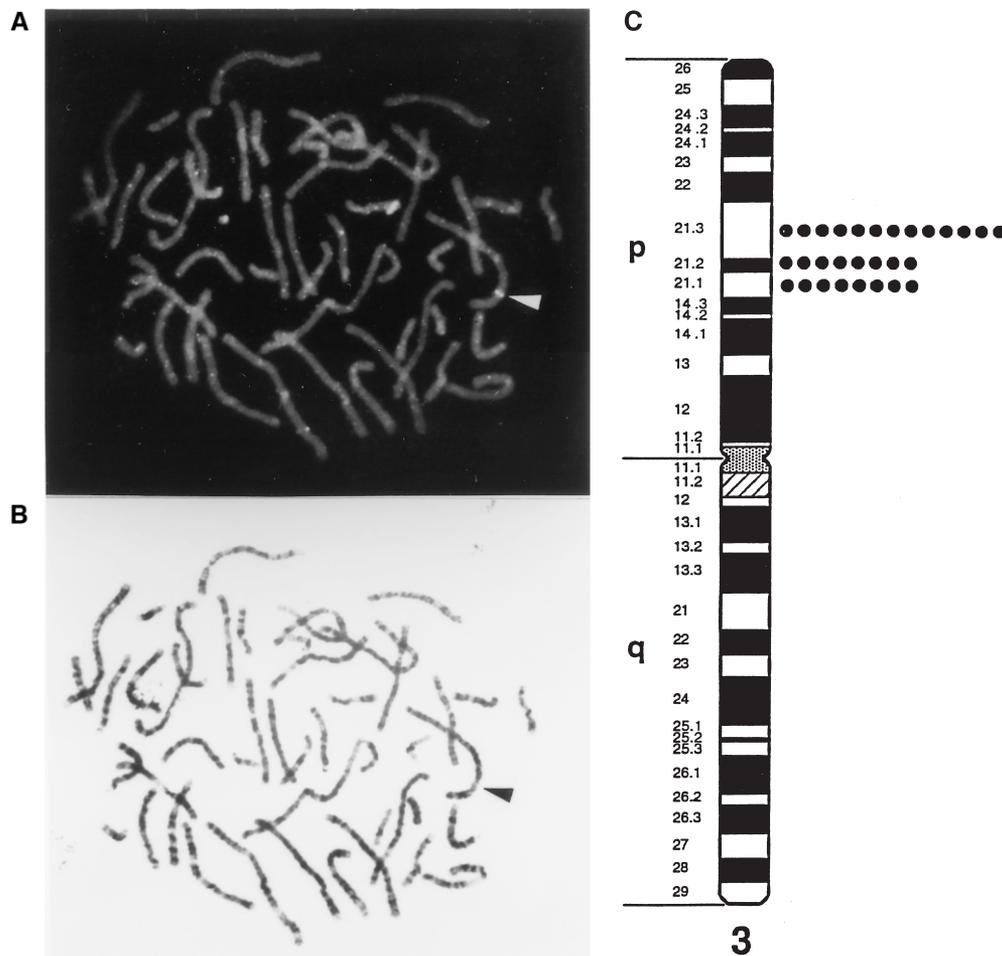


Figure 10. Chromosomal localization of SZF1 by FISH. (A) P1 plasmids encompassing the SZF1 gene were nick-translated with biotin-14 dATP and hybridized to metaphase chromosome spreads from normal lymphocytes cultured with BrdU. The specific paired FISH signals of the SZF1 gene are shown with an arrow on chromosome 3 (DAPI stained). (B) Metaphases were G-banded and photographed prior to FISH. The arrow indicates band 3p21, corresponding to the FISH signals seen in (A). (C) The chromosome ideogram of paired signals from FISH. Each dot represents a paired signal seen on metaphase chromosomes.

riety of cell types. For this reason, the bands seen on Northern blotting represent SZF1-2 transcripts. There are several differences in the predicted protein products coded by the two transcripts. SZF1-1 is missing the final histidine residue of its fourth zinc finger and has a different 73 amino acid carboxy terminus, as compared to SZF1-2. These changes could alter the DNA-binding specificity or protein-protein interactions. The unique and dominant expression of SZF1-1 in CD34⁺ cells suggests a potential role in the transcriptional control of genes expressed in CD34⁺ cells. SZF1-1 is less effective in the repression of the FLT3/FLK-2/STK-1 promoter compared with that of SZF1-2. It appears SZF1-2 is expressed only at low levels in CD34⁺ cells. One interpretation is that SZF1-1 loses much of the repressional activity by the alternative splicing that occurs. This mechanism might be present in CD34⁺ cells to protect the expression of FLT3/FLK-2/STK-1 or perhaps other genes important for

the preservation of this cell population. Alternatively, SZF1-1 might be repressing the expression of genes involved in lineage commitment or differentiation. Perhaps this is why SZF1-1 is not expressed past the level of CD34 expression in bone marrow cells. There is precedence for differential expression of different spliced zinc finger proteins in hematopoietic cells. MZF-1 and 2 are splice products of the same gene and have different expression patterns [73,74]. In addition, overall expression of the SZF1 gene seems to decrease with commitment of CD34⁺ cells (Fig. 5B) and with forced differentiation of the HL60 and ML-1 cell lines (Fig. 6). The decreased expression of SZF1 with forced differentiation implicates a lack of role for SZF1-2 once the cells are terminally committed, at least for these two leukemic-derived cell lines. Determination of the cognate binding sites of SZF1 and analysis of SZF1 domain function in hematopoietic cells will help to elucidate its function.

Acknowledgments

This work was supported by grants from the National Institute of Health (DK48709 and HL50383 to Dr. Small; P30CA06973 to Drs. Small and Griffin), American Cancer Society (to Dr. Small), and Gryphon Pharmaceuticals (to Drs. Small and Civin). The authors would like to thank Dr. Alan Friedman for his critical reading of the manuscript. Gryphon Pharmaceuticals, Inc. provided funding for the study described in this article. The sequences reported in this article have been deposited in the GenBank database (accession numbers AF114816 and AF114817).

References

- Orkin SH (1995) Transcription factors and hematopoietic development. *J Biol Chem* 270:4955
- Shivdasani RA, Orkin SH (1996) The transcriptional control of hematopoiesis. *Blood* 87:4025
- Mucenski ML, McLain K, Kier AB, Swerdlow SH, Schreiner CM, Miller TA, Pietryga DW, Scott WJ, Potter S (1991) A functional *c-myb* gene is required for normal murine fetal hepatic hematopoiesis. *Cell* 65:677
- Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR (1996) AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* 84:321
- Shivdasani RA, Mayer EL, Orkin SH (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncogene tal-1/SCL. *Nature* 373:432
- Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH (1996) The T cell leukemia oncogene SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* 86:47
- Bain G, Robanus-Maandag E, te Riele HPJ, Izon D, Amsen D, Kruisbeek A, Krop I, Schlissel MS, Feeney AJ, van Roon M, van der Valk M, Berns A, et al (1994) E2A proteins are required for proper B cell development and initiations of immunoglobulin gene rearrangement. *Cell* 79:885
- Urbanek P, Wang Z-Q, Fetka I, Wagner E, Busslinger M (1994) Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* 79:901
- Busslinger M, Urbanek P (1995) The role of BSAP (Pax-5) in B-cell development. *Curr Opin Genet Dev* 5:595
- Zhang D-E, Zhang P, Wang N-D, Hetherington CJ, Darlington GJ, Tenen DG (1997) Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α -deficient mice. *Proc Natl Acad Sci U S A* 94:569
- Scott EW, Simon MC, Anastasi J, Singh H (1994) Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 265:1573
- Shivdasani RA, Rosenblatt MF, Franklin DZ, Jackson CW, Hunt P, Saris CJM, Orkin SH (1995) Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* 81:695
- Pandolfi PP, Roth ME, Karis A, Leonard MW, Dzierzak E, Grosveld FG, Douglas E, Lindenbaum MH (1995) Targeted disruption of the *GATA-3* gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nature Genet* 11:40
- Tsai F-Y, Keller G, Kuo FC, Weiss M, Chen J, Rosenblatt M, Alt FW, Orkin SH (1994) An early haematopoietic defect in mice lacking the transcription factor *GATA-2*. *Nature* 371:221
- Georgopoulos K, Bigby M, Wang J-H, Molnar A, Wu P, Winandy S, Sharpe A (1994) The *Ikaros* gene is required for the development of all lymphoid lineages. *Cell* 79:143
- Warren AJ, Colledge WH, Carlton MBL, Evans MJ, Smith AJH, Rabbits TH (1994) The oncogenic cysteine-rich LIM domain protein Rbn2 is essential for erythroid development. *Cell* 78:45
- Krishnaraju K, Nguyen HQ, Liebermann DA, Hoffman B (1995) The zinc finger transcription factor *Egr-1* potentiates macrophage differentiation of hematopoietic cells. *Mol Cell Biol* 15:5499
- Morris JF, Rauscher FJ III, Davis B, Klemsz M, Zu D, Tenen D, Hromas R (1995) The myeloid zinc finger gene, *MZF-1*, regulates the *CD34* promoter in vitro. *Blood* 86:3640
- Perrotti D, Melotti P, Skorski T, Casella I, Peschle C, Calabretta B (1995) Overexpression of the zinc finger protein *MZF1* inhibits hematopoietic development from embryonic stem cells: correlation with negative regulation of *CD34* and *c-myb* promoter activity. *Mol Cell Biol* 15:6075
- Small D, Levenstein M, Kim E, Carow C, Amin S, Rockwell P, Witte L, Burrow C, Ratajczak M, Gewirtz AM, Civin C (1994) STK-1, the human homolog of Flk-2/FLT-3, is selectively expressed in *CD34*+ human bone marrow cells and is involved in the proliferation of early progenitor/stem cells. *Proc Natl Acad Sci U S A* 91:459
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1993) Current protocols in molecular biology. New York: John Wiley & Sons, Inc
- Alschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. New York: Cold Spring Harbor Laboratory Press
- Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonucleases fragments to high specific activity. *Anal Biochem* 132:6
- Bhatt B, Burns J, Flannery D, McGee J (1988) Direct visualization of single copy genes on banded metaphase chromosomes by non-isotopic in situ hybridization. *Nucleic Acids Res* 16:3951
- Lichter P, Tang C, Call K, Hermanson G, Evans G, Houseman D, Ward D (1990) High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 247:64
- Gore SD, Kastan MB, Civin CI (1991) Normal human bone marrow precursors that express terminal deoxynucleotidyl transferase include T-cell precursors and possible lymphoid stem cells. *Blood* 77:1681
- Horiguchi-Yamada J, Yamada H (1993) Differing responses of G2-related genes during differentiation of HL60 cells induced by TPA or DMSO. *Mol Cell Biochem* 119:29
- Horiguchi-Yamada J, Yamada H, Nakada S, Ochi K, Nemoto T (1994) Changes of G1 cyclins, cdk2, and cyclin A during the differentiation of HL60 cells induced by TPA. *Mol Cell Biochem* 132:31
- Kozopas KM, Buchan HL, Craig RW (1990) Improved coupling between proliferation-arrest and differentiation-induction in ML-1 human myeloblastic leukemia cells. *J Cell Physiol* 145:575
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium. *Anal Biochem* 162:156
- Suzow J, Friedman AD (1993) The murine myeloperoxidase promoter contains several functional elements—one element binds a cell type-restricted transcription factor, myeloid nuclear factor 1 (MyNF1). *Mol Cell Biol* 13:2141
- Knochel W, Potting A, Koster M, el Baradi T, Niefeld W, Bouwmeester TP, Pieler T (1989) Evolutionary conserved modules associated with zinc fingers in *zenopus laevis*. *Proc Natl Acad Sci U S A* 86:6097
- Rosati M, Marino M, Franze A, Tramontano A, Grimaldi G (1991) Members of the zinc finger protein gene family sharing a conserved N-terminal module. *Nucleic Acids Res* 19:5661
- Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA (1991) The evolutionary conserved kruppel-associated box domain defines a subfamily of eukaryotic multi-fingered proteins. *Proc Natl Acad Sci U S A* 88:3608
- Chavrier P, Lemaire P, Revelant O, Bravo R, Charnay P (1988) Characterization of a mouse multigene family that encodes zinc finger structures. *Mol Cell Biol* 8:1319
- Pinna LA (1990) Casein kinase 2: an "eminence grise" in cellular recognition? *Biochim Biophys Acta* 1054:267

38. Kennelly PJ, Krebs EG (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J Biol Chem* 266:15555
39. Brawerman G (1990) Mechanisms of mRNA decay. *Trends Biotechnol* 8:171
40. Rogers S, Wells R, Reichsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234:364
41. Tommerup N, Vissing H (1995) Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* 27:259
42. Witzgall R, O'Leary E, Gessner R, Ouellette AJ, Bonventre JV (1993) *Kid-1*, a putative renal transcription factor: regulation during ontogeny and in response to ischemia and toxic injury. *Mol Cell Biol* 13:1933
43. Rosnet O, Marchetto S, DeLapeyriere O, Brinbaum D (1991) Murine FLT3, a gene encoding a novel tyrosine kinase receptor of the PDGFR/CSF1R family. *Oncogene* 6:1641
44. Matthews W, Jordan CT, Wiegand GW, Pardoll D, Lemischka IR (1991) A receptor tyrosine kinase specific to hematopoietic stem and progenitor cell-enriched populations. *Cell* 65:1143
45. Gotze KS, Ramirez M, Tabor K, Small D, Matthews W, Civin CI (1997) Flt3^{high} and Flt3^{low} CD34⁺ progenitor cells isolated from human bone marrow are functionally distinct. *Blood* 91:1947
46. Lyman S, James L, Zappone J, Sleath P, Beckmann M, Bird T (1993) Characterization of the protein encoded by the Flt3 (Flk2) receptor-like tyrosine kinase gene. *Oncogene* 8:815
47. Lyman SD, James L, Johnson L, Brasel K, de Vries P, Escobar SS, Downey H, Splett RR, Beckmann MP, McKenna HJ (1994) Cloning of the human homologue of the murine Flt3 ligand: a growth factor for early hematopoietic progenitor cells. *Blood* 83:2795
48. Lyman SD, Brasel K, Rousseau AM, Williams DE (1994) The Flt3 ligand: a hematopoietic stem cell factor whose activities are distinct from steel factor. *Stem Cells* 12:99
49. Lyman SD (1995) Biology of Flt3 ligand and receptor. *Int J Hematol* 62:63
50. Hannum C, Culpepper J, Campbell D, McClanahan T, Zurawski S, Bazan JF, Kastelein R, Hudak S, Wagner J, Mattson J, Luh J, Duda G, et al (1994) Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of haematopoietic stem cells and is encoded by variant RNAs. *Nature* 368:643
51. Kok K, Osinga J, Carritt B, Davis MB, van der Hout AH, van der Veen AY, Landsvater RM, de Leij LF, Berendsen HH, Postmus PE, Poppema S, Buys CHCM (1987) Deletion of a DNA sequence at the chromosomal region 3p21 in all major types of lung cancer. *Nature* 330:578
52. Naylor SL, Johnson BE, Minna JD, Sakaguchi AY (1987) Loss of heterozygosity of chromosome 3p markers in small-cell lung cancer. *Nature* 329:451
53. Brauch H, Johnson B, Hovis J, Yano T, Gazdar A, Pettengill OS, Graziaano S, Soprenson GD, Poiesz BJ, Binna J, Linehan M, Zbar B (1987) Molecular analysis of the short arm of chromosome 3 in small-cell and non-small cell carcinoma of the lung. *N Engl J Med* 317:1109
54. Zbar B, Braunch H, Talmadge C, Linehan M (1987) Loss of alleles of loci on the short arm of chromosome 3 in renal cell carcinoma. *Nature* 327:721
55. Devilee P, van den Broek M, Kulpers-Dijkshoorn M, Kolluri R, Khan PM, Pearson PL, Cornelisse CJ (1989) At least four different chromosomal regions are involved in loss of heterozygosity in human breast carcinoma. *Genomics* 5:554
56. el-Naggar AK, Lee MS, Wang G, Luna MA, Goepfert H, Batsakis JG (1993) Polymerase chain reaction-based restriction fragment length polymorphism analysis of the short arm of chromosome 3 in primary head and neck squamous carcinoma. *Cancer* 72:81
57. Huret JL, Tanzer J, Henry-Amar M (1986) Aberrant breakpoints in chronic myelogenous leukaemia: oncogenes and fragile sites. *Hum Genet* 74:447
58. Ohyashiki K, Ohyashiki JH, Kinniburgh AJ, Han T, Sandberg AA (1986) Recurrent chromosome changes at 3p21 in Epstein-Barr virus-transformed cells derived from human polymorphous leukemia. *Cancer Res* 46:5282
59. Ohyashiki K, Ohyashiki JH, Takeuchi J, Han T, Henderson ES, Sandberg AA (1987) Cytogenetic studies in hairy cell leukemia. *Cancer Genet Cytogenet* 24:109
60. Shi G, Weh HJ, Martensen S, Seeger D, Hossfeld DK (1996) 3p21 is a recurrent treatment-related breakpoint in myelodysplastic syndrome and acute myeloid leukemia. *Cytogenet Cell Genet* 74:295
61. Johansson B, Billstrom R, Kristofferson U, Akerman M, Garwicz S, Ahlgren T, Malm C, Mitelman F (1997) Deletion of chromosome arm 3p in hematologic malignancies. *Leukemia* 11:1207
62. Bellefroid EJ, Lecocq PJ, Benhida A, Poncelet DA, Belayew A, Martial JA (1989) The human genome contains hundreds of genes coding for finger proteins of the kruppel type. *DNA* 8:377
63. Pellegrino GR, Berg JM (1991) Identification and characterization of "zinc-finger" domains by the polymerase chain reaction. *Proc Natl Acad Sci U S A* 88:671
64. Margolin JF, Friedman JR, Meyer WK, Vissing H, Theisen HJ, Rauscher FJ (1994) Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci USA* 91:4509
65. Pengue G, Calabro V, Bartoli PC, Pagliuca A, Lania L (1994) Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res* 22:2908
66. Vissing H, Meyer WK-H, Aagaard L, Tommerup N, Thiesen H-J (1995) Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Lett* 369:153
67. Witzgall R, O'Leary E, Leaf A, Onaldi D, Bonventre JV (1994) The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc Natl Acad Sci U S A* 91:4514
68. Pengue G, Lania L (1996) Kruppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements. *Proc Natl Acad Sci U S A* 93:1015
69. Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang X-P, Neilson EG, Rauscher FJ III (1996) KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev* 10:2067
70. Crossley M, Orkin SH (1994) Phosphorylation of the erythroid transcription factor GATA-1. *J Biol Chem* 269:16589
71. Klocke B, Knochel W (1995) Proteins of the xenopus laevis zinc finger multigene family as targets for CK II phosphorylation. *Mol Cell Biochem* 142:49
72. Towatari M, May GE, Marais R, Perkins GR, Marshall CJ, Cowley S, Enver T (1995) Regulation of GATA-2 phosphorylation by mitogen-activated protein kinase and interleukin-3. *J Biol Chem* 270:4101
73. Murai K, Murakami H, Nagata S (1997) A novel form of the myeloid-specific zinc-finger protein (MZF-2). *Genes Cells* 2:581
74. Murai K, Murakami H, Nagata S (1998) Myeloid-specific transcriptional activation by murine myeloid zinc-finger protein 2. *Proc Natl Acad Sci U S A* 95:3461