MOLECULAR AND CELLULAR BIOLOGY, Nov. 1991, p. 5462–5469
0270-7306/91/115462-08$02.00/0

Vol. 11, No. 11

Structural Characterization of *SIL*, a Gene Frequently Disrupted in T-Cell Acute Lymphoblastic Leukemia

PETER D. APLAN,[1,2] DONALD P. LOMBARDI,[1] AND ILAN R. KIRSCH[1]*

Naval Medical[1] and Pediatric Oncology[2] Branches, National Cancer Institute, Bethesda, Maryland 20889-5105

Received 3 June 1991/Accepted 7 August 1991

The *SIL* (*SCL* interrupting locus) gene was initially discovered at the site of a genomic rearrangement in a T-cell acute lymphoblastic leukemia cell line. This rearrangement, which occurs in a remarkably site-specific fashion, is present in the leukemic cells of 16 to 26% of patients with T-cell acute lymphoblastic leukemia. We have now cloned a normal *SIL* cDNA from a cell line which does not carry the rearrangement. The *SIL* cDNA has a long open reading frame of 1,287 amino acids, with a predicted molecular size of 143 kDa. The predicted protein is not homologous with any previously described protein; however, a potential eukaryotic topoisomerase I active site was identified. Cross-species hybridization using a *SIL* cDNA probe indicated that the *SIL* gene was conserved in mammals. A survey of human and murine cell lines and tissues demonstrated *SIL* mRNA to be ubiquitously expressed, at low levels, in hematopoietic cell lines and tissues. With the exception of 11.5-day-old mouse embryos, *SIL* mRNA was not detected in nonhematopoietic tissues. The genomic structure of *SIL* was also analyzed. The gene consists of 18 exons distributed over 70 kb, with the 5′ portion of the gene demonstrating alternate exon utilization.

Nonrandom chromosomal translocations or deletions present in specific malignancies have been recognized for the past 30 years, since the association of the Ph chromosome with chronic myelogenous leukemia (20). The analysis of these recurring chromosomal abnormalities at a molecular level has identified numerous proto-oncogenes and growth-affecting genes (for a review, see reference 23). These rearrangements often occur at the sites of transcriptionally active DNA, which may be in a more open chromatin configuration and thus more susceptible to chromosomal breakage and rejoining (17). Chromosomal breakpoints often involve genes that are important for the growth or development of the cell that undergoes the translocation; the classic example is seen in Burkitt's lymphoma, in which the c-*myc* and immunoglobulin genes are disrupted and brought into chromosomal contiguity (18).

Traditionally, these chromosomal abnormalities have been identified cytogenetically on preparations of metaphase chromosomes. Recently, while investigating the genomic structure of the newly described *SCL* (*TCL5*, *tal-1*) gene (6, 7, 9, 13), a member of the basic helix-loop-helix family of transcription factors, we identified a frequent, site-specific chromosomal deletion that involved *SCL* and a previously unidentified locus that we called *SIL* (for *SCL* interrupting locus) (4). This interstitial deletion is the first known instance whereby two genes, neither one of which is an antigen receptor gene, are joined through the action of the V(D)J recombinase system. The deletion occurs below the level of conventional cytogenetic detection and leads to a fusion mRNA between *SIL* and *SCL*. A 5.5-kb *SIL* transcript, distinct from *SCL*, was identified in normal tissues by Northern (RNA) blot hybridization (4). To better understand how disruption of the *SIL* gene may be relevant to the development of these T-cell leukemias and what its normal function might be, we proceeded to clone the *SIL* cDNA, determine its genomic structure, and examine its spectrum of expression.

* Corresponding author.

MATERIALS AND METHODS

DNA and RNA isolation and analysis. DNA and RNA were isolated from cell lines and tissues by the guanidinium isothiocyanate method (11). DNA samples (10 μg) were digested to completion with restriction endonucleases as recommended by the supplier (Bethesda Research Laboratories), size fractionated on 0.8% agarose gels, and transferred to nitrocellulose membranes by the Southern method (25). RNA samples [10 μg of total RNA or 2 μg of poly(A) RNA] were size fractionated on a 1.0% agarose–formaldehyde gel and transferred to nitrocellulose membranes (11). Hybridizations were performed with nick-translated ^{32}P-labeled probes and washed under standard conditions (6), with the highest-stringency washes consisting of 0.1% sodium dodecyl sulfate and 0.1× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 52°C. Low-stringency hybridizations, using a human probe on nonhuman Southern and Northern blots, were performed in a similar fashion, with the highest stringency wash at 42 instead of 52°C. mRNA half-life was determined by incubating K562 cells in the presence of actinomycin D (50 μg/ml; Sigma Chemical Co.) and harvesting the cells at 0.5, 1.0, 2.0, 4.0, and 8.0 h. RNA was extracted, and 10 μg from each time point was run on a 1.0% agarose–formaldehyde gel, transferred to a nitrocellulose membrane, and hybridized to a *SIL* probe. mRNA half-life was determined by measuring the decline in *SIL* signal on the autoradiograph.

Genomic and cDNA cloning. cDNA phage clones were obtained from a human bone marrow cDNA library and a cDNA library constructed from the SUPT1 T-cell lymphoma cell line (15). Both libraries were constructed in lambda ZAP II (Stratagene) as previously described (3). Genomic phage clones were obtained by screening a human placental genomic library constructed in lambda FIX II (Stratagene). Relevant cDNA and genomic restriction fragments were subcloned into the pGem7zf (Promega) and pBluescript (Stratagene) plasmid vectors.

Sequence analysis. Plasmid inserts were sequenced on both strands of DNA by the dideoxy-chain termination method, using Sequenase polymerase (U.S. Biochemical) and protocols. Both nested deletion mutants (Erase-A-Base system; Promega) and synthetic oligonucleotide primers were used. An IBM PS/2 with PC-Gene (Intelligenetics) was used for sequence analysis and manipulation. Genomic and cDNA sequences were compared with those in the GenBank data base. Protein sequences were compared with those in the SWISSPROT data base.

RNase protection assay. Relevant genomic or cDNA restriction fragments were subcloned into plasmids, and uniformly ^{32}P labeled antisense RNA was synthesized by using T7, T3, or SP6 RNA polymerase and Gemini riboprobe reagents (Promega). The radiolabeled antisense RNA (2×10^5 cpm) was hybridized to 30 μg of sample RNA for 12 to 16 h at 50°C. The samples were then subjected to RNase digestion and size fractionated on a 6% acrylamide–7 M urea denaturing gel (22).

Oligonucleotide synthesis. Oligonucleotides for sequencing and polymerase chain reaction (PCR) amplification were synthesized with an Applied Biosystems DNA synthesizer (model 380B) and used without further purification.

RACE PCR cloning of the *SIL* 5′ end. A modification of the RACE (rapid amplification of cDNA ends) technique (14) was used to clone the 5′ end of *SIL*. Total cellular RNA (10 μg) from the T-cell line Jurkat (16) or HSB-2 (1) was annealed to 10 ng of a *SIL* exon 5 antisense oligonucleotide, 5′-AGTCGGATGGTCTTCTCAGT-3′, by heating the mixture to 65°C for 5 min and then chilling it on ice. Moloney murine leukemia virus reverse transcriptase (200 U; Bethesda Research Laboratories) was used for first-strand cDNA synthesis, according to the supplier's recommended protocol, at 45°C for 1 h. First-strand cDNA was separated from the reaction by using a Sephacryl 300 spun column (Pharmacia) as recommended by the manufacturer. An oligo(dA) tail was added to the 3′ end of the cDNA by incubating 15 μl of the recovered first-strand cDNA with 0.5 μl of 10 mM dATP, 4 μl of 5× tailing buffer, and 15 U of terminal deoxynucleotidyltransferase for 5 min at 37°C and 5 min at 65°C. The tailed first-strand cDNA was then diluted to 200 μl, and 1 to 3 μl was used for PCR amplification. Amplification was performed by using a *SIL* exon 4 antisense oligonucleotide, 5′-CTGTAGTAACTGAGATGTA-3′, and a universal 5′ oligonucleotide, 5′-GACTCGAGTCG ACATCGAT$_{17}$-3′, containing *Sal*I, *Cla*I, and *Xho*I restriction sites. Thermal cycling was carried out for 35 cycles of 95°C for 45 s, 48°C for 45 s, and 72°C for 2 min. The PCR products were then extracted with phenol-chloroform, digested with *Kpn*I and *Cla*I, and ligated into pBluescript II.

Nucleotide sequence accession number. The nucleotide sequence of the composite *SIL* cDNA (see Fig. 2) has been deposited in GenBank under accession number M74558.

RESULTS

***SIL* cDNA cloning.** Previous experiments demonstrated that the normal *SIL* transcript unit was 5.5 kb (4). One million recombinant phage clones from a human bone marrow cDNA library were screened with a probe containing a portion of the *SIL* transcript unit as previously described (4). A single hybridizing clone (clone 31.4; Fig. 1) was purified, subcloned, and sequenced. RNase protection assays using the insert from clone 31.4 as a probe did not protect the 5′ portion of the clone in any test RNA (data not shown), suggesting that the 5′ portion of this single clone most likely was a cloning artifact. A nonreiterated fragment (1.1KR; Fig. 1) derived from a portion of the bone marrow cDNA clone was then used as a probe to screen 10^6 recombinant clones from an unamplified SUPT1 (a T-cell lymphoma cell line [15]) cDNA library. Three hybridizing clones were purified and subcloned. All had inserts of approximately 5 kb, and sequence analysis along with extensive restriction mapping indicated that the three clones were all quite similar. The 5′ ends of all three clones lay within 12 nucleotides of one another, while the 3′ ends were all within 26 nucleotides of one another. The longest clone (7S; Fig. 1) was sequenced on both strands of DNA in its entirety; selected portions of the other two clones were sequenced. RNase protection transcript mapping experiments (described below) indicated that none of the cDNA clones were full length; therefore, the RACE technique (14) was used to generate clones, such as clone 96 (Fig. 1), extending further 5′. The sequence and predicted protein(s) of a composite *SIL* cDNA are presented in Fig. 2; alternate splicing events are described in the figure legend. An ATG triplet in good context for protein initiation (ATCATGG) is seen at nucleotide 366; this ATG is preceded by stop codons in all three reading frames. There is a long open reading frame of 3,861 nucleotides, potentially encoding a protein of 1,287 amino acids, with a predicted molecular size of 143 kDa.

RNase protection analysis and the RACE cloning strategy (14) were used to determine the *SIL* transcript initiation site. A total of 18 independent RACE clones from HSB-2 and Jurkat mRNAs were sequenced. The majority of these clones ended within 1 nucleotide of nucleotide 1 in Fig. 2, four additional clones ended at a second cluster 28 nucleotides upstream, and the remainder were scattered within 55 nucleotides of nucleotide 1. To confirm that nucleotide 1 represented the major transcript initiation site, a 0.4-kb genomic fragment encompassing exon 1 was isolated and sequenced (Fig. 3A). When this genomic fragment was used as an RNase protection probe, a 112-bp fragment was protected (Fig. 3B) in all mRNA samples tested. This 112-bp protected fragment corresponds precisely to nucleotide 1 as defined by the RACE clones. The sequence surrounding this nucleotide, CTCAGTTCC, is in good agreement with the consensus transcription start sequence of PyPyC(A/G)PyPy PyPy (10). A very similar sequence, TTCAGTTTC, is found 28 nucleotides upstream, where four additional clones ended. Several CCAAT boxes and a potential Sp1 binding site are indicated in Fig. 3B.

None of the *SIL* cDNA clones contained a poly(A) tail. However, since the clones were obtained from an oligo(dT)-

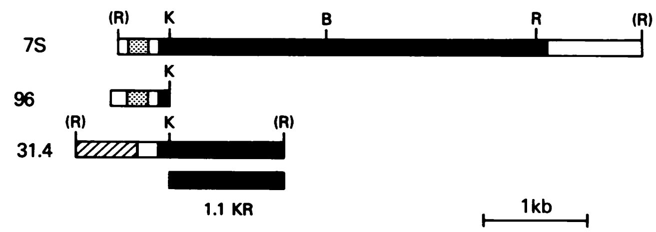


FIG. 1. Restriction map of three representative *SIL* cDNA clones. Clone 7S is from a SUPT1 cDNA library, clone 31.4 is from a human bone marrow library, and clone 96 was generated from Jurkat mRNA by the RACE protocol. Coding sequence is indicated by solid bars; 5′ and 3′ untranslated regions are represented by open bars. The stippled segment represents an alternatively spliced exon (exon 2; see below). The cross-hatched region of clone 31.4 represents non-*SIL* sequence found only in a single clone from the human bone marrow library and most likely represents a cloning artifact. Restriction sites: B, *Bam*HI; K, *Kpn*I; R, *Eco*RI. (R) indicates *Eco*RI sites introduced in the cloning process.

```
   1 AGTTCCCGCGACCCCAACGTCCCAGAGGCGGGGCCGGAGTCGGCGGTGGCGCTCCTTGGAGCCGGCTCCCGC
     TCCTACCCTGCAAACAGACCTCAGCTCCGCGGAAGTTGCGAGACGGGGTTTCACCATGTTGGTCGGGCTGGT
     CTGGAAATCCTGACTTCAGGTGATCCACCCGCCTCGGCCTCCCAAAATGCTGGGATTACAAGCGTGAGCCAC
     CGCCCCTGACATGAGCCATTGACTTTTAAAGCAGGAGAATAATTTGGATCAGATTTATATGGAAACACTCTT
     CTAGCAGCATTATGGGGACTTTTCCATAAGTCTGGATACTGAGGATTTGGAATTAAAGAAATCATTCACCAG
 361 ACATC
 366 ATGGAGCCTATATATCCTTTTGCACGGCCCCAGATGAATACCAGGTTTCCTTCAAGCAGGATGGTACCTTTC
   1  M  E  P  I  Y  P  F  A  R  P  Q  M  N  T  R  F  P  S  S  R  M  V  P  F
     CACTTTCCTCCATCAAAATGTGCACTTTGGAACCCAACGCCAACTGGAGATTTCATCTACTTACATCTCAGT
  25  H  F  P  P  S  K  C  A  L  W  N  P  T  P  T  G  D  F  I  Y  L  H  L  S
     TACTACAGAAATCCAAAGCTTGTGGTGACTGAGAAGACCATCCGACTTGCTTATCGTCATGCTAACGAGAAT
  49  Y  Y  R  N  P  K  L  V  V  T  E  K  T  I  R  L  A  Y  R  H  A  N  E  N
     AAAAAAAATTCGTCATGCTTTTTACTTGGTTCTCTGACAGCAGACGAAGATGAAGAAGGTGTAACATTGACA
  73  K  K  N  S  S  C  F  L  L  G  S  L  T  A  D  E  D  E  E  G  V  T  L  T
     GTAGATCGCTTTGATCCTGGTCGAGAAGTACCTGAATGCCTAGAAATAACCCCTACTGCTTCTCTTCCTGGG
  97  V  D  R  F  D  P  G  R  E  V  P  E  C  L  E  I  T  P  T  A  S  L  P  G
 726 GACTTTTTGATTCCATGCAAAGTTCATACTCAAGAACTTTGTTCAAGAGAAATGATAGTTCACAGTGTAGAT
 121  D  F  L  I  P  C  K  V  H  T  Q  E  L  C  S  R  E  M  I  V  H  S  V  D
     GACTTCAGTTCAGCTTTAAAGGCTCTACAGTGCCATATATGTAGCAAAGATTCCTTGGACTGTGGTAAGCTG
 145  D  F  S  S  A  L  K  A  L  Q  C  H  I  C  S  K  D  S  L  D  C  G  K  L
     CTTTCCCTAAGAGTTCATATCACTTCCAGGGAGAGTTTGGACAGTGTGGAATTTGACTTGCATTGGGCAGCA
 169  L  S  L  R  V  H  I  T  S  R  E  S  L  D  S  V  E  F  D  L  H  W  A  A
     GTAACTCTAGCAAATAACTTTAAATGCACACCTGTGAAGCCCATCCCCATTATTCCAACAGCTCTGGCAAGA
 193  V  T  L  A  N  N  F  K  C  T  P  V  K  P  I  P  I  I  P  T  A  L  A  R
     AACTTGAGCAGTAATCTGAATATTTCTCAAGTTCAAGGGACTTATAAATATGGATATCTTACCATGGATGAA
 217  N  L  S  S  N  L  N  I  S  Q  V  Q  G  T  Y  K  Y  G  Y  L  T  M  D  E
1086 ACACGCAAATTGTTACTTTTGTTGGAATCTGATCCCAAGGTTTATTCTCTACCATTGGTGGGAATTTGGCTG
 241  T  R  K  L  L  L  L  L  E  S  D  P  K  V  Y  S  L  P  L  V  G  I  W  L
     TCTGGAATTACACATATCTATAGTCCTCAGGTATGGGCTTGCTGTTTGCGATACATATTCAATTCTTCTGTT
 265  S  G  I  T  H  I  Y  S  P  Q  V  W  A  C  C  L  R  Y  I  F  N  S  S  V
     CAAGAAAGGGTTTTTTCAGAATCTGGAAATTTCATCATAGTTCTCTATTCTATGACACATAAGGAACCTGAG
 289  Q  E  R  V  F  S  E  S  G  N  F  I  I  V  L  Y  S  M  T  H  K  E  P  E
     TTTTATGAATGCTTCCCTTGTGATGGCAAGATACCTGACTTTCGGTTTCAGTTGCTAACCAGTAAGGAAACA
 313  F  Y  E  C  F  P  C  D  G  K  I  P  D  F  R  F  Q  L  L  T  S  K  E  T
     TTACATCTTTTCAAAAATGTTGAACCTCCTGACAAAAATCCAATCCGTTGTGAACTGAGCGCTGAAAGCCAA
 337  L  H  L  F  K  N  V  E  P  P  D  K  N  P  I  R  C  E  L  S  A  E  S  Q
1446 AATGCAGAAACAGAGTTTTTCAGTAAGGCTTCCAAGAATTTTTCAATTAAGAGGTCTTCCCAAAAGTTATCT
 361  N  A  E  T  E  F  F  S  K  A  S  K  N  F  S  I  K  R  S  S  Q  K  L  S
     TCTGGGAAGATGCCAATACATGATCACGACTCTGGTGTTGAAGATGAAGATTTTTCTCCAAGACCAATTCCT
 385  S  G  K  M  P  I  H  D  H  D  S  G  V  E  D  E  D  F  S  P  R  P  I  P
     AGTCCTCATCCAGTGAGTCAGAAGATTTCTAAGATCCAACCATCAGTTCCTGAACTTTCACTTGTGTTGGAT
 409  S  P  H  P  V  S  Q  K  I  S  K  I  Q  P  S  V  P  E  L  S  L  V  L  D
     GGCAATTTCATAGAATCAAACCCTCTGCCTACTCCATTGGAAATGGTGAATAATGAAAATCCTCCTTTGATT
 433  G  N  F  I  E  S  N  P  L  P  T  P  L  E  M  V  N  N  E  N  P  P  L  I
```

FIG. 2. Composite nucleotide sequence and predicted protein of the *SIL* cDNA. Nucleotide 1 represents the 5′ extent of the *SIL* cDNA, as defined by RACE cloning and RNase protection (see text). The initiation ATG codon, at nucleotide 366, is preceded by stop codons (underlined) in all three reading frames. Splice junctions are indicated by brackets. Exon 2 (nucleotides 113 to 322) is spliced out of some mRNA species. The boxed region of exon 7 (nucleotides 862 to 1002) is spliced out of a minority of mRNA species (see text) and is bounded by GT splice donor and AG splice acceptor sequences. A single cDNA clone extended exon 5 133 nucleotides beyond the usual splice donor site into the contiguous intron and subsequently spliced at the normal exon 6 splice acceptor site. This exon, containing all of exon 5 and a portion of the adjacent intron, is referred to as exon 5a. The protein predicted by this cDNA clone is referred to as *SIL* form C (see text).

primed cDNA library, and three independent clones began within 26 nucleotides of one another, it seemed likely that these clones were near the 3′ end. When a genomic clone encompassing exon 18 (the terminal *SIL* exon; see below) was sequenced and compared with the longest *SIL* cDNA clone, an AATAAA polyadenylation signal was noted 11 nucleotides 3′ of the end of the cDNA clone. These data indicate that the 3′ end of a full-length *SIL* cDNA clone is, in all likelihood, 20 to 30 nucleotides 3′ of this polyadenylation signal.

***SIL* mRNA expression.** Northern blot analysis of mRNA extracted from a variety of human cell lines and tissues indicated that the 5.5-kb *SIL* transcript was expressed ubiquitously, at low levels, in all hematopoietic cell lines and hematopoietic tissues studied (Table 1). The half-life of the *SIL* message, based on actinomycin D inhibition experiments, was 2 h in the K562 cell line. *SIL* message was not detected in any other human tissue studied. To extend this spectrum of expression, a variety of murine tissues were studied by Northern blotting. A human *SIL* cDNA probe (1.1KR; Fig. 1) identified a single 5.5-kb transcript in several murine cell lines, including uninduced murine erythroleukemia, F9 teratocarcinoma, and D3 embryonic stem cells. Table 1 demonstrates that the only murine tissues (of those studied) expressing *SIL* mRNA were thymus and 11.5-day whole mouse embryos.

***SIL* genomic structure.** To determine the exon/intron structure of the *SIL* gene, a human placental genomic library was screened with several *SIL* cDNA probes. The overlapping genomic phage clones were compared with the *SIL* cDNA, and the exon/intron structure is shown in Fig. 4. The exon/intron borders were sequenced; all splice acceptor sequences contained the AG dinucleotide sequence. However, a single splice donor had a GC dinucleotide instead of the usual GT splice donor sequence. The nucleotide sequence surrounding this splice donor is AAGgcaagt, which is in very good agreement with the consensus sequence of (C/A)AGgt(a/g)agt except for the C-for-T replacement. Two splice donor sequences that violate the GT rule have been reported before (21), the variant dinucleotide being GC in both instances.

Examination of the genomic structure of both *SIL* and *SCL* (3) indicates that the previously reported *SIL/SCL* rearrangement (4) juxtaposes *SIL* intron 1 with *SCL* intron 1. This leads to formation of a fusion *SIL/SCL* mRNA, with *SIL* exon 1 splicing to *SCL* exon 3, in a head-to-tail fashion

```
      AACCACTTGGAACACTTGAAGCCATTGCAACCCCAGCTTTATGATGAGAAACACAGTCCAGAAGTTGAAGCT
457    N  H  L  E  H  L  K  P  L  Q  P  Q  L  Y  D  E  K  H  S  P  E  V  E  A
 1806 GGAGAGCCTTCCTTGAGAGGAATACCAAATCAGTTAAACCAGGATAAACCAGCTCTTTTGAGACACTGCAAA
481    G  E  P  S  L  R  G  I  P  N  Q  L  N  Q  D  K  P  A  L  L  R  H  C  K
      GTAAGACAGCCACCTGCCTATAAGAAAGGGAACCCCCATACCAGGAACAGTATTAAACCATCTTCTCATAAT
505    V  R  Q  P  P  A  Y  K  K  G  N  P  H  T  R  N  S  I  K  P  S  S  H  N
      GGGCCATCTCATGATATATTTGAAAAGCTCCAAACAGTTTCTGCTGGAAATGTACAAAACGAAGAGTATCCT
529    G  P  S  H  D  I  F  E  K  L  Q  T  V  S  A  G  N  V  Q  N  E  E  Y  P
      ATAAGACCCTCCACACTTAATTCTAGGCAGTCTTCTCTTGCCCCGCAGTCCCAACCACACGATTTTGTTTTT
553    I  R  P  S  T  L  N  S  R  Q  S  S  L  A  P  Q  S  Q  P  H  D  F  V  F
      TCACCCCATAATTCAGGAAGACCAATGGAACTTCAGATACCTACTCCCCCACTGCCATCTTACTGTTCCACA
577    S  P  H  N  S  G  R  P  M  E  L  Q  I  P  T  P  P  L  P  S  Y  C  S  T
 2166 AACGTTTGCAGGTGTTGTCAGCATCATAGTCATATTCAATATAGTCCGCTAAATTCTTGGCAAGGAGCAAAC
601    N  V  C  R  C  C  Q  H  H  S  H  I  Q  Y  S  P  L  N  S  W  Q  G  A  N
      ACAGTTGGATCCATTCAAGATGTCCAGTCTGAAGCCCTTCAAAAGCATTCATTATTTCACCCAAGTGGATGT
625    T  V  G  S  I  Q  D  V  Q  S  E  A  L  Q  K  H  S  L  F  H  P  S  G  C
      CCAGCCCTGTACTGTAATGCATTCTGTTCTTCAAGTAGTCCTATAGCCTTGAGACCTCAGGGAGATATGGGC
649    P  A  L  Y  C  N  A  F  C  S  S  S  S  P  I  A  L  R  P  Q  G  D  M  G
      AGTTGTTCTCCCCACAGCAATATTGAACCATCGCCTGTGGCAAGACCGCCTTCACATATGGACTTATGTAAC
673    S  C  S  P  H  S  N  I  E  P  S  P  V  A  R  P  P  S  H  M  D  L  C  N
      CCACAGCCTTGCACAGTGTGCATGCACACACCCAAGACTGAGTCAGATAATGGAATGATGGGACTATCTCCA
697    P  Q  P  C  T  V  C  M  H  T  P  K  T  E  S  D  N  G  M  M  G  L  S  P
 2526 GATGCATATCGGTTCCTCACAGAACAAGACAGACAGCTAAGACTACTTCAGGCACAGATTCAGCGTTTGTTG
721    D  A  Y  R  F  L  T  E  Q  D  R  Q  L  R  L  L  Q  A  Q  I  Q  R  L  L
      GAAGCACAGTCTCTGATGCCCTGTTCCCCTAAGACAACTGCTGTTGAAGACACAGTGCAAGCTGGAAGACAA
745    E  A  Q  S  L  M  P  C  S  P  K  T  T  A  V  E  D  T  V  Q  A  G  R  Q
      ATGGAGTTGGTTTCTGTGGAAGCACAGTCTTCCCCTGGCCTGCACATGAGAAAAGGTGTAAGCATTGCTGTG
769    M  E  L  V  S  V  E  A  Q  S  S  P  G  L  H  M  R  K  G  V  S  I  A  V
      AGCACAGGTGCTAGCTTGTTTTGGAATGCAGCAGGTGAGGATCAAGAGCCTGACTCTCAAATGAAGCAAGAT
793    S  T  G  A  S  L  F  W  N  A  A  G  E  D  Q  E  P  D  S  Q  M  K  Q  D
      GATACCAAAATTTCCAGTGAGGACATGAATTTTTCTGTCGATATTAATAATGAAGTCACAAGTCTTCCAGGT
817    D  T  K  I  S  S  E  D  M  N  F  S  V  D  I  N  N  E  V  T  S  L  P  G
 2886 AGTGCATCTTCATTAAAAGCAGTTGATATTCCCAGTTTTGAAGAGAGCAACATTGCTGTGGAAGAAGAATTT
841    S  A  S  S  L  K  A  V  D  I  P  S  F  E  E  S  N  I  A  V  E  E  E  F
      AACCAGCCACTTTCTGTATCCAACTCTTCTCTAGTTGTGAGAAAAGAACCTGATGTACCTGTGTTCTTTCCA
865    N  Q  P  L  S  V  S  N  S  S  L  V  V  R  K  E  P  D  V  P  V  F  F  P
      AGTGGCCAGCTGGCAGAAAGTGTAAGCATGTGTTTACAGACTGGACCAACAGGGGGTGCCAGTAACAATTCT
889    S  G  Q  L  A  E  S  V  S  M  C  L  Q  T  G  P  T  G  G  A  S  N  N  S
      GAAACATCAGAGGAACCAAAAATTGAGCATGTAATGCAACCCTTGCTTCATCAACCATCAGATAACCAGAAA
913    E  T  S  E  E  P  K  I  E  H  V  M  Q  P  L  L  H  Q  P  S  D  N  Q  K
      ATTTACCAGGATTTATTGGGTCAAGTAAACCACCTATTAAATAGTTCCTCCAAGGAAACTGAGCAGCCGTCT
937    I  Y  Q  D  L  L  G  Q  V  N  H  L  L  N  S  S  S  K  E  T  E  Q  P  S
 3246 ACCAAAGCAGTAATTATCAGTCATGAATGCACCAGAACCCAAAACGTTTACCATACAAAGAAAAAAACACAT
961    T  K  A  V  I  I  S  H  E  C  T  R  T  Q  N  V  Y  H  T  K  K  K  T  H
      CATTCAAGACTGGTGGACAAAGATTGTGTCCTTAATGCAACTCTTAAGCAACTAAGAAGCCTTGGAGTAAAA
985    H  S  R  L  V  D  K  D  C  V  L  N  A  T  L  K  Q  L  R  S  L  G  V  K
      ATTGATTCTCCCACTAAAGTGAAGAAAAATGCACATAACGTGGATCACGCCAGTGTGTTGGCATGCATCAGC
1009   I  D  S  P  T  K  V  K  K  N  A  H  N  V  D  H  A  S  V  L  A  C  I  S
      CCAGAAGCAGTGATCTCTGGATTAAACTGCATGTCATTTGCTAATGTTGGCATGAGCGGCTTAAGCCCCAAT
1033   P  E  A  V  I  S  G  L  N  C  M  S  F  A  N  V  G  M  S  G  L  S  P  N
      GGTGTGGATTTGAGCATGGAGGCAAATGCTATAGCTCTGAAATATTTAAATGAAAATCAGCTGTCACAACTG
1057   G  V  D  L  S  M  E  A  N  A  I  A  L  K  Y  L  N  E  N  Q  L  S  Q  L
 3606 TCTGTCACTCGATCGAACCAAAATAATTGTGACCCATTCAGCCTTCTCCATATTAATACAGACAGAAGCACA
1081   S  V  T  R  S  N  Q  N  N  C  D  P  F  S  L  L  H  I  N  T  D  R  S  T
      GTGGGGCTTAGTTTAATTTCACCAAACAACATGTCATTTGCAACCAAAAAATATATGAAGAGATATGGACTC
1105   V  G  L  S  L  I  S  P  N  N  M  S  F  A  T  K  K  Y  M  K  R  Y  G  L
      CTACAAAGCAGTGACAATAGTGAAGATGAAGAGGAACCTCCCGACAATGCAGATAGCAAGAGTGAATATTTA
1129   L  Q  S  S  D  N  S  E  D  E  E  E  P  P  D  N  A  D  S  K  S  E  Y  L
      TTGAATCAGAACCTTAGGTCCATACCCGAACAGCTTGGTGGTCAGAAAGAGCCTTCTAAGAATGACCATGAA
1153   L  N  Q  N  L  R  S  I  P  E  Q  L  G  G  Q  K  E  P  S  K  N  D  H  E
      ATAATTAATTGTTCTAACTGTGAATCTGTGGGGACCAACGCAGATACGCCAGTATTGAGAAATATTACAAAT
1177   I  I  N  C  S  N  C  E  S  V  G  T  N  A  D  T  P  V  L  R  N  I  T  N
 3966 GAAGTTTTGCAGACAAAAGCAAAACAGCAGTTGACTGAAAAGCCAGCTTTCTTAGTAAAGAACCTTAAACCA
1201   E  V  L  Q  T  K  A  K  Q  Q  L  T  E  K  P  A  F  L  V  K  N  L  K  P
      AGTCCTGCAGTGAACCTTCGAACCGGGAAAGCAGAGTTCACTCAACATCCTGAGAAAGAAAATGAAGGGGAC
1225   S  P  A  V  N  L  R  T  G  K  A  E  F  T  Q  H  P  E  K  E  N  E  G  D
      ATTACAATTTTTCCTGAAAGTTTGCAACCTTCTGAAACGCTAAAGCAGATGAATAGCATGAATTCAGTAGGC
1249   I  T  I  F  P  E  S  L  Q  P  S  E  T  L  K  Q  M  N  S  M  N  S  V  G
      ACCTTCTTAGATGTAAAACGTCTCAGACAGTTACCAAAATTATTTTAA
1273   T  F  L  D  V  K  R  L  R  Q  L  P  K  L  F  -
      CCTTTTAACTCCCTGCCCTTTTAATACAGGGACAGGGTGTCTCCTGAAGATACTTAGGGAAAACAGGAGCCT
 4326 ACCACAAGGCTCCTGATCATTCTGGAGTCACTGTTTCTTGGTAGCAGCCAATTGGGAAGAGTGACTTCTGTG
      AGATGGCTGGCTGGTGATAGGACTAAGTTCTCATTGTTCAAATAGAGCTGTTCAACATCACTGAAACCTTTA
      AGAAAAGCCCTGAGATCAGTTATTCCTACAAGTTTAAGTAGTAGACAGATACTATCCAGCTCTAAGTCTCAA
      CTGCTCTTTTATACTGTACTTTTTTTTTTGAGACGGAGTTTTGCTCTTGTAGCCCAGGCTGGAGTGCAATGG
      CAGGATCTCAGATCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTTTCCTGCTTCATCTTCCCAGGTAGC
 4686 TGGGATTACAGGCATGTGCCACAACGCCTGGCTAATTTTGTATTTTTAGTAGAGACTGGTTTCTCCATGTTG
      GTCAGGCTGGTCTCAAACTCCCGACCTCAGGTGATCCGCCCGCCTCGGCCTCCTAAAGTGCTGGGATTACAG
      GCGTGAGCCACTGCCCAGCTATACTGTATATTTTAAGAAGTTCCAGCATGTTGCATCTCTGCATTATCCTAT
      ATCATAAAAGAAGCATAAGTTATCATGGTGTTGGGTAAATTAGCGAAATCAACCGCTTCCTAAGTTTAAGGG
      AAAAGTTATTTTTAAAAACAACTTAATAAAAACTTACACTCTTATACAAGAGTGTATTTCCCCTTAATTAGG
 5046 ATGCATGTTGATTAAACTCGAGATACAGCTTTTTGCAGTATGGTGGGTTGGTTTTGGTGTAACATCTTCAAC
      ATGTCACACTGGCTATCAAAGAATAAGAAAATTATTGAGTATGAGTGTGTTTTATAAACTTTCTGAGTTTTT
      CAGATGTCTTAATATTTTT
```

FIG. 2—*Continued.*

A

```
-202 GTCCGCCCGC AGTTCTCCAA GAAGACTTGG GATTGGTCGA GCGCGGAACC AGTGCGGGGC
-142 GCTGATTGGT CGGCACACCA ATACGTAACG GCGACCGTGC GCGGCTCTCT AGCACCACCC
                                                                  *
 -82 CCGCTCCCTG ACTGGCGAGG TTTCTGACCA GTCAGCAGGC GTGGCGCGGC CTTCAGTTTC
                         *
 -22 GCGAGCTTGT GTTTGCCGCC TCAGTTCCCG CGACCCCAAC GTCCCAGAGG CGGGGCCGGA
 +39 GTCGGCGGTG GCGCTCCTTG GAGCCGGCTC CCGCTCCTAC CCTGCAAACA GACCTCAGCT
 +99 CCGCGGAAGT TGCG
```

B

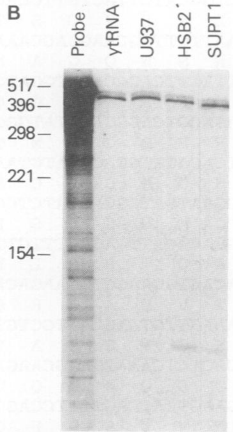


FIG. 3. (A) Nucleotide sequence 5′ of the *SIL* transcript initiation site. Exon 1 is boxed; nucleotide 1 is marked with an asterisk and represents the major transcript initiation site. A minor transcript initiation site (see text) at nucleotide −28 is also marked with an asterisk. Three CCAAT boxes are underlined; an Sp1 binding site is underlined with dashes. (B) Mapping the *SIL* transcript initiation site by RNase protection. A 0.5-kb *Not*I-*Apa*I genomic fragment encompassing *SIL* exon 1 was used as a probe. A 112-bp fragment (arrow) is seen in the U937, HSB-2, and SUPT1 lanes but not the ytRNA (yeast tRNA) control lane. Sizes are indicated in nucleotides.

as shown in Fig. 5. *SCL* exon 1b is a transcript initiation exon (3) and lacks a splice acceptor site. Although *SIL* exon 1 could potentially splice to *SCL* exon 2b, we did not detect this mRNA form. This splicing pattern is similar to that seen with the normal *SCL* mRNA (3), in which exon 2b is incorporated into only a minority of mature *SCL* mRNA forms. As both *SIL* exon 1 and *SCL* exon 3 are 5′ untranslated exons, the *SIL*/*SCL* fusion mRNA predicts a full-length SCL protein.

The predicted SIL protein. Three related but distinct forms of the SIL protein are encoded by the *SIL* cDNA clones. RNase protection analysis (data not shown) was used to determine the relative abundance of these different forms in human bone marrow and a variety of hematopoietic cell lines, including Jurkat, SUPT11, HSB-2, and K562. Form A (Fig. 6A) was the most prevalent mRNA species in all cell lines studied and encodes a protein of 1,287 amino acids, residues 1 to 1287 in Fig. 2. Form B differed from form A by an in-frame internal deletion of 141 nucleotides (Fig. 2) and was less abundant but easily detectable by RNase protection. Form B predicts a protein of 1,240 amino acids, residues 1 to 165 and 213 to 1287, with residues 166 to 212 being deleted. Form C, which predicts an amino-terminal truncated SIL protein of 1,150 amino acids (residues 138 to 1287), was undetectable by RNase protection, indicating that it represented a low-abundance transcript in the cell lines and tissues examined. All three forms of the predicted SIL protein were identical for amino acid residues 213 to 1287. None of the three forms showed significant homology with any of the protein sequences deposited in the SWISSPROT data base. The predicted SIL protein is rich in serine, proline, and asparagine residues, has a predicted isoelectric point of 5.92, and has no membrane-spanning domains. A potential eukaryotic topoisomerase I active site was identified by using the PROSITE function (5) of PC-Gene. The consensus sequence for eukaryotic topoisomerase I proteins, Ex(8)SKx(2)Y(L/I/M), with the tyrosine residue being the active site (12, 19), is indicated in Fig. 6B. All of the protein sequences deposited in the SWISSPROT data base containing this consensus sequence are known eukaryotic topoisomerase I proteins (5).

TABLE 1. *SIL* mRNA expression[a]

Cell line or tissue type	*SIL* expression[b]
Human cell lines	
Jurkat	+
Molt 4	+
Hut 78	+
DU 528	+
SUPT1	+
SUPT11	+
HSB-2	+
CEM	+
RPMI 8402	+
K562	+
HEL	+
KK124	+
HL60	+
U937	+
Human tissues	
Fetal liver	+
Thymus	+
Bone marrow	+
Peripheral blood mononuclear cells	+
Placenta	−
Brain	−
Liver	−
Murine tissues	
Embryo (11.5 days)	+
Thymus	+
Liver	−
Brain	−
Heart	−
Lung	−
Stomach	−
Kidney	−
Muscle	−
Spleen	−

[a] Ten micrograms of total RNA or 2 μg of poly(A) RNA was analyzed by Northern blotting. Integrity of the RNA was assessed by ethidium staining (total RNA) or hybridization to an actin probe [poly(A) RNA] or both. Jurkat, Molt 4, Hut 78, SUPT11, HSB-2, CEM, and RPMI 8402 are T-cell leukemia cell lines, SUPT1 is a T-cell lymphoma cell line, DU 528 is a stem cell leukemia cell line, K562 is a chronic myelogenous leukemia cell line, HEL is an erythroleukemia cell line, KK124 is a Burkitt's lymphoma cell line, HL60 is a promyelocytic leukemia cell line, and U937 is a monocytic cell line.

[b] +, detectable signal on the autoradiogram after hybridization to the 1.1KR *SIL* cDNA probe; −, no signal obtained even after a 1-week exposure of the autoradiogram.

DISCUSSION

In many instances of recurrent, nonrandom chromosomal translocations associated with specific malignancies, the translocations disrupt genes that are essential for the growth or differentiation of the involved cell. The *SIL* gene, which is frequently disrupted in the leukemic cells from patients with

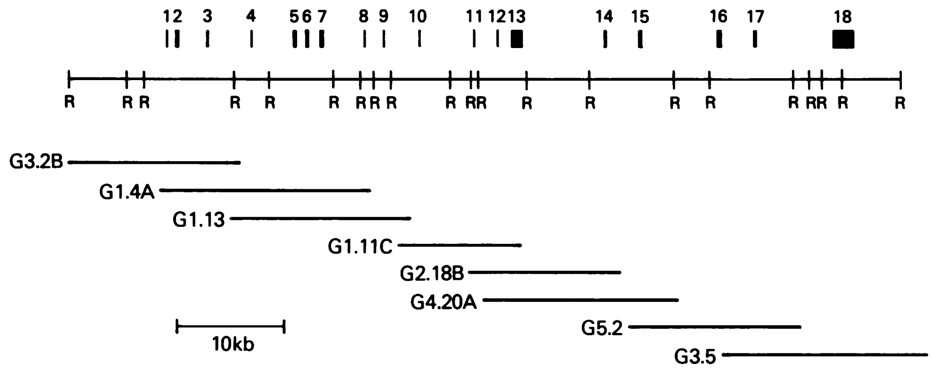


FIG. 4. Genomic structure of *SIL*. Eight overlapping genomic phage clones, encompassing 80 kb of genomic DNA, are shown. *Eco*RI sites (R) are indicated. Exons are as shown; the smaller exons are not drawn to scale.

T-cell acute lymphoblastic leukemia (ALL), has now been cloned and sequenced. On the basis of the cDNA sequence, three distinct but similar forms of the SIL protein are predicted. One form (form A) predominates in all tissues and cell lines studied, and sequence analysis predicts a protein of 143 kDa. The predicted SIL protein has a potential eukaryotic topoisomerase I active site. On the basis of protein sequence similarities to known topoisomerases, a recent review (26) has speculated that two proteins known to be involved in intrachromosomal DNA recombination, mammalian RAG1 (recombinase-activating gene I) (24) and the yeast HPR1 gene product (2), may have topoisomerase I activity. Similar to the topoisomerases and these two proteins, it is possible that the SIL gene product functions in either DNA recombination or replication. There is no other similarity yet seen among SIL and these other proteins.

The genomic structure of *SIL* demonstrates a fairly large gene, consisting of 18 exons distributed over 70 kb. Transcripts that differ at their 5′ ends are generated by alternate mRNA splicing in both the 5′ untranslated region and the predicted coding region. This type of alternate splicing had been reported for other genes (reference 3 and references therein) and may be relevant to the function of the SIL gene product(s). As shown in Fig. 5, the *SIL/SCL* rearrangement leads to deletion of the body of the *SIL* gene on the rearranged allele.

A limited survey of cell lines and tissues indicated that *SIL* mRNA expression could be detected only in hematopoietic or embryonic cell lines and tissues. Both Northern and RNase protection analysis indicated that *SIL* was expressed at low levels in all hematopoietic tissues and cell lines studied. However, with the exception of whole mouse embryos and embryonic cell lines, we did not detect *SIL* transcripts in any nonhematopoietic tissues. This limited spectrum of expression suggests that the SIL gene product may play a role exclusively in hematopoietic growth or differentiation. Alternatively, *SIL* may be expressed in nonhematopoietic tissues, at levels too low to be detected by Northern or RNase protection analysis. Preliminary experiments (3a) indicate that *SIL* mRNA expression decreases to undetectable levels when certain cell lines (HL60 and U937) are treated with agents that induce terminal differentiation, suggesting that the SIL gene product is active when cells are in the proliferative but not the terminally differentiated state.

It remains unclear how the *SIL/SCL* rearrangement, seen in the leukemic cells from 16 to 26% of T-cell ALL patients (4a, 8), may contribute to leukemogenesis. T-cell ALL patients with a t(1;14) translocation involving *SCL* all show either a functional or structural disruption of the normal *SCL* 5′ regulatory region, leading to inappropriate *SCL* mRNA expression (reference 3 and references therein). We have previously speculated that inappropriate expression of SCL

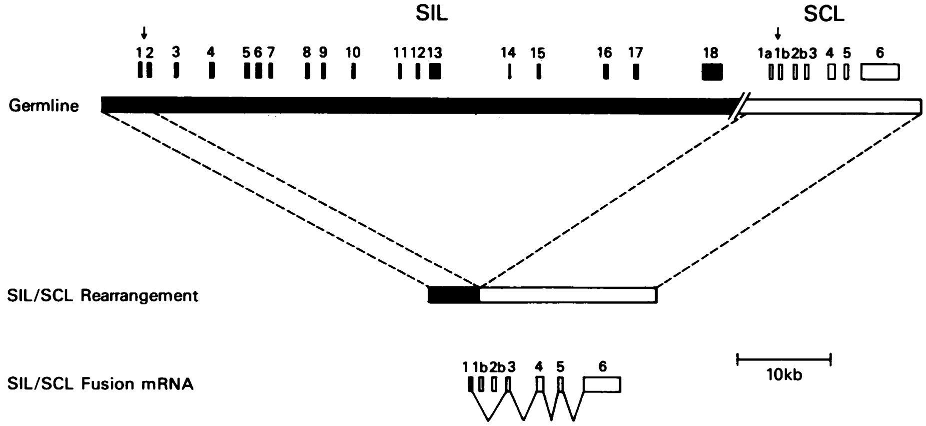
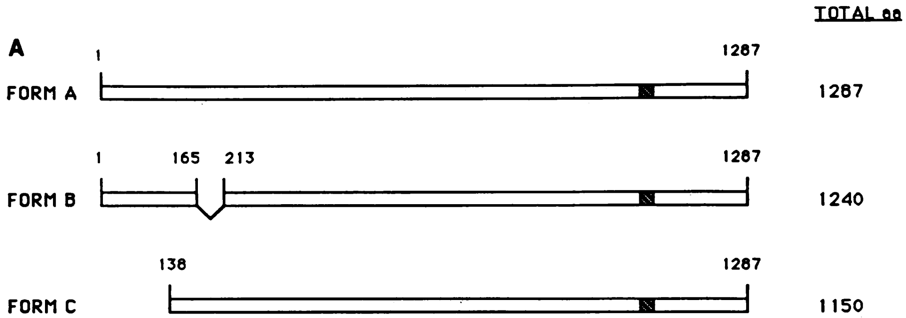


FIG. 5. Schematic representation of *SIL/SCL* fusion mRNA. The germ line *SIL* (solid boxes) and *SCL* (open boxes) genomic structures are shown. The deletion breakpoints are indicated with arrows. The *SIL/SCL* genomic rearrangement, as previously reported (4), is indicated below. The *SIL/SCL* fusion mRNA is formed by *SIL* exon 1 (solid box) splicing to *SCL* exon 3 (open box) in a head-to-tail fashion.




```
B
                                 L
Consensus  E                SK  YI
                                 M

SIL        EEEPPDNADSKSEYL
Human      ENKQIALGTSKLNYL
Vaccinia   EVVGHTPSISKRAYM
S. cer.    ENSQVSLGTSKINYI
```

FIG. 6. (A) Predicted SIL proteins. Forms A, B, and C correspond to the different *SIL* mRNA species, as discussed in the text. Form A contains amino acid (aa) residues 1 to 1287 of Fig. 2, form B contains residues 1 to 165 and 213 to 1287, and form C contains residues 138 to 1287. All forms retain the predicted eukaryotic topoisomerase I active site (▧). (B) Conserved amino acid residues surrounding the eukaryotic topoisomerase I active site compared with the amino acid sequences from SIL, human topoisomerase I (Human), vaccinia virus topoisomerase I (Vaccinia), and *S. cerevisiae* topoisomerase I (S. cer.).

in these patients may contribute to malignant transformation, in a manner analogous to inappropriately expressed c-*myc* associated with Burkitt's lymphoma (18). It is possible that the *SIL/SCL* deletion contributes to malignant transformation in a manner similar to that hypothesized for patients with t(1;14) translocations, as the *SIL/SCL* rearrangement also disrupts the 5′ regulatory region of *SCL*. The *SIL/SCL* rearrangement produces a fusion mRNA (4) potentially encoding a full-length SCL protein. Furthermore, this *SIL/SCL* fusion mRNA is not an artifact of tissue culture. We have recently studied mRNA from the leukemic blasts of several patients with newly diagnosed T-cell ALL, and all express an identical *SIL/SCL* fusion message (4a). The net effect of this rearrangement is to put *SCL* transcription under control of the *SIL* 5′ regulatory region. In this light, it is important to note that while *SCL* is not normally expressed in T cells (7), *SIL* is. Therefore, in T cells that have undergone a *SIL/SCL* deletion, the 5′ regulatory region of *SCL*, a gene not normally expressed in T cells, is replaced by the 5′ regulatory region of *SIL*, which normally is expressed in T cells, leading to inappropriate *SCL* expression. While it is possible that *SIL* functions as a tumor suppressor gene and that the deletion of one copy of *SIL* contributes to malignant transformation, we have no reason to believe that this is the case. In sum, we have taken advantage of a commonly occurring chromosomal deletion, associated specifically with T-cell ALL, to identify and characterize the cDNA and genomic structure of two distinct genes, *SCL* (3) and *SIL* (this report). It seems very likely that the fusion of these two genes contributes significantly to the malignant transformation of T cells.

REFERENCES

1. **Adams, R. A., A. Flowers, and B. J. Davis.** 1968. Direct implantation and serial transplantation of human acute lymphoblastic leukemia in hamsters, SB-2. Cancer Res. **28:**1121–1125.
2. **Aguilera, A., and H. L. Klein.** 1990. HPR1, a novel yeast gene that prevents intrachromosomal excision, shows carboxy-terminal homology to the *Saccharomyces cerevisiae TOP1* gene. Mol. Cell. Biol. **10:**1439–1451.
3. **Aplan, P. D., C. G. Begley, V. L. Bertness, M. Nussmeier, A. Ezquerra, J. Coligan, and I. R. Kirsch.** 1990. The *SCL* gene is formed from a transcriptionally complex locus. Mol. Cell. Biol. **10:**6426–6435.

3a. **Aplan, P. D., and I. R. Kirsch.** Unpublished data.

4. **Aplan, P. D., D. P. Lombardi, A. M. Ginsberg, J. Cossman, V. L. Bertness, and I. R. Kirsch.** 1990. Disruption of the human SCL locus by "illegitimate" V(D)J recombinase activity. Science **250:**1426–1429.

4a. **Aplan, P. D., D. P. Lombardi, G. Reaman, and I. R. Kirsch.** Submitted for publication.

5. **Bairoch, A.** 1990. PROSITE: a dictionary of protein sites and patterns, 6th release. University of Geneva, Geneva, Switzerland.
6. **Begley, C. G., P. D. Aplan, M. P. Davey, K. Nakahara, K. Tchorz, J. Kurtzberg, M. Hershfield, B. F. Haynes, D. I. Cohen, T. A. Waldmann, and I. R. Kirsch.** 1989. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc. Natl. Acad. Sci. USA **86:**2031–2035.
7. **Begley, C. G., P. D. Aplan, S. M. Denning, B. F. Haynes, T. A. Waldmann, and I. R. Kirsch.** 1989. The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc. Natl. Acad. Sci. USA **86:**10128–10132.
8. **Brown, L., J.-T. Cheng, Q. Chen, M. J. Siciliano, W. Crist, G. Buchanan, and R. Baer.** 1990. Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. EMBO J. **9:**3343–3351.
9. **Chen, Q., J.-T. Cheng, L.-H. Tsai, N. Schneider, G. Buchanan, A. Carroll, W. Crist, B. Ozanne, M. J. Siciliano, and R. Baer.** 1990. The tal gene undergoes chromosome translocation in T-cell leukemia and potentially encodes a helix-loop-helix protein. EMBO J. **9:**415–424.
10. **Corden, J., B. Wasylyk, P. Buchwalder, P. Sossone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein coding genes. Science **209:**1406–1413.
11. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. Basic methods in molecular biology. Elsevier Science Publishing Co., New York.
12. **Eng, W.-K., S. D. Pandit, and R. Sternglanz.** 1989. Mapping of the active site tyrosine of eukaryotic DNA topoisomerase I. J. Biol. Chem. **264:**13373–13376.
13. **Finger, L. R., J. Kagan, G. Christopher, J. Kurtzberg, M. S. Herschfield, P. C. Nowell, and C. M. Croce.** 1989. Involvement of the TCL5 gene on human chromosome one in T-cell leukemia and melanoma. Proc. Natl. Acad. Sci. USA **86:**5039–5043.
14. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:**8998–9002.
15. **Hecht, F., R. Morgan, B. K. M. Hecht, and S. D. Smith.** 1984. Common region on chromosome 14 in T-cell leukemia and lymphoma. Science **226:**1445–1447.
16. **Kaplan, J., J. Tilton, and W. D. Peterson.** 1976. Identification of T-cell lymphoma tumor antigens on human T-cell lines. Am. J. Hematol. **1:**219–223.
17. **Kirsch, I. R., J. A. Brown, J. Lawrence, S. J. Korsmeyer, and**

C. C. Morton. 1985. Translocations that highlight chromosomal regions of differentiated activity. Cancer Genet. Cytogenet. **18:**159–171.

18. **Leder, P., J. Battey, G. Lenoir, C. Moulding, W. Murphy, H. Potter, T. Stewart, and R. Taub.** 1983. Translocations among antibody genes in human cancer. Science **222:**765–771.
19. **Lynn, R. M., M.-A. Bjornsti, P. R. Caron, and J. C. Wang.** 1989. Peptide sequencing and site-directed mutagenesis identify tyrosine-727 as the active site tyrosine of *Saccharomyces cerevisiae* DNA topoisomerase I. Proc. Natl. Acad. Sci. USA **86:**3559–3563.
20. **Nowell, P. C., and D. A. Hungerford.** 1960. A minute chromosome in human chronic granulocytic leukemia. Science **132:** 1417–1419.
21. **Padgett, R. A., P. J. Grabowski, M. M. Konerska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55:**1119–1150.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Sandberg, A.** 1990. The chromosomes in human cancer and leukemia, 2nd ed. Elsevier Scientific Publishing Co., New York.
24. **Schatz, D. G., M. A. Oettinger, and D. Baltimore.** 1989. The V(D)J recombination activating gene, RAG1. Cell **59:**1035–1048.
25. **Southern, E. M.** 1975. Detection of specific DNA sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
26. **Wang, J. C., P. R. Caron, and R. A. Kim.** 1990. The role of DNA topoisomerases in recombination and genome stability: a double-edged sword? Cell **62:**403–406.