

# ***MSF* (MLL septin-like fusion), a fusion partner gene of *MLL*, in a therapy-related acute myeloid leukemia with a t(11;17)(q23;q25)**

(chromosome translocation)

MITSUHIKO OSAKA, JANET D. ROWLEY, AND NANCY J. ZELEZNIK-LE\*

Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL 60637

Contributed by Janet D. Rowley, March 22, 1999

**ABSTRACT** *MLL* (*ALL1*, *Htrx*, *HRX*), which is located on chromosome band 11q23, frequently is rearranged in patients with therapy-related acute myeloid leukemia who previously were treated with DNA topoisomerase II inhibitors. In this study, we have identified a fusion partner of *MLL* in a 10-year-old female who developed therapy-related acute myeloid leukemia 17 months after treatment for Hodgkin's disease. Leukemia cells of this patient had a t(11;17)(q23;q25), which involved *MLL* as demonstrated by Southern blot analysis. The partner gene was cloned from cDNA of the leukemia cells by use of a combination of adapter reverse transcriptase-PCR, rapid amplification of 5' cDNA ends, and BLAST database analysis to identify expressed sequence tags. The full-length cDNA of 2.8 kb was found to be an additional member of the septin family, therefore it was named *MSF* (MLL septin-like fusion). Members of the septin family conserve the GTP binding domain, localize in the cytoplasm, and interact with cytoskeletal filaments. A major 4-kb transcript of *MSF* was expressed ubiquitously; a 1.7-kb transcript was found in most tissues. An additional 3-kb transcript was found only in hematopoietic tissues. By amplification with *MLL* exon 5 forward primer and reverse primers in *MSF*, the appropriately sized products were obtained. *MSF* is highly homologous to *hCDCrel-1*, which is a partner gene of *MLL* in leukemias with a t(11;22)(q23;q11.2). Further analysis of *MSF* may help to delineate the function of *MLL* partner genes in leukemia, particularly in therapy-related leukemia.

Chemotherapeutic drugs are recognized as double-edged swords. Cancer chemotherapy is a cause of acute myeloid leukemia (AML) in the young. Secondary or therapy-related AML (t-AML) accounts for 10–20% of all leukemia cases (1). DNA topoisomerase II inhibitors are widely used for cancer treatment, but they also are recognized to evoke secondary leukemias with a short latency period.

*MLL* (*ALL1*, *Htrx*, *HRX*), which is located on chromosome band 11q23 (2–6), frequently is rearranged by chromosome translocations in *de novo* leukemias and also in t-AML patients who previously received DNA topoisomerase II inhibitors for a primary malignant disease (7–11). The wild-type *MLL* protein with a predicted molecular mass of 431 kDa has AT-hook DNA binding domains, transcriptional activation and repression domains, multiple plant homeodomain (PHD) finger domains, and a region of homology to mammalian DNA methyltransferases. The C terminus of *MLL* contains a [Su(var)3-9, enhancer of zeste, and trithorax] (SET) domain, which along with the PHD domain, is conserved with the *Drosophila trithorax* (*trx*) gene. *Trx* is involved in segment determination in *Drosophila* through regulation of homeotic genes. *MLL* is thought to be the human homolog of *trx* because

mice with mutant *Mll* demonstrate homeotic transformation (12). It is suggested that *MLL* plays an important role in hematopoiesis and leukemogenesis through its ability to regulate the expression of *Hox* genes. *Mll*-null mutant mice have defects in early stages of hematopoiesis such as hematopoietic progenitor proliferation or recruitment (13). Maturation arrest was observed in hematopoietic differentiation *in vitro* of single or double *Mll* knockout embryonic stem cells (14). Chromosome translocations involving *MLL* yield in-frame fusion products of the N-terminal portion of *MLL* with different partner proteins (15–18); terminal deletions of *MLL* have not been identified in leukemia cells. Nineteen partner genes have been cloned to date (19–34). There has been no shared common structure among any of the partner gene products; however, it is thought that the partner gene is critical for leukemogenesis for several reasons. All partner genes are fused in-frame to *MLL*, resulting in a novel fusion protein, rather than simply a truncation of *MLL*. In a “knock-in” model system, only mice with a fusion partner identical to that seen in human leukemia developed myeloid-lineage leukemias, suggesting that the fusion product is critical for leukemogenesis (35). Furthermore, in a murine bone marrow immortalization system, the *ENL* partner gene was critical for immortalization and leukemogenesis (36). However, it is not understood how the fusion products participate in leukemogenesis.

In the present study, we have identified a gene that was not present in the GenBank database as a fusion partner of *MLL* from a patient who had t-AML with a t(11;17)(q23;q25). This pattern of chromosome translocation has been reported in several cases of acute leukemia and myelodysplastic syndrome (37). This partner gene was named *MSF* (MLL septin-like fusion) because it is highly conserved with genes of the septin family. The septin family proteins are located in the cytoplasm, are characterized by the GTP binding domain, and are involved in cytokinesis caused by interaction with cytoskeletal filaments (38–41). *MSF* is highly homologous to and conserves the GTP binding domain with *hCDCrel-1*, which is a partner gene of *MLL* in leukemias with a t(11;22)(q23;q11.2) (33). We discuss here possible mechanisms for the way in which the *MLL-MSF* fusion product contributes to leukemogenesis based on these findings.

## MATERIALS AND METHODS

**Patient.** A 10-year-old female was treated for Hodgkin's disease with four courses of COPP (600 mg/m<sup>2</sup> of cytoxan, 1.5 mg/m<sup>2</sup> of vincristine, 60 mg/m<sup>2</sup> per day of procarbazine, and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: AML, acute myeloid leukemia; t-AML, therapy-related AML; RT, reverse transcriptase; 5' RACE, rapid amplification of 5' cDNA ends; EST, expressed sequence tag.  
Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF123052).

\*To whom reprint requests should be addressed at: Section of Hematology/Oncology, Department of Medicine, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637. e-mail: njzelezn@mcis.bsd.uchicago.edu.

100 mg/m<sup>2</sup> per day of prednisone) and ABV (35 mg/m<sup>2</sup> of adriamycin, 10 mg/m<sup>2</sup> of bleomycin, and 6 mg/m<sup>2</sup> of vinblastine). Adriamycin is a DNA topoisomerase II inhibitor. After the chemotherapy, she received 24 Gy of radiation in the mini-mantle area. She developed t-AML with t(11;17)(q23;q25) 17 months after completing the therapy. Leukemia cells from the peripheral blood were obtained with informed consent and were subjected to the following analyses.

**Cloning of Chimeric cDNA.** Total RNA was isolated by use of Tri Reagent (Molecular Research Center, Cincinnati). First-strand cDNA was synthesized from 1 µg of total RNA with random hexamer primers, and second-strand cDNA was synthesized by using the PCR-select cDNA subtraction kit (CLONTECH). Double-stranded cDNA was digested with *Rsa*I and ligated with adapters (5'-CTA ATA CGA CTC ACT ATA GGG CTC GAG CGG CCG CCC GGG CAG GT-3') or (5'-CTA ATA CGA CTC ACT ATA GGG CAG CGT GGT CGC GGC CGA GGT-3') according to the manufacturer's protocol. PCR was used for amplification of the fusion gene with biotin-labeled *MLL* exon 5 primer EX5VNP (5'-CCT GAA TCC AAA CAG GCC ACC ACT-3') and adapter specific primers AP1 (5'-TCG AGC GGC CGC CCG GGC AGG T-3') or AP2 (5'-AGC GTG GTC GCG GCC GAG GT-3'). A 50-µl PCR contained 0.5 µg of cDNA, 100 pmol of each primer, 50 µM dNTPs, 1× reaction buffer (50 mM KCl/10 mM Tris-HCl, pH 9.0/1 mM MgCl<sub>2</sub>), and 2.5 units of *Taq* DNA polymerase (Boehringer Mannheim). The PCR program was: 4 min at 94°C, 30 cycles of 30 sec at 94°C, 30 sec at 55°C, 2 min at 72°C followed by 5 min at 72°C. One-tenth of the product was enriched on streptavidin-magnetic beads (Dynal, Oslo) and used as template for a second round of PCR with *MLL* exon 5 primer EX5NP (5'-GAT AAG CTT CCA GGA AGT CAA GCA AGC AGG-3') and adapter primers. The PCR product was fractionated on a 1% Tris/acetate/EDTA (TAE)-agarose gel, purified by using the QIAquick gel extraction kit (Qiagen, Chatsworth, CA), and ligated into pGEM T-vector (Promega). Plasmids were sequenced by use of the Applied Biosystems PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer) and an automated sequencer (model 377A, Applied Biosystems). The reaction was: 2 min at 94°C, 30 cycles of 30 sec at 94°C, 30 sec at 42°C, 4 min at 60°C. Both strands were completely sequenced by using sense and antisense primers. Searches for sequence similarity were performed with the GenBank nucleic acid and protein databases by using the BLAST program (42, 43).

**Southern and Northern Blot Analysis.** DNA was extracted from leukemia cells by standard techniques. Eight micrograms of genomic DNA was digested with *Bam*HI or *Hind*III, electrophoresed on 0.8% agarose gels, and transferred to Hybond-N<sup>+</sup> nylon membranes (Amersham Pharmacia). The 0.74-kb *Bam*HI fragment (16) was used to detect rearrangement of *MLL*, and a 614-bp PCR product amplified by primers MSFF2 (5'-TGG GCG TGA AGA ACT CAG AA-3') and MSFR3 (5'-TTC TCG TTC CGT GAT GCA AG-3') was used to detect *MSF*. DNA probes were labeled with α-<sup>32</sup>P-dCTP (Amersham Pharmacia) by the Megaprime DNA Labeling System (Amersham Pharmacia). The filters were prehybridized and hybridized by using standard conditions and were exposed to x-ray film (Kodak) for 24–48 hr. Human multiple tissue Northern blots were purchased from CLONTECH.

**Rapid Amplification of 5' cDNA Ends (5' RACE).** To map the transcriptional start site of the partner gene, we used the 5' RACE system (GIBCO/BRL). One microgram of total RNA of the K562 cell line was reverse-transcribed with primer MSFR3, homopolymeric tails of dCTP were added, and the dC-tailed cDNA was used as template for PCR with an anchor primer (5'-GGC CAC GCG TCG ACT AGT AC-3') and primer MSFR4 (5'-TGA ATT CTC CAA GGT CTG GG-3'). The reaction was: 4 min at 94°C, 30 cycles of 30 sec at 94°C,

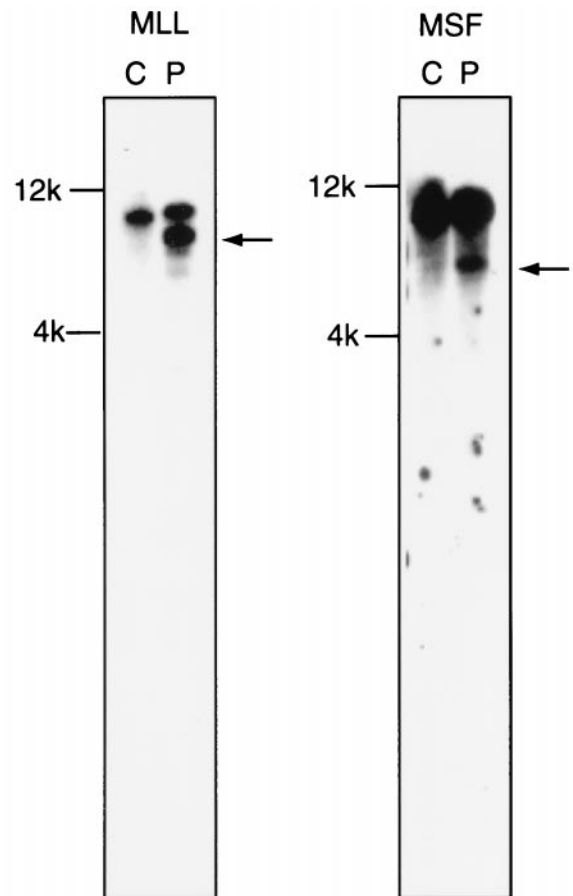


Fig. 1. Southern blot analysis. *MLL* and *MSF* are involved in the t(11;17). DNAs digested with *Hind*III were hybridized with the *MLL* probe (Left) and *MSF* probe (Right). Rearranged bands are found in both hybridizations (indicated with an arrow), demonstrating that *MLL* and *MSF* are involved in the chromosome translocation. C, control human placenta DNA (Sigma); P, patient's DNA.

30 sec at 55°C, 2 min at 72°C followed by 5 min at 72°C. The PCR product was purified, ligated into pGEM T-vector, and sequenced as described above.

**Reverse Transcriptase (RT)-PCR.** cDNA was synthesized from 1 µg of total RNA of the patient's leukemia cells with random hexamer primers with the First Strand cDNA Kit (GIBCO/BRL). Total RNA of leukemia cell line K562 was the

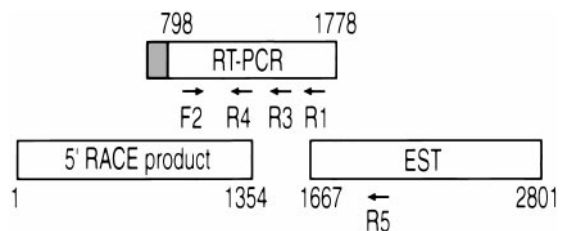


Fig. 2. *MSF* cDNAs used for sequencing. A chimeric RT-PCR product of *MLL* exon 5 (gray box) fused with unknown sequences initially was isolated. BLAST search showed that these sequences were identical to a portion of EST54371 (GenBank accession no. 347984), but did not exactly match any known genes. This EST was sequenced; 112 bp matched with the RT-PCR sequence exactly and extended it 1,023 bp downstream 3'. To obtain the sequence upstream of the breakpoint, we performed 5' RACE using cDNA of the leukemia cell line K562 as template and MSFR4 in *MSF* as the primer. 5' RACE yielded a 1.4-kb product that had sequences overlapping the initial sequence obtained from RT-PCR. Arrows show location of the primers used in this study. The numbers indicate the nucleotide positions shown in Fig. 5.

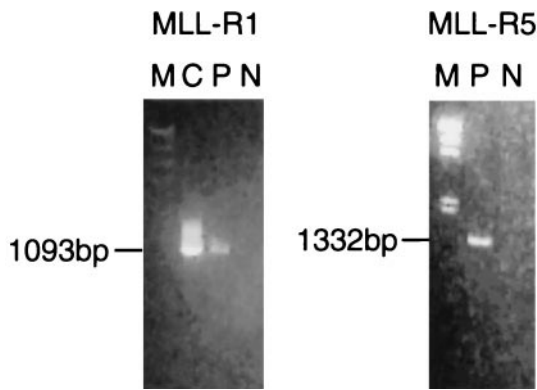


Fig. 3. RT-PCR of chimeric *MLL-MSF*. RT-PCR with an *MLL* exon 5 forward primer (EX5NP) and reverse primers MSFR1 (Left) and MSFR5 (Right) in *MSF* (see Fig. 2). Amplification with EX5NP and MSFR1 yielded a 1,093-bp product in the positive control (plasmid containing the first chimeric PCR product of *MLL* exon 5 and partial *MSF*) and in the patient's cDNA. A product of 1,332 bp was amplified with EX5NP and MSFR5 in the patient's cDNA. The MSFR5 sequence was in the EST. M,  $\lambda$ HindIII digest marker; C, positive control; P, patient's cDNA; N, negative control (K562 cDNA).

control. For amplification of the fusion gene, a 50- $\mu$ l PCR contained 0.5  $\mu$ g of cDNA, 100 pmol of each primer, 200  $\mu$ M dNTPs, 1 $\times$  reaction buffer (50 mM KCl/10 mM Tris-HCl, pH 9.0/1 mM MgCl<sub>2</sub>), and 2.5 units of *Taq* DNA polymerase (Boehringer Mannheim). Primers were EX5NP in exon 5 of *MLL* and MSFR1 or MSFR5 (5'-CAT AAA CTC GAT GTC CAG GG-3') in *MSF*. The reaction was: 4 min at 94°C, 25 cycles of 30 sec at 94°C, 30 sec at 55°C, 90 sec at 72°C followed by 4 min at 72°C. To check a reciprocal transcription of *MSF-MLL*, we used MSF301 (5'-ATC CGT AGG AAT GGA GAG GG-3') and EX8BOT (5'-CTG GCA CAG AGA AAG CAA ACC AC-3') located in *MLL* exon 8. The same program was used for amplification.

## RESULTS

The t-AML patient's leukemia cells had a t(11;17)(q23;q25). Southern blot analysis of *Hind*III-digested DNA hybridized with the *MLL* probe revealed a rearranged band, suggesting that *MLL* was involved in the chromosome translocation (Fig. 1). One partner gene of *MLL* previously had been cloned from chromosome 17, but from a different cytogenetic band (*AF17* at 17q21) (23). Although a different region of chromosome 17 seemed to be involved in this patient, we used *MLL* forward primers and *AF17* reverse primers in RT-PCR to determine whether *AF17* was involved in the translocation in this patient. No product was amplified, even though controls worked (data not shown). This finding suggested that a new partner gene of *MLL* might be involved in this patient.

To clone the gene fused to *MLL*, double-stranded cDNA was synthesized from total RNA of leukemia cells with random primers, digested with *Rsa*I and ligated to an adapter. First amplification used biotinylated *MLL* exon 5 primer and the adapter primer. Biotinylated PCR products were enriched by magnetic beads, but no product was evident by electrophoresis and ethidium bromide staining. A secondary amplification used a nested *MLL* primer and adapter primer. Four products between 300 bp and 1.1 kb were obtained. After cloning and sequencing, we determined that three products were wild-type *MLL*, but one was *MLL* exon 5 fused with unknown sequences (Fig. 2). A BLAST search showed that these sequences were highly homologous to EST54371 (GenBank no. 347984), but did not exactly match any genes already identified. We sequenced this expressed sequence tag (EST); 112 bp matched the sequence exactly and extended it 1,023 bp 3' (Fig. 2). A portion of the sequence was highly homologous to the septin gene family, therefore we named the gene *MSF* for *MLL* septin-like fusion gene. We also confirmed involvement of *MSF* in the chromosome translocation by Southern blot analysis of *Hind*III-digested DNA. A rearranged band was found in the DNA of the leukemia cells (Fig. 1).

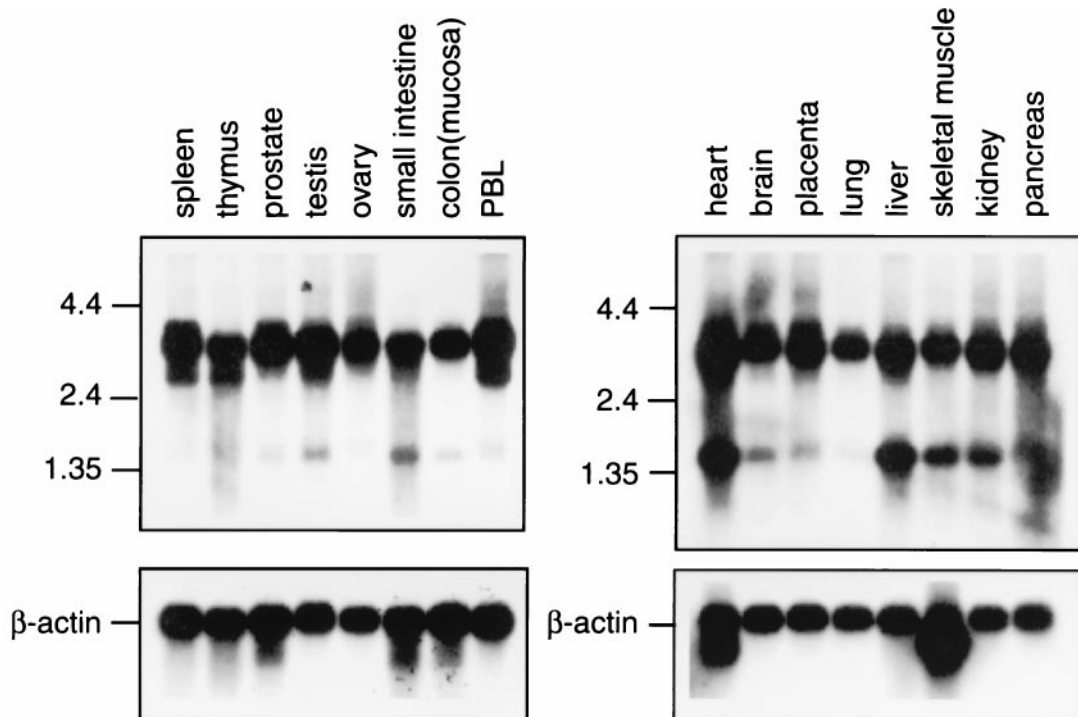


Fig. 4. Northern blot analysis of *MSF* in normal tissues. Multiple-tissue Northern blots (CLONTECH) were hybridized with a 614-bp PCR product amplified by MSFF2 and MSFR3 (see Fig. 2). A major 4-kb transcript was expressed ubiquitously. An additional 3-kb less abundant transcript was found in spleen, thymus, and peripheral blood leukocytes (PBL). A 1.7-kb transcript also was found in many tissues with a relatively higher level of expression in heart, liver, skeletal muscle, and kidney. Blots were rehybridized with a  $\beta$ -actin probe.

To confirm that the sequences obtained for *MSF* outside the original RT-PCR cloned fragment were part of the *MLL* partner gene in the patient sample, we carried out RT-PCR with an *MLL* exon 5 forward primer and various reverse primers in *MSF*, either the MSFR1 primer present in the initial RT-PCR product or the MSFR5 in the EST clone. Appropriately sized products were obtained from cDNA of the patient's leukemia cells, while they were negative in control K562 cDNA (Fig. 3). We also checked for a reciprocal chimeric transcript of 5' *MSF* and 3' *MLL*, but no product was obtained (data not shown).

Expression of wild-type *MSF* was analyzed by Northern blotting. A 614-bp probe including the central portion of *MSF* was used for detection. As shown in Fig. 4, a major 4-kb transcript was expressed ubiquitously. An additional 3-kb transcript was found in spleen, thymus, and peripheral blood leukocytes. The 3-kb transcript was less abundant than the main transcript. A 1.7-kb transcript also was found in most tissues with a higher level of expression in heart, liver, skeletal muscle, and kidney.

To obtain the full-length sequence of this gene, including the sequence upstream of the breakpoint, we performed 5' RACE by using cDNA of the leukemia cell line K562. 5' RACE yielded five products between 500 bp and 1.4 kb, all of which were cloned and sequenced. Two of them, 1.4-kb and 1.1-kb products, had sequences overlapping the initial sequence obtained from RT-PCR. The 1.1-kb product was identical to the 1.4-kb product, but lacked the first 384 bp (Fig. 2).

The total sequence of *MSF* consists of 2,801 bp (Fig. 5, GenBank accession no. AF123052). The ORF starts at nucleotide 776 and encodes a putative protein of 568 aa with a predicted molecular mass of 63.6 kDa. The nucleotide sequence and deduced amino acid sequence were compared with other genes by the BLAST search program. The amino acid sequence of *MSF* from codon 287 to the C terminus is highly homologous to CDC10 and brain protein H5 (hPNU2) of humans, Nedd5 and Diff6 of the mouse, peanut (Pnut) of *Drosophila melanogaster*, and CDC10 in *Saccharomyces cerevisiae*. The C-terminal portion of *MSF* is 40–55% identical and 60–80% similar to these proteins (Fig. 6). Within this homologous region, all of these proteins completely conserve the amino acid consensus sequence for a GTP binding domain, GX<sub>4</sub>GKS–DX<sub>2</sub>G–KXD. *MSF* has 50–60% identity to the others in the nucleotide sequence of this GTP binding region. No significant homology to the N terminus was found.

**DISCUSSION**

We demonstrate here that a previously uncharacterized gene, *MSF*, is fused in-frame to *MLL* exon 5 in a t-AML patient with a t(11;17)(q23;q25). The distribution of *MLL* breakpoints in t-AML tends to be located in the telomeric half of the breakpoint cluster region (BCR) of *MLL* (44, 45). In this case, however, the breakpoint is located in the centromeric half of the BCR. This finding is not unique, however, because other cases of t-AML have 5' *MLL* breaks (44, 45). It is likely that the genomic breakpoint is located in *MLL* intron 5, but we cannot exclude the possibility that the fusion transcript is alternatively spliced. The genomic structure of *MSF* and the *MLL-MSF* fusion must be analyzed to confirm where the break occurs in this case. The *MLL-MSF* fusion product retains the N terminus of *MLL*, which contains the AT-hook DNA binding domain and repression domain, but it loses the activation domain of *MLL*. The C terminus of the fusion product contains all but the first seven amino acids of *MSF*. *MSF* is highly homologous to the genes of the septin family, including *hCDC10* (46) and *hCDCrel-1* (47, 48) in human, and *Nedd5* (38), *Diff6*, and *H5* in the mouse. All of these genes completely conserve the GTP binding domain and are considered to be involved in cytokinesis.

```

1 TGACTTCTCAGGGTCTCTCC TGGGAGGTGCTCCGAAACC TTTTCTCAGACGCTGGCCCT GGGGACGGCCCTCTCCCT
81 GCCCACCACCTTGGGGTAC AGGGTGAAGAAGGGCTGGGG CCAGCCAGGACAGAGGAAG GCGAGGACGAGCAGGAA
161 CTGGGCTTTTAAACCCTTA AGCCCAAGGAATCTGATA TCCGGGACAGGGAAGATAA AAGACTTTGGAACTGCTAG
241 GAATTTGACTCTGTGAGTTG GTTTCAGAGTCTAAGTTA AGCATCTCCAAGTGGATAT AAAAAGGAGCAGCAAGCCTC
321 GGGGCGGGGGGGCTGGAGG AGGTGGAGAGGAGGGCTCG CGGAAGCCGACTCGGGACC TCTGACCCACCCAGCACAC
401 GGGGCGGGGGACTCTGGG ACTCTCGAGGAGGAGCCGG TGGTCTGCCGACTCTCGG GGCCTACTCGGGCCCTCTC
481 TCTCGCCTCTATTTTGGG TTTCTCTCTTTGCTCCCT TTTCTCCCGTTTGAAGAG ACAATGCTACTCTCAGTTGG
561 AGCACAACATATGATCAGC ACATGGAAATGTGTAATTC GGATGCACTCTGATTTGCAA CAGATGAAGAATAAGACC
641 AGACAAGAGTGTTTTGA GAGGAGGAGGAGGAGGAGG AGGCTCAGAGAGGGAGGGCC ACGGGGGTGAGAAGGGGAG
721 GCCCTCTCAGCGGGGACGC CGGGACTCCCCCGCTGCTA AATATACTCGTAGGAATGA GAGGGACGGATCTCAAGCT
801 TGAAGAATCTTTTGAAGTC GAGGAGTTCGAGACCAAAA CTCCACCCACCCCGGAGGG TCCAGACTCCCTACTCCGA
881 GCACCTGTGGCCAGCTCCAC CCAGAAATTCAGAGGACTGG GCGTGAAGAACTCAGAACC TCGGCCCGCATGTGGACTC
961 CCTAAGCCCAAGCTCCCECA AGGCTCCCTCGGAGGGTG GAGCTCTCGGGCCCAAGGC GGCAGGCGGCTGCTCCGCG
1041 GCACCTGACTCTCATTGAC ATCTCTCGAAGCAGGTGGA GAACGCGGGGCGGCTGGCC CGTCCCGGTTCCGGCTCAAG
1121 AGGGCCGAGTGTGGGCCA CAAGACGCGAGAACCGCCC CTCGGAGAGCGAGATCAC ATCTGCAACCCGAGGCTC
1201 AGCCCAAGCGAGGTGGAGC CCCCCTGCTCCAGGCTCCC GAGGTCTCCGCTCCGCTCCG CACCGACCCCTCCCAAGA
1281 AHRRME P P A S K V P E V T A P A T D A A P K 158
1361 GGTGGGAGTCAGATGCC AACCTGCTGAGGCGCCAC GCCCCCGCCAGCCAGGA CCTTGAGAATTCAGAGCTC
1441 RVEIQMP K P A E A P T A P S P A Q T L E N S E P 185
1521 GCGCTCTGTCTCAGCTGCA GAGCAGGCTGGAGCCAAAG CCGACCCCGCTGTGCTGAG GCTACACCCGGAGCAGGA
1601 A P V S Q L Q S R L E P K P Q P P V A E A T P R S O E 212
1681 GCGCCTCAGGCGGCTCCCA GCTGCTGGCGACATGCG GACACCCCGAGAGTCCGG GCTCAAGCAGGCTCCGCTGAT
1761 A T E A A P S C V G D M A D T P R D A G L K Q A P A 238
1841 CACGGAACGAGAGGCGCCG GTGCACTTCGGCTACGTGG GATTGACCTCATCTCGGAGC AGATGCGCCGAAAGCCATG
1921 S R N E K A P V D F G Y V G I D S I L E Q M R R K A M 265
2001 AAGCAGGCTCTCAGTTCAA CATCATGGTGGTGGCGAGA CGGGCTGGGTAAATCACC TTAATCAACACCTCTTCAA
2081 K Q G F E F N I M V Y G Q S E L L G K S T L I N T L F K 292
2161 ATCCAAAATCAGCGGAAGT CGGTGACGCCACTCAGAG GAGGCTACCCCAAGACCAT CAGATCAAGTCCATCACG
2241 S K I S R K S V Q P T S E E R I P K T I E I K S I T 318
2321 ACATATTGAGGAGAAGGCG GTCCGATGAAGCTCAGAG GATTGACCGACAGCGTTCCG GGGACACATCAAGCAAG
2401 H D I E E K G V R M K L T Y I D T P G F G D H I N N E 345
2481 AAAGCGCATCCCGGACACCC GAAGTTCATCAATGACAGT ACAGAGAACTCTGCAAGG GAGGTCAACATCAACCGAA
2561 N C W Q P I M K F I N D Q Y E K Y L Q E V N I N R K 372
2641 GAAGCGCATCCCGGACACCC GCGTCACTGCTGCTCTAC TTCATCCCGCCGACGGCCA CACACTCACCTCGGAGCA
2721 K R I P D T R V H C C L Y F I P A T G H S L R P L D 398
2801 TCGAGTTTATGAAGCGCTG AGCAAGTGGTCAACATCG CCGTGTATCGCTCAAGCGCG ACACACTCACCTCGGAGG
2881 I E F M K R L S K V V N I V P V I A K A D T L T L E 425
2961 AGCGTCCACTTCAACAGCG GATCACCGAGACTGCTGT CCAAGGCGATCGACTGTAC CCCCAGAGGAATTTGATG
3041 R V H F K Q R I T A D L L S N G I D V V P Q K E F D E 452
3121 GNACTCGGAGACCGGCTGG TGAACGAGAGTCCGGAG ATGATCGCCATTGTGCTGGT GGGCACTGACCAAGTACC
3201 D S E D R L V N E K F R E M I P F A V V G S D H E Y 478
3281 AGGTCAACGAGAGGAGCTC CTGGGAGGAGCAAGTGG GGGTACCTATCGATTTGAAA ACACACACACAGCTGAGTTT
3361 Q V N G K R I L G R K T K G T I E V N T T H C F E 505
3441 GCTCATCTCGGGACTCTTC CATCAGGACGCATGTCAGA ACATCAAGGACTCACCCAGC AGCATCACTTCGAGGCGTA
3521 A Y L R D L L I R T H M Q N I K D I T S I H F E A Y 532
3601 CCGTGTGGCGGCTCAAGC AGGGCAGSAGGCCCATGCC AACGGCTGAGGAGAAGGA GCCAGAGCCCGGAGATG
3681 R V K R L N E G S S A M A N G V E E K E P E A P E M 568
3761 AGACGCGCCCTGCCACCC CCGGATGCTCGCCCAAGT CAITTCGTCGCCCCAGC GCGCTCCACACCCCAATT
3841
3921
4001
4081
4161
4241
4321
4401
4481
4561
4641
4721
4801
4881
4961
5041
5121
5201
5281
5361
5441
5521
5601
5681
5761
5841
5921
6001
6081
6161
6241
6321
6401
6481
6561
6641
6721
6801
6881
6961
7041
7121
7201
7281
7361
7441
7521
7601
7681
7761
7841
7921
8001
8081
8161
8241
8321
8401
8481
8561
8641
8721
8801
8881
8961
9041
9121
9201
9281
9361
9441
9521
9601
9681
9761
9841
9921
10001

```

FIG. 5. Sequence of the *MSF* cDNA. The total sequence of *MSF* consists of 2,801 bp. The ORF starts at nucleotide 776 and encodes a putative protein of 568 aa. An in-frame stop codon is present 139 bp upstream of the putative start codon. An arrow indicates the location of the translocation breakpoint. The GTP binding domain is shaded in gray. Numbers to the left refer to nucleotide position, and those to the right indicate amino acid position.

It is particularly interesting that *hCDCrel-1* also is involved in a translocation with *MLL*. It was identified as being fused to *MLL* in infant twins suffering from leukemia with a t(11;22)(q23;q11.2) (33). In this case, *MLL* exon 7 was fused in-frame to *hCDCrel-1* exon 3, which is almost the entire *hCDCrel-1* gene. The fusion would exclude only the first 18 aa of *hCDCrel-1* and would retain the GTP binding region, which is similar to the *MLL-MSF* fusion in our patient. The wild-type *hCDCrel-1* gene is expressed at a high level in brain, heart, and platelets whereas it is expressed at a low level in liver and placenta (48, 49). *hCDCrel-1* is localized in the cytoplasm and is cofractionated with SNAP-25 and synaptosomes in neurons of adult human brain, suggesting that *hCDCrel-1* may contribute to synaptic vesicle transport (49). Among the mamma-

|                 |     |   |
|-----------------|-----|---|
| human MSF       | 287 | G Q S G L G K S T L I N T L F K S K I S R K S V Q P T - S E E R I P K T I E I K S I T H D I E E K G |
| human CDC10     | 38  | G E S G L G K S T L I N S L F L T D L Y S P E - Y P G - P S H R I K K T V O V E O S K V L I K E G G |
| human CDCrel1   | 51  | G E S G L G K S T L V H S L F L T D L Y K D R K L L S - A E E R I S Q T V E I L K H T V D I E E K G |
| human PNUTL2    | 151 | G E S G L G K S T L V N S L F L T D L Y R D R K L L G - A E E R I M Q T V E I T K H A V D I E E K G |
| Drosophila pnut | 149 | G A S G L G K S T L I N S M F L S D I Y N A E Q Y P G - P S L R K K K T V A V E A T K V M L K E N G |
| yeast CDC10     | 39  | G Q S G L G K S T L I N T L F A S H L I D S A T G D D I S A L P V T K T T E M K I S T H T L V E D R |
|                 |     |   |
| human MSF       | 336 | V R M K L T V I D T P G F G D H I N N E N C W Q P I M K F I N D O Y E K Y L Q E E V N I N R K K R I |
| human CDC10     | 86  | V Q L L L T I V D T P G F G D A V D N S N C W Q P V I D Y I D S K F E D Y L N A E S R V N R R O - M |
| human CDCrel1   | 100 | V K L K L T I V D T P G F G D A V N N T E C W K P I T D Y V D Q Q F E Q Y F R D E S G L N R K N - I |
| human PNUTL2    | 200 | V R L R L T I V D T P G F G D A V N N T E C W K P V A E Y I D Q Q F E Q Y F R D E S G L N R K N - I |
| Drosophila pnut | 198 | V N L T L T V V D T P G F G D A V D N S N C W V P I L E Y V D S K Y E E Y L T A E S R V Y R K T - I |
| yeast CDC10     | 89  | V R L N I N V I D T P G F G D F I D N S K A W E P I V K Y I K E Q H S Q Y L R K E L T A Q R E R F I |
|                 |     |   |
| human MSF       | 386 | P D T R V H C C L Y F I A P A T G H S L R P L D I E F M K R L S K V V N I V P V I A K A D           |
| human CDC10     | 135 | P D N R V Q C C L Y F I A P S G H G L K P L D I E F M K R L H E K V N I I P L I A K A D             |
| human CDCrel1   | 149 | Q D N R V H C C L Y F I S P F G H G L R P V D V G F M K A L H E K V N I V P L I A K A D             |
| human PNUTL2    | 249 | Q D N R V H C C L Y F I S P E F G H G L R P L D V E F M K A L H Q R V N I V P I L A K A D           |
| Drosophila pnut | 247 | S D N R V H C C L Y F I A P S G H G L L P L D I A C M O S L S D K V N L V P V I A K A D             |
| yeast CDC10     | 139 | T D T R V H A I L Y F L Q P N G K E L S R L D V E A L K R L T E I A N V I P V I G K S D             |

Fig. 6. The C-terminus sequence of MSF is compared with others of the septin family. Lines above the columns indicate the GTP binding domain (GX<sub>4</sub>GKS-DX<sub>2</sub>G-KXD). White letters in black boxes represent amino acid identity. All of these proteins have 100% sequence identity to a GTP binding domain within this homologous region. Numbers refer to amino acid position of each protein.

lian septin family proteins, murine Nedd5 has been studied functionally. In contrast to *hCDCrel-1*, *Nedd5* is expressed ubiquitously. Depending on the growth state of the cell, the Nedd5 protein accumulates near the contractile ring from anaphase through telophase and condenses into the midbody. In interphase cells, it localizes to fibrous and granular structures and interacts with the actin-based cytoskeletal system. GTP hydrolysis is required for fibrous distribution of Nedd5 (38).

The septin family was first recognized in yeast. The cell division cycle genes *CDC3*, *CDC10*, *CDC11*, and *CDC12* in *S. cerevisiae* encode a family of structural proteins with no significant sequence similarity to other previously known filamentous proteins. They encode component proteins of the bud neck 10-nm filament, which localizes in a ring near the cell surface before bud emergence, remains as a ring at the neck as the bud forms and grows, and persists as a ring after cytokinesis (39–41). Mutations in these genes result in a deficiency in 10-nm bud filaments and cause a disruption of cytokinesis, giving rise to multinucleated cells with abnormal bud growth (50, 51). These proteins play an important role in the cytokinesis of yeast cells. The *Drosophila pnut* gene, which is highly homologous to the septin family, also has been well characterized. The Pnut protein is localized to the cleavage furrow of dividing cells and to the intercellular bridge connecting postmitotic daughter cells (52). In *pnut* mutants, imaginal tissues fail to proliferate and instead develop clusters of large, multinucleated cells (52). A purified complex comprised of three septin family proteins (Pnut, Sep1, and Sep2) forms filaments and hydrolyzes GTP (53). GTP binding and hydrolysis may regulate the polymerization of the septin family or the interaction with other proteins. These findings suggest that the function of the septin family of proteins is conserved from yeast to mammals.

Considering the functional conservation of the septin family, MSF is expected to be located in the cytoplasm and to be associated with cytoskeletal filaments. The MLL-MSF protein produced as a result of the translocation not only probably could alter the Hox cluster genes or other target gene expression regulated by MLL, but also could potentially affect cytokinesis. Among the other *MLL* partner genes cloned to date, AF6 and *hCDCrel-1* have been localized to the cytoplasm. AF6 has been shown to bind to GTP-Ha-Ras (54). In this instance, the chimeric product of MLL and AF6 shifts into the nuclei, and it is suggested that the AT-hook domain of MLL may be important in nuclear localization of the chimeric product (55). The normal function of AF6 or how that function

is altered upon fusion to MLL is not yet known. Localization of the MLL-*hCDCREL-1* fusion still remains to be determined experimentally. Studies on another fusion protein product, which is the result of the *inv(16)*, also may yield useful insights into MLL-MSF function. The *CBFβ* gene, which encodes a subunit of a transcription factor, is fused with *MYH11*, a smooth-muscle myosin heavy chain gene in the *inv(16)*. *CBFβ* localizes to the nucleus, whereas *MYH11* alone localizes with the F-actin on stress fibers and the vinculin in membrane processes (56). The *CBFβ*-*MYH11* fusion protein also colocalizes with actin in the cytoplasm and forms filament-like structure (57–59). In this case, the myosin heavy-chain component of the *CBFβ*-*MYH11* is responsible for sequestering *CBFβ* away from the nucleus, thereby inactivating the *CBFα/β* transcription factor. Whether MLL-MSF is targeted to the nucleus similar to MLL-AF6, or rather, whether it is targeted to the cytoplasm similar to *CBFβ*-*MYH11* remains to be determined. The cellular localization is likely to give insights into the mechanism of leukemogenesis.

This work was supported by grants from the National Institutes of Health (PO1 CA40046 to J.D.R. and N.J.Z.-L. and CA42557 to J.D.R.).

- Smith, M. A., McCaffrey, R. P. & Karp, J. E. (1996) *J. Natl. Cancer Inst.* **88**, 407–418.
- Cimino, G., Moir, D. T., Canaani, O., Williams, K., Crist, W. M., Katzav, S., Cannizzaro, L., Lange, B., Nowell, P. C., Croce, C. M. & Canaani, E. (1991) *Cancer Res.* **51**, 6712–6714.
- Ziemin-van der Poel, S., McCabe, N. R., Gill, H. J., Espinosa, R. III, Patel, Y., Harden, A., Rubinelli, P., Smith, S. D., Le Beau, M. M., Rowley, J. D. & Diaz, M. O. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10735–10739.
- Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D. & Evans, G. A. (1992) *Nat. Genet.* **2**, 113–118.
- Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M. & Canaani, E. (1992) *Cell* **71**, 701–708.
- Tkachuk, D. C., Kohler, S. & Cleary, M. L. (1992) *Cell* **71**, 691–700.
- Gill-Super, H. J., McCabe, N. R., Thirman, M. J., Larson, R. A., Le Beau, M. M., Pedersen-Bjergaard, J., Philip, P., Diaz, M. O. & Rowley, J. D. (1993) *Blood* **82**, 3705–3711.
- Hunger, S. P., Tkachuk, D. C., Amylon, M. D., Link, M. P., Carroll, A. J., Welborn, J. L., Willman, C. L. & Cleary, M. L. (1993) *Blood* **81**, 3197–3203.
- Pedersen-Bjergaard, J. & Rowley, J. D. (1994) *Blood* **83**, 2780–2786.
- Domer, P. H., Head, D. R., Renganathan, N., Raimondi, S. C., Yang, E. & Atlas, M. (1995) *Leukemia* **9**, 1305–1312.

11. Felix, C. A., Lange, B. J., Hosler, M. R., Fertala, J. & Bjornsti, M. A. (1995) *Cancer Res.* **55**, 4287–4292.
12. Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. (1995) *Nature (London)* **378**, 505–508.
13. Hess, J. L., Yu, B. D., Li, B., Hanson, R. & Korsmeyer, S. J. (1997) *Blood* **90**, 1799–1806.
14. Fidanza, V., Melotti, P., Yano, T., Nakamura, T., Bradley, A., Canaani, E., Calabretta, B. & Croce, C. M. (1996) *Cancer Res.* **56**, 1179–1183.
15. Rowley, J. D. (1993) in *Seminars in Cancer Biology*, ed. Rabbitts, T. H. (Academic, London), pp. 377–385.
16. Thirman, M. J., Gill, H. J., Burnett, R. C., Mbangkollo, D., McCabe, N. R., Kobayashi, H., Ziemmin-van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A. A., *et al.* (1993) *N. Engl. J. Med.* **329**, 909–914.
17. Bernard, O. A. & Berger, R. (1995) *Genes Chromosomes Cancer* **13**, 75–85.
18. Waring, P. M. & Cleary, M. L. (1997) *Curr. Top. Microbiol. Immunol.* **220**, 1–23.
19. Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R. P., Lange, B., Crist, W. M., Nowell, P. C., *et al.* (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4631–4635.
20. Prasad, R., Gu, Y., Alder, H., Nakamura, T., Canaani, O., Saito, H., Huebner, K., Gale, R. P., Nowell, P. C., Kuriyama, K., *et al.* (1993) *Cancer Res.* **53**, 5624–5628.
21. Bernard, O., Mauchauffe, M., Mecucci, C., Van Den Berghe, H. & Berger, R. (1994) *Oncogene* **9**, 1039–1045.
22. Parry, P., Wei, Y. & Evans, G. (1994) *Genes Chromosomes Cancer* **11**, 79–84.
23. Prasad, R., Leshkowitz, D., Gu, Y., Alder, H., Nakamura, T., Saito, H., Huebner, K., Berger, R., Croce, C. M. & Canaani, E. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 8107–8111.
24. Rubnitz, J. E., Morrissey, J., Savage, P. A. & Cleary, M. L. (1994) *Blood* **84**, 1747–1752.
25. Thirman, M. J., Levitan, D. A., Kobayashi, H., Simon, M. C. & Rowley, J. D. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 12110–12114.
26. Chaplin, T., Bernard, O., Beverloo, H. B., Saha, V., Hagemeijer, A., Berger, R. & Young, B. D. (1995) *Blood* **86**, 2073–2076.
27. Chaplin, T., Ayton, P., Bernard, O. A., Saha, V., Valle, V. D., Hillion, J., Gregorini, A., Lillington, D., Berger, R. & Young, B. D. (1995) *Blood* **85**, 1435–1441.
28. Tse, W., Zhu, W., Chen, H. S. & Cohen, A. (1995) *Blood* **85**, 650–656.
29. Hillion, J., Le Coniat, M., Jonveaux, P., Berger, R. & Bernard, O. (1997) *Blood* **90**, 3714–3719.
30. So, C. W., Caldas, C., Liu, M. M., Chen, S. J., Huang, Q. H., Gu, L. J., Sham, M. H., Wiedemann, L. M. & Chan L. C. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 2563–2568.
31. Sobulo, O. M., Borrow, J., Tomek, R., Reshmi, S., Harden, A., Schlegelberger, B., Housman, D., Doggett, N. A., Rowley, J. D. & Zeleznik-Le, N. J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 8732–8737.
32. Taki, T., Sako, M., Tsuchida, M. & Hayashi, Y. (1997) *Blood* **89**, 3945–3950.
33. Megonigal, M. D., Rappaport, E. F., Jones, D. H., Williams, T. M., Lovett, B. D., Kelly, K. M., Lerou, P. H., Moulton, T., Budarf, M. L. & Felix C. A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6413–6418.
34. Taki, T., Shibuya, N., Taniwaki, M., Hanada, R., Morishita, K., Bessho, F., Yanagisawa, M. & Hayashi, Y. (1998) *Blood* **92**, 1125–1130.
35. Corral, J., Lavenir, I., Impey, H., Warren, A. J., Forster, A., Larson, T. A., Bell, S., McKenzie, A. N., King, G. & Rabbitts, T. H. (1996) *Cell* **85**, 853–861.
36. Lavau, C., Szilvassy, S. J., Slany, R. & Cleary, M. L. (1997) *EMBO J.* **16**, 4226–4237.
37. Mitelman, F., Mertens, F. & Johansson, B. (1997) *Nat. Genet.* **15** Spec., 417–474.
38. Kinoshita, M., Kumar, S., Mizoguchi, A., Ide, C., Kinoshita, A., Haraguchi, T., Hiraoka, Y. & Noda, M. (1997) *Genes Dev.* **11**, 1535–1547.
39. Haarer, B. K. & Pringle, J. R. (1987) *Mol. Cell. Biol.* **7**, 3678–3687.
40. Ford, S. K. & Pringle, J. R. (1991) *Dev. Genet.* **12**, 281–292.
41. Kim, H. B., Haarer, B. K. & Pringle, J. R. (1991) *J. Cell. Biol.* **112**, 535–544.
42. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
43. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., *et al.* (1991) *Science* **252**, 1651–1656.
44. Broecker, P. L. S., Super, H. G., Thirman, M. J., Pomykala, H., Yonebayashi, Y., Tanabe, S., Zeleznik-Le, N. & Rowley, J. D. (1996) *Blood* **87**, 1912–1922.
45. Felix, C. A. (1998) *Biochim. Biophys. Acta* **1400**, 233–255.
46. Nakatsuru, S., Sudo, K. & Nakamura, Y. (1994) *Biochem. Biophys. Res. Commun.* **202**, 82–87.
47. Zieger, B., Hashimoto, Y. & Ware, J. (1997) *J. Clin. Invest.* **99**, 520–525.
48. Yagi, M., Zieger, B., Roth, G. J. & Ware, J. (1998) *Gene* **212**, 229–236.
49. Caltagarone, J., Rhodes, J., Honer, W. G. & Bowser, R. (1998) *NeuroReport* **9**, 2907–2912.
50. Hartwell, L. H. (1971) *Exp. Cell Res.* **69**, 265–271.
51. Flescher, E. G., Madden, K. & Snyder, M. (1993) *J. Cell. Biol.* **122**, 373–386.
52. Neufeld, T. P. & Rubin, G. M. (1994) *Cell* **77**, 371–379.
53. Field, C. M., al-Awar, O., Rosenblatt, J., Wong, M. L., Alberts, B. & Mitchison, T. J. (1996) *J. Cell. Biol.* **133**, 605–616.
54. Kuriyama, M., Harada, N., Kuroda, S., Yamamoto, T., Nakafuku, M., Iwamatsu, A., Yamamoto, D., Prasad, R., Croce, C., Canaani, E. & Kaibuchi, K. (1996) *J. Biol. Chem.* **271**, 607–610.
55. Joh, T., Yamamoto, K., Kagami, Y., Kakuda, H., Sato, T., Yamamoto, T., Takahashi, T., Ueda, R., Kaibuchi, K. & Seto, M. (1997) *Oncogene* **15**, 1681–1687.
56. Tanaka, Y., Watanabe, T., Chiba, N., Niki, M., Kuroiwa, Y., Nishihira, T., Satomi, S., Ito, Y. & Satake, M. (1997) *Oncogene* **15**, 677–683.
57. Lu, J., Maruyama, M., Satake, M., Bae, S. C., Ogawa, E., Kagoshima, H., Shigesada, K. & Ito, Y. (1995) *Mol. Cell. Biol.* **15**, 1651–1661.
58. Wijmenga, C., Gregory, P. E., Hajra, A., Schrock, E., Ried, T., Eils, R., Liu, P. P. & Collins, F. S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 1630–1635.
59. Adya, N., Stacy, T., Speck, N. A. & Liu, P. P. (1998) *Mol. Cell. Biol.* **18**, 7432–7443.