SHORT COMMUNICATION

Identification of a New Murine *runt* Domain-Containing Gene, *Cbfa3*, and Localization of the Human Homolog, *CBFA3*, to Chromosome 1p35–pter

CISCA WIJMENGA,* NANCY A. SPECK,† NICHOLAS C. DRACOPOLI,‡
MARTEN H. HOFKER,§ PU LIU,* AND FRANCIS S. COLLINS*,[1]

*Laboratory of Gene Transfer and ‡Laboratory of Genetic Disease Research, National Center for Human Genome Research, 9000 Rockville Pike, Bethesda, Maryland 20892; †Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire; and §MGC-Department of Human Genetics, Leiden University, Leiden, The Netherlands

Received October 18, 1994; revised December 11, 1994

Core binding factor (CBF) is a heterodimeric transcription factor composed of two distinct subunits. The monomeric β subunit is ubiquitously expressed, whereas expression of the three α subunits isolated previously seems to be restricted mainly to hematopoietic tissues. To isolate additional α genes, degenerate oligonucleotides derived from the *runt* domain—a region shared by all α genes—were used for screening cDNA libraries. A 228-bp fragment was isolated from a mouse thymus cDNA library, which showed 82 and 76% DNA sequence identity to the previously isolated murine α genes, *Cbfa1* and *Cbfa2*. This novel α gene was named *Cbfa3*. The corresponding sequence from the human homolog *CBFA3* was obtained by cosmid cloning and sequencing of the appropriate restriction fragment. The corresponding regions of mouse *Cbfa3* and human *CBFA3* show 91% nucleotide identity and 100% protein identity. *In situ* hybridization and physical mapping of somatic cell hybrids localized *CBFA3* to chromosome 1p35–pter.

Core binding factor (CBF; also known as PEBP2) is a novel transcriptional regulatory protein detected in various lymphoid and myeloid cells and implicated in the regulation of viral and cellular genes specifically expressed in these cells (7, 20, 24–26, 28). Although CBF is undetectable in F9 embryonal carcinoma cells, it is induced following differentiation of these cells by retinoic acid (6, 10), suggesting that CBF may play a role in mammalian development as well.

CBF is a heterodimer composed of α and β subunits (22, 23, 31). Thus far, two distinct genes encoding CBF α subunits were identified, *Cbfa2/CBFA2* (formerly *aml1/AML1;* from mouse and human) (1, 5) and *Cbfa1* (formerly *pebp2aA;* from mouse; 23). Recently, two additional human α genes have been isolated: *AML2* and *AML3*, the latter apparently representing the human homolog of *Cbfa1* (11). In keeping with the current nomenclature, *AML2* and *AML3* will be referred to as *CBFA3* and *CBFA1*, respectively. The α subunits all share a 128-amino-acid region that is highly homologous to the product of the *Drosophila* segmentation gene *runt* (9). This *runt* domain is responsible for dimerizing with the β subunit (17, 23) and binds to DNA by recognizing the consensus sequence PuACCPuCA (8, 15).

Various sizes of *Cbfa2* transcripts, which probably originate from alternative RNA splicing, have been detected at different expression levels in T and B cells (1), lung, heart, spleen, thymus, and ovary. Transcripts have not been detected in brain, liver, kidney, or testis (18). A very similar pattern of expression was observed for human *CBFA2* and *CBFA3* in multiple hematopoietic lineages (11, 18). However, the expression of *Cbfa1* seems to be highly cell type-specific: it is abundantly expressed in T-lymphocyte-derived cell lines but is undetectable in B-cell lines (23).

CBFβ does not bind DNA directly but increases the affinity of the α subunit for DNA (22). In contrast to any of the identified α subunits, the β subunit, *CBFB*, is ubiquitously expressed (31). One possible explanation for this inconsistency might be the existence of additional differentially expressed, and as-yet unidentified, α subunit genes. In this communication we describe the discovery of a new murine *runt* domain-containing gene (*Cbfa3*) and the mapping of its human homolog (*CBFA3*) to 1p35–pter.

Using a 64-fold 5′ degenerate oligonucleotide primer (CGGGATCCGTNGCNTTYAARGT) and a 128-fold 3′ degenerate oligonucleotide primer (CGGAATTCTANGTNGCNACYTG) to the amino acids VAFKV (aa 130–134) and QVATY (aa 201–205) from the *runt* homology region (aa positions are according to the *Cbfa1* sequence; 23), a 228-bp fragment was obtained by PCR amplification from λ phage DNA isolated from a liquid phage lysate of a mouse thymus Lambda ZAP cDNA library (Stratagene) following the protocol described

[1] To whom correspondence should be addressed at the National Center for Human Genome Research, Building 38A, Room 605, 38 Library Drive MSC6050, Bethesda, MD 20892-6050. Telephone: (301) 496-0844. Fax: (301) 402-0837.

GENOMICS **26,** 611–614 (1995)
0888-7543/95 $6.00

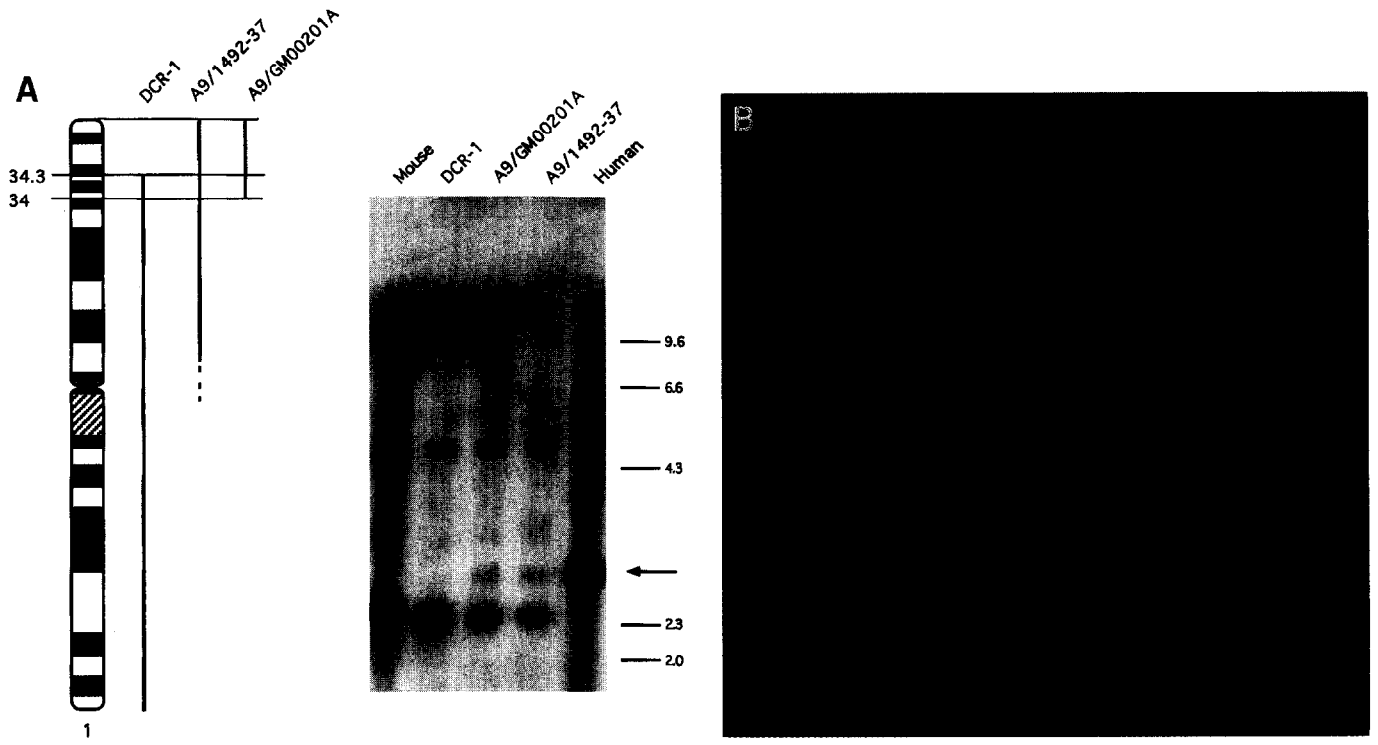

FIG. 1. Mapping of *CBFA3* to distal 1p. (**A**) Southern analysis with pR1 of *Hin*dIII-digested DNA of mouse, human, and three hybrid cell lines containing different parts of chromosome 1. The human-specific fragment (indicated by an arrow) is present in cell lines A9/1492-37 and A9/GM00201A, indicating that the *CBFA3* must reside in the region 1p35–pter. (**B**) *In situ* hybridization of a *CBFA3* cosmid to normal human metaphase spreads. Fluorescent signals are located on distal 1p.

previously (31). This PCR product was subcloned into the *Bam*HI and *Eco*RI sites of pBluescript SK(+) (pR1) and sequenced using the dideoxy-chain termination method with reagents provided in a Sequenase version 2.0 kit (U.S. Biochemical). Sequence comparison revealed 82 and 76% identity to *Cbfa1* and *Cbfa2,* respectively, implying that the PCR product was amplified from a novel murine gene. However, the deduced amino acid sequence showed 100% identity with the recently discovered *CBFA3,* implying that we have identified the murine homolog of *CBFA3.* Therefore, this gene was called *Cbfa3.*

Expression of *Cbfa3* in various mouse tissues was examined by Northern blot analysis, as described previously (31). However, no signal was obtained in any lane, suggesting that the gene is expressed at very low levels, or not at all, in the tissues examined, which did not include any hematopoietic cell lines. Unless expressed exclusively in hematopoietic tissues, the failure to detect a signal might also be due to the small size of probe pR1.

The chromosomal localization of the human homolog of *Cbfa3* was determined by analyzing a somatic cell hybrid panel. A human/rodent somatic cell hybrid panel was obtained from BIOS Laboratories (New Haven, CT). The presence or absence of *CBFA3* was determined by Southern blot analysis. The results showed that the *CBFA3* gene is located on chromosome 1 (0% discordance; data not shown). Subsequently, a more precise localization was obtained using mouse/human hybrid cell lines containing different parts of chromosome 1. These hybrids included DCR-1 (16), A9/GM00201A (Human Genetic Cell Repository, Camden, NJ), and A9/1492-37 (4). The breakpoints of these hybrid cell lines and their representative idiograms are depicted in Fig. 1A. Based on these data, the human *CBFA3* gene maps to 1p35–pter. This agrees well with the results reported by Levanon *et al.* (11), which appeared as this paper was being prepared.

In addition, human cosmid clones were isolated by screening a total human genomic cosmid library (MHH, unpublished) using pR1. A total of 11 positive clones that all share multiple restriction fragments using various enzymes were obtained (data not shown). To investigate the authenticity of the cosmids, a 2.3-kb *Pst*I fragment present in all cosmids and hybridizing with pR1 was subcloned in pBluescript SK(+) and sequenced using an ABI 373A DNA sequencer (Applied Biosystems). Synthetic oligonucleotides were designed to complete the entire sequence. The sequence (GenBank Accession No. U14520) was used to screen the GenBank and EMBL databases. There is a 161-bp region with 77 to 79% similarity to the *runt* domain of *CBFA2, Cbfa2,* and *Cbfa1.* In addition, this 161-bp region revealed 91% homology with the newly identified *Cbfa3* and represents a single exon. The deduced amino acid sequence of this exon shows between 93 and 100% identity with the *Cbfa1* and *CBFA3* gene and the *CBFA2/Cbfa2* genes (Fig. 2).

Two of the *CBFA3*-cosmid clones were used to per-

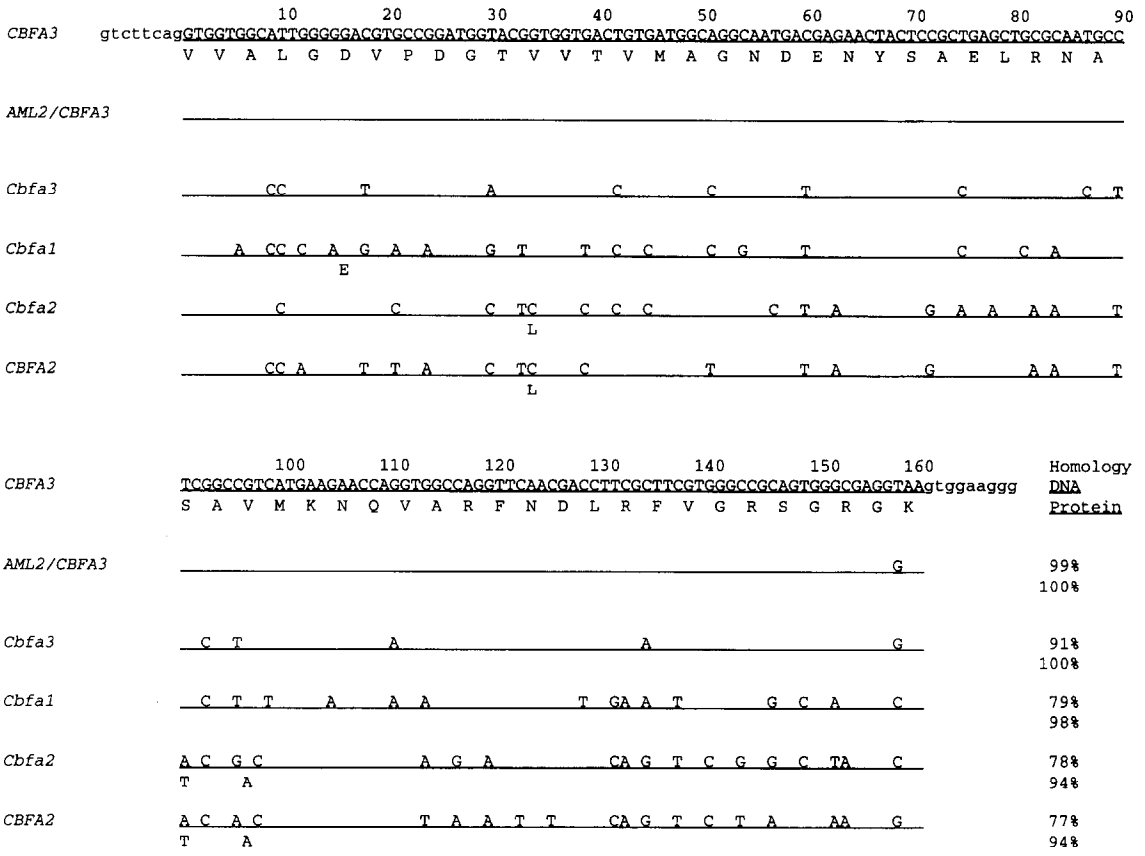

FIG. 2. Alignment of the nucleotide and predicted amino acid sequence of the 161-bp *CBFA3* exon (capital letters) with *AML2/CBFA3* (11), *Cbfa3, Cbfa1, Cbfa2,* and *CBFA2.* Only those residues that are not conserved between the sequences are shown. Adjacent *CBFA3* intronic sequences are shown in lowercase letters.

form fluorescence *in situ* hybridization (FISH), according to standard protocols (12). Analysis of 30 metaphases confirmed in all cases a localization of the *CBFA3* gene to the distal part of the long arm of chromosome 1 (Fig. 1B).

In this report a new murine *runt* domain-containing gene, *Cbfa3,* that is identified as the human homolog of *CBFA3* has been discovered. The chromosomal localization of *CBFA3* to 1p35–pter has been confirmed. Based on comparative linkage mapping, *Cbfa3* is predicted to be localized to mouse chromosome 4.

In parallel with the previously isolated α subunits, it might be expected that the *runt* domain in CBFα3/Cbfα3 forms a heterodimer with PEBP2β/CBFβ as well. Whether these different α genes have distinct functions remains to be elucidated. However, it is attractive to assume that each α/β combination forms a cell-type/tissue-specific transcription factor whose specificity is determined by the α gene.

Both *CBFA2* and *CBFB* are implicated in myeloid leukemia through chromosomal abnormalities leading to fusion transcripts (5, 13, 21). Translocations involving 1p36 are found in a variety of malignant diseases, both solid tumors and hematopoietic malignancies. Translocations t(1;3)(p36;q21) and t(1;7)(p36;q32) occur in patients with myeloproliferative diseases (19, 27, 30); the t(1;3) is also found in patients with acute myeloid leukemia (2). A t(1;13)(p36;q14) has been reported in alveolar rhabdomyosarcoma (3). Finally, a t(1;2)(p36;p24) has been observed in a small number of uterine leiomyomas (14, 29). It will be interesting to investigate whether the *CBFA3* gene is rearranged in any of these malignancies.

REFERENCES

1. Bae, S. C., Yanaguchi-Iwai, Y., Ogawa, E., Maruyama, M., Inuzuka, M., Kagoshima, H., Shigesada, K., Satake, M., and Ito, Y. (1993). Isolation of PEBP2αB cDNA representing the mouse homolog of human acute myeloid leukemia gene, AML1. *Oncogene* **8:** 809–814.
2. Bloomfield, C. D., Garson, O. M., Voilin, L., Knuutila, S., and de la Chapelle, A. (1985). t(1;3)(p36;q21) in acute nonlymphocytic leukemia: A new cytogenetic-clinicopathologic association. *Blood* **66:** 1409–1413.
3. Douglass, E. C., Rowe, S. T., Valentine, M., Parham, D. M., Berkow, R., and Bowman, W. P. (1991). A second non-random translocation t(1;3)(p36;q14) in alveolar rhabdomyosarcoma. *Cancer Genet. Cytogenet.* **56:** 121.
4. Dracopoli, N. C., Stanger, B. Z., Ito, C. Y., Call, K. M., Lincoln, S. E., Lander, E. S., and Housman, D. E. (1988). A genetic linkage map of 27 loci from PND to FY on the short arm of human chromosome 1. *Am. J. Hum. Genet.* **43:** 462–470.
5. Erickson, P., Gao, J., Chang, K.-S., Look, T., Whisenant, E., Raimondi, S., Lasher, R., Trujillo, J., Rowley, J., and Drabkin, H. (1992). Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML/ETO,

with similarity to *Drosophila* segmentation gene, *runt*. *Blood* **80:** 1825–1831.

6. Furukawa, K., Yamaguchi, Y., Ogawa, E., Shigesada, K., Satake, M., and Ito, Y. (1990). A ubiquitous repressor interacting with an F9 cell-specific silencer and its functional suppression by differentiated cell-specific positive factors. *Cell Growth Differ.* **1:** 135–147.
7. Hsiang, Y. H., Spencer, D., Wang, S., Speck, N. A., and Raulet, D. H. (1993). The role of viral enhancer core motif-related sequences in regulating T-cell receptor-γ and -δ gene expression. *J. Immunol.* **150:** 3905–3916.
8. Kamachi, Y., Ogawa, E., Asano, M., Ishida, S., Murakami, Y., Satake, M., Ito, Y., and Shigesada, K. (1990). Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J. Virol.* **64:** 4808–4819.
9. Kania, M. A., Bonner, A. S., Duffy, J. B., and Gergen, J. P. (1990). The *Drosophila* segmentation gene *runt* encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. *Genes Dev.* **4:** 1701–1713.
10. Kryszke, M.-H., Piette, J., and Yaniv, M. (1987). Induction of a factor that binds to the polyomavirus A enhancer on differentiation of embryonal carcinoma cells. *Nature* **328:** 254–256.
11. Levanon, D., Negreanu, V., Bernstein, Y., Bar-Am, I., Avivi, L., and Groner, Y. (1994). AML1, AML2, and AML3, the human members of the *runt* domain gene-family: cDNA structure, expression and chromosomal localization. *Genomics* **23:** 425–432.
12. Liu, P., Claxton, D. F., Marlton, P., Hajra, A., Siciliano, J., Freedman, M., Chandrasekharappa, S. C., Yanagisawa, K., Stallings, R. L., Collins, F. S., and Siciliano, M. (1993). Identification of yeast artificial chromosomes containing the inversion 16 p-arm breakpoint associated with acute myelomonocytic leukemia. *Blood* **82:** 716–721.
13. Liu, P., Tarle, S. A., Hajra, A., Claxton, D. F., Marlton, P., Freedman, M., Siciliano, M. J., and Collins, F. S. (1993). Fusion between transcription factor CBFβ/PEBP2β and a myosin heavy chain in acute myeloid leukemia. *Science* **261:** 1041–1044.
14. Mark, J., Havel, G., Grepp, C., Dahlenfors, R., and Wedell, B. (1988). Cytogenetical observations in human benign uterine leiomyomas. *Anticancer Res.* **8:** 621–626.
15. Melnikova, I. N., Crute, B. E., Wang, S., and Speck, N. A. (1993). Sequence specificity of the core-binding factor. *J. Virol.* **67:** 2408–2411.
16. Menon, A. G., Ledbetter, D. H., Rich, D. C., Seizinger, B. R., Rouleau, G. A., Michels, V. F., Schmidt, M. A., Dewald, G., DallaTorre, C. M., Haines, J. L., and Gusella, J. F. (1989). Characterization of a translocation within the von Recklinghausen neurofibromatosis region of chromosome 17. *Genomics* **5:** 245–249.
17. Meyers, S., Downing, J. R., and Hiebert, S. W. (1993). Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: The *runt* homology domain is required for DNA binding and protein–protein interactions. *Mol. Cell. Biol.* **13:** 6336–6345.
18. Miyoshi, H., Kozu, T., Shimizu, K., Enomopo, K., Maseki, N., Kaneko, Y., Kamada, N., and Ohki, M. (1993). The t(8;21) translocation in acute myeloid leukemia results in production of an AML1–MTG8 fusion transcript. *EMBO J.* **12:** 2715–2721.
19. Moir, D. J., Jones, P. A. E., Pearson, J., Duncan, J. R., Cook, P., and Buckle, V. J. (1984). A new translocation t(1;3)(p36;q21) in myelodysplastic disorders. *Blood* **64:** 243–248.
20. Nuchprayoon, I., Meyers, S., Scott, L. M., Suzow, J., Hiebert, S., and Friedman, A. D. (1994). PEBP2/CBF, the murine homolog of the human myeloid AML1 and PEBP2β/CBFβ proto-oncoproteins, regulates the murine myeloperoxidase and neutrophil elastase genes in immature myeloid cells. *Mol. Cell. Biol.* **14:** 5558–5568.
21. Nucifora, G., Begy, C. R., Kobayashi, H., Roulston, D., Claxton, D., Pedersen-Bjergaard, J., Parganas, E., Ihle, J. N., and Rowley, J. D. (1994). Consistent intergenic splicing and production of multiple transcripts between AML1 at 21q22 and unrelated genes at 3q26 in (3;21)(q26;q22) translocations. *Proc. Natl. Acad. Sci. USA* **91:** 4004–4008.
22. Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto, M., Ito, Y., and Shigesada, K. (1993). Molecular cloning and characterization of PEBP2β, the heterodimeric partner of a novel *Drosophila runt*-related DNA binding protein PEBP2α. *Virology* **194:** 314–331.
23. Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigessada, K., and Ito, Y. (1993). PEBP2/PEA2 represents a new family of transcription factor homologous to the products of the Drosophila runt and the human AML1. *Proc. Natl. Acad. Sci. USA* **90:** 6859–6863.
24. Redondo, J. M., Pfohl, J. L., Hernandez-Munain, C., Wang, S., Speck, N. A., and Krangel, M. S. (1992). Indistinguishable nuclear factor binding to functional core sites of the T-cell receptor δ and murine leukemia virus enhancers. *Mol. Cell. Biol.* **12:** 4817–4823.
25. Satake, M., Inuzuka, M., Shigesada, K., Oikawa, T., and Ito, Y. (1992). Differential expression of subspecies of polyomavirus and murine leukemia virus enhancer core binding protein, PEBP2, in various hematopoietic cells. *Jpn. J. Cancer Res.* **83:** 714–722.
26. Speck, N. A., Renjifo, B., Golemis, E., Frederickson, T. N., Hartley, J. W., and Hopkins, N. (1990). Mutations of the core or adjacent LVb elements of the Moloney murine leukemia virus enhancer alters disease specificity. *Genes Dev.* **4:** 233–242.
27. Stefaescu, D. T., Colita, D., Nicoara, S., and Calin, G. (1994). t(1;7)(p36;q32): A new recurring abnormality in primary myelodysplastic syndrome. *Cancer Genet. Cytogenet.* **75:** 103–105.
28. Thornell, A., Hallberg, B., and Grundstrom, T. (1988). Differential protein binding in lymphocytes to a sequence in the enhancer of the mouse retrovirus SL3-3. *Mol. Cell. Biol.* **8:** 1625–1637.
29. Vanni, R., Dal Cin, P., and Van Den Berghe, H. (1990). Is the chromosome band 1p36 another hot-spot for rearrangements in uterine leiomyoma? *Genes Chrom. Cancer* **2:** 255–256.
30. Viguie, F., Marie, J.-P., Poler, F., and Barnadou, A. (1986). Three cases of preleukemic myelodysplastic disorders with the same translocation t(1;3). *Cancer Genet. Cytogenet.* **19:** 213–218.
31. Wang, S., Wang, Q., Crute, B. E., Melnikova, I. N., Keller, S. R., and Speck, N. A. (1993). Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor (1993). *Mol. Cell. Biol.* **13:** 3324–3339.